

# The Blackwell Companion to Substance Dualism

# Blackwell Companions to Philosophy

This outstanding student reference series offers a comprehensive and authoritative survey of philosophy as a whole. Written by today's leading philosophers, each volume provides lucid and engaging coverage of the key

figures, terms, topics, and problems of the field. Taken together, the volumes provide the ideal basis for course use, representing an unparalleled work of reference for students and specialists alike.

## Already published in the series

1. The Blackwell Companion to Philosophy, Second Edition  
*Edited by Nicholas Bunnin and Eric Tsui-James*
2. A Companion to Ethics  
*Edited by Peter Singer*
3. A Companion to Aesthetics, Second Edition  
*Edited by Stephen Davies, Kathleen Marie Higgins, Robert Hopkins, Robert Stecker, and David E. Cooper*
4. A Companion to Epistemology, Second Edition  
*Edited by Jonathan Dancy, Ernest Sosa and Matthias Steup*
5. A Companion to Contemporary Political Philosophy (two-volume set), Second Edition  
*Edited by Robert E. Goodin and Philip Pettit*
6. A Companion to Philosophy of Mind  
*Edited by Samuel Guttenplan*
7. A Companion to Metaphysics, Second Edition  
*Edited by Jaegwon Kim, Ernest Sosa and Gary S. Rosenkrantz*
8. A Companion to Philosophy of Law and Legal Theory, Second Edition  
*Edited by Dennis Patterson*
9. A Companion to Philosophy of Religion, Second Edition  
*Edited by Charles Taliaferro, Paul Draper, and Philip L. Quinn*
10. A Companion to the Philosophy of Language  
*Edited by Bob Hale and Crispin Wright*
11. A Companion to World Philosophies  
*Edited by Eliot Deutsch and Ron Bontekoe*
12. A Companion to Continental Philosophy  
*Edited by Simon Critchley and William Schroeder*
13. A Companion to Feminist Philosophy  
*Edited by Alison M. Jaggar and Iris Marion Young*
14. A Companion to Cognitive Science  
*Edited by William Bechtel and George Graham*
15. A Companion to Bioethics, Second Edition  
*Edited by Helga Kuhse and Peter Singer*
16. A Companion to the Philosophers  
*Edited by Robert L. Arrington*
17. A Companion to Business Ethics  
*Edited by Robert E. Frederick*
18. A Companion to the Philosophy of Science  
*Edited by W. H. Newton-Smith*
19. A Companion to Environmental Philosophy  
*Edited by Dale Jamieson*
20. A Companion to Analytic Philosophy  
*Edited by A. P. Martinich and David Sosa*
21. A Companion to Genethics  
*Edited by Justine Burley and John Harris*
22. A Companion to Philosophical Logic  
*Edited by Dale Jacquette*
23. A Companion to Early Modern Philosophy  
*Edited by Steven Nadler*
24. A Companion to Philosophy in the Middle Ages  
*Edited by Jorge J. E. Gracia and Timothy B. Noone*
25. A Companion to African-American Philosophy  
*Edited by Tommy L. Lott and John P. Pittman*
26. A Companion to Applied Ethics  
*Edited by R. G. Frey and Christopher Heath Wellman*
27. A Companion to the Philosophy of Education  
*Edited by Randall Curren*
28. A Companion to African Philosophy  
*Edited by Kwasi Wiredu*
29. A Companion to Heidegger  
*Edited by Hubert L. Dreyfus and Mark A. Wrathall*
30. A Companion to Rationalism  
*Edited by Alan Nelson*
31. A Companion to Pragmatism  
*Edited by John R. Shook and Joseph Margolis*
32. A Companion to Ancient Philosophy  
*Edited by Mary Louise Gill and Pierre Pellegrin*
33. A Companion to Nietzsche  
*Edited by Keith Ansell Pearson*
34. A Companion to Socrates  
*Edited by Sara Ahbel-Rappe and Rachana Kamtekar*
35. A Companion to Phenomenology and Existentialism  
*Edited by Hubert L. Dreyfus and Mark A. Wrathall*
36. A Companion to Kant  
*Edited by Graham Bird*
37. A Companion to Plato  
*Edited by Hugh H. Benson*
38. A Companion to Descartes  
*Edited by Janet Broughton and John Carriero*
39. A Companion to the Philosophy of Biology  
*Edited by Sahotra Sarkar and Anya Plutynski*
40. A Companion to Hume  
*Edited by Elizabeth S. Radcliffe*
41. A Companion to the Philosophy of History and Historiography  
*Edited by Aviezer Tucker*
42. A Companion to Aristotle  
*Edited by Georgios Anagnostopoulos*
43. A Companion to the Philosophy of Technology  
*Edited by Jan-Kyrre Berg Olsen, Stig Andur Pedersen, and Vincent F. Hendricks*
44. A Companion to Latin American Philosophy  
*Edited by Susana Nuccetelli, Ofelia Schutte, and Otávio Bueno*
45. A Companion to the Philosophy of Literature  
*Edited by Garry L. Hagberg and Walter Jost*
46. A Companion to the Philosophy of Action  
*Edited by Timothy O'Connor and Constantine Sandis*
47. A Companion to Relativism  
*Edited by Steven D. Hales*
48. A Companion to Hegel  
*Edited by Stephen Houlgate and Michael Baur*
49. A Companion to Schopenhauer  
*Edited by Bart Vandenabeele*
50. A Companion to Buddhist Philosophy  
*Edited by Steven M. Emmanuel*
51. A Companion to Foucault  
*Edited by Christopher Falzon, Timothy O'Leary, and Jana Sawicki*
52. A Companion to the Philosophy of Time  
*Edited by Heather Dyke and Adrian Bardon*
53. A Companion to Donald Davidson  
*Edited by Ernest Lepore and Kirk Ludwig*
54. A Companion to Rawls  
*Edited by Jon Mandle and David Reidy*
55. A Companion to W.V.O. Quine  
*Edited by Gilbert Harman and Ernest Lepore*
56. A Companion to Derrida  
*Edited by Zeynep Direk and Leonard Lawlor*
57. A Companion to David Lewis  
*Edited by Barry Loewer and Jonathan Schaffer*
58. A Companion to Kierkegaard  
*Edited by Jon Stewart*
59. A Companion to Locke  
*Edited by Matthew Stuart*
60. The Blackwell Companion to Hermeneutics  
*Edited by Niall Keane and Chris Lawn*
61. A Companion to Ayn Rand  
*Edited by Allan Gotthelf and Gregory Salmieri*
62. The Blackwell Companion to Naturalism  
*Edited by Kelly James Clark*
63. A Companion to the Philosophy of Language (two-volume set), Second Edition  
*Edited by Bob Hale*

# THE BLACKWELL COMPANION TO SUBSTANCE DUALISM

*Edited by*

Jonathan J. Loose, Angus J. L. Menuge,  
and  
J. P. Moreland

**WILEY Blackwell**

This edition first published 2018  
© 2018 John Wiley and Sons Inc.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, except as permitted by law. Advice on how to obtain permission to reuse material from this title is available at <http://www.wiley.com/go/permissions>.

The right of Jonathan J. Loose, Angus J. L. Menuge, and J. P. Moreland to be identified as the authors of the editorial material in this work has been asserted in accordance with law.

*Registered Office*

John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, USA

*Editorial Office*

9600 Garsington Road, Oxford, OX4 2DQ, UK

For details of our global editorial offices, customer services, and more information about Wiley products visit us at [www.wiley.com](http://www.wiley.com).

Wiley also publishes its books in a variety of electronic formats and by print-on-demand. Some content that appears in standard print versions of this book may not be available in other formats.

*Limit of Liability/Disclaimer of Warranty*

While the publisher and authors have used their best efforts in preparing this work, they make no representations or warranties with respect to the accuracy or completeness of the contents of this work and specifically disclaim all warranties, including without limitation any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives, written sales materials or promotional statements for this work. The fact that an organization, website, or product is referred to in this work as a citation and/or potential source of further information does not mean that the publisher and authors endorse the information or services the organization, website, or product may provide or recommendations it may make. This work is sold with the understanding that the publisher is not engaged in rendering professional services. The advice and strategies contained herein may not be suitable for your situation. You should consult with a specialist where appropriate. Further, readers should be aware that websites listed in this work may have changed or disappeared between when this work was written and when it is read. Neither the publisher nor authors shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

*Library of Congress Cataloging-in-Publication Data*

Names: Loose, Jonathan, editor.

Title: The Blackwell companion to substance dualism / edited by Jonathan Loose, Angus Menuge, and J.P. Moreland.

Description: Hoboken : Wiley, 2018. | Series: Blackwell companions to philosophy | Includes bibliographical references and index. |

Identifiers: LCCN 2017049916 (print) | LCCN 2017059533 (ebook) |

ISBN 9781119375296 (pdf) | ISBN 9781119375302 (epub) | ISBN 9781119375265 (cloth)

Subjects: LCSH: Mind and body. | Dualism. | Materialism.

Classification: LCC B105.M53 (ebook) | LCC B105.M53 B53 2018 (print) | DDC 147/.4--dc23

LC record available at <https://lcn.loc.gov/2017049916>

Cover image: dianaarturovna/Gettyimages

Cover design by Wiley

Set in 10/12.5 pt MinionPro-Regular by Thomson Digital, Noida, India

# Contents

<i>Notes on Contributors</i>	<i>ix</i>
1 Introduction: Substance Dualism and Its Physicalist Rivals <i>Jonathan J. Loose, Angus J. L. Menuge, and J. P. Moreland</i>	1
2 Redressing Substance Dualism <i>William G. Lycan</i>	22
<b>Part I. Articulating Substance Dualism</b>	<b>41</b>
3 Substance Dualism: A Defense <i>Charles Taliaferro</i>	43
<b><i>Debating Emergent Dualism</i></b>	<b>61</b>
4 The Case for Emergent Dualism <i>William Hasker</i>	62
5 Against Emergent Dualism <i>Brandon L. Rickabaugh</i>	73
<b><i>Debating Thomistic Dualism</i></b>	<b>87</b>
6 Aquinas on the Human Soul <i>Edward Feser</i>	88
7 In Defense of a Thomistic-like Dualism <i>J. P. Moreland</i>	102
8 A Critique of Thomistic Dualism <i>William Hasker</i>	123
<b><i>Debating Cartesian Dualism</i></b>	<b>132</b>
9 Cartesian Substance Dualism <i>Richard Swinburne</i>	133

10	Against Cartesian Dualism <i>Jaegwon Kim</i>	152
11	Non-Cartesian Substance Dualism <i>E. J. Lowe</i>	168
	<b><i>Debating the Unity of Consciousness</i></b>	<b>183</b>
12	Substance Dualism and the Unity of Consciousness <i>J. P. Moreland</i>	184
13	Problems with Unity of Consciousness Arguments for Substance Dualism <i>Tim Bayne</i>	208
	<b><i>Debating Near-Death Experiences</i></b>	<b>226</b>
14	Evidential Near-Death Experiences <i>Gary R. Habermas</i>	227
15	The Phenomenology of Near-Death and Out-of-Body Experiences: No Heavenly Excursion for “Soul” <i>Michael N. Marsh</i>	247
16	Why Reject Substance Dualism? <i>Ian Ravenscroft</i>	267
	<b>Part II. Alternatives to Substance Dualism</b>	<b>283</b>
17	Why Should a Christian Embrace Materialism (about Human Persons)? <i>Kevin Corcoran</i>	285
	<b><i>Debating Animalism</i></b>	<b>296</b>
18	For Animalism <i>Eric T. Olson</i>	297
19	Against Animalism <i>Stewart Goetz</i>	307
	<b><i>Debating Nonreductive Physicalism</i></b>	<b>316</b>
20	For Nonreductive Physicalism <i>Nancey Claire Murphy</i>	317
21	Against Nonreductive Physicalism <i>Joshua Rasmussen</i>	328
	<b><i>Debating Constitutionalism</i></b>	<b>340</b>
22	Constitutionalism: Alternative to Substance Dualism <i>Lynne Rudder Baker</i>	341
23	Against Constitutionalism <i>Ross Inman</i>	351

<b><i>Debating Emergent Individualism</i></b>	<b>368</b>
24 For Emergent Individualism <i>Timothy O'Connor</i>	369
25 Against Emergent Individualism <i>Robert C. Koons</i>	377
26 Why Reject Christian Physicalism? <i>Angus J. L. Menuge</i>	394
 <b>Part III. Substance Dualism, Theology, and the Bible</b>	 <b>411</b>
<b><i>Debating Biblical Anthropology</i></b>	<b>412</b>
27 Biblical Anthropology is Holistic and Dualistic <i>John W. Cooper</i>	413
28 The Strange Case of the Vanishing Soul <i>Joel B. Green</i>	427
 <b><i>Debating the Incarnation</i></b>	 <b>439</b>
29 Dualism Offers the Best Account of the Incarnation <i>Luke Van Horn</i>	440
30 The Word Made Flesh: Dualism, Physicalism, and the Incarnation <i>Trenton Merricks</i>	452
 <b><i>Debating the Resurrection</i></b>	 <b>469</b>
31 Materialism Most Miserable: The Prospects for Dualist and Physicalist Accounts of Resurrection <i>Jonathan J. Loose</i>	470
32 I Look for the Resurrection of the Dead and the Life of the World to Come <i>Peter van Inwagen</i>	488
 <i>Index</i>	 501





# Notes on Contributors

**Lynne Rudder Baker** is Professor of Philosophy Emeritus (formerly distinguished professor) at the University of Massachusetts Amherst and is the author of *Naturalism and the First-Person Perspective* (Oxford, 2013), *The Metaphysics of Everyday Life* (Cambridge, 2007), *Persons and Bodies* (Cambridge, 2000), *Explaining Attitudes* (Cambridge, 1995), and *Saving Belief* (Princeton, 1987), as well as numerous articles and book chapters in the philosophy of mind, metaphysics, and philosophical theology. She also has spoken widely in over twenty countries on five continents. See <http://people.umass.edu/lrb>.

**Tim Bayne** is a philosopher of mind and cognitive science, with a particular interest in the nature of consciousness. He is currently professor of philosophy at Monash University (Melbourne), having taught previously at Macquarie University, the University of Western Ontario, the University of Manchester, and the University of Oxford. He is the author of *The Unity of Consciousness* (2010) and *Thought: A Very Short Introduction* (2013); and an editor of *Delusion and Self-Deception* (Psychology Press, 2008), *The Oxford Companion to Consciousness* (Oxford University Press, 2009), and *Cognitive Phenomenology* (Oxford University Press, 2011).

**John W. Cooper** taught philosophy at Calvin College and since 1985 has been professor of philosophical theology at Calvin Theological Seminary. Among his publications are *Body, Soul and Life Everlasting: Biblical Anthropology and the Monism-Dualism Debate* (Eerdmans, 1989) and *Panentheism – The Other God of the Philosophers: From Plato to the Present* (Baker Academic, 2006).

**Kevin Corcoran** is Professor of Philosophy at Calvin College in Grand Rapids, Michigan. He is the author or editor of three books – *Soul, Body and Survival* (Cornell University Press, 2001), *Rethinking Human Nature* (Baker Academic, 2006) and *Church in the Present Tense* (Baker, 2011)—as well as many articles, mostly dealing with the metaphysics of human nature and the philosophy of mind. In 2012, he was included in Princeton Review's *The Best 300 Professors*.

**Edward Feser** is Associate Professor of Philosophy at Pasadena City College in Pasadena, California. His current research interests are in metaphysics and philosophy of nature,

natural theology, and philosophical psychology. His books include *Philosophy of Mind* (Oneworld, 2013), *Aquinas* (Bolinda Audio, 2012), *Scholastic Metaphysics: A Contemporary Introduction* (Editiones Scholasticae, 2014), and *Neo-Scholastic Essays* (St. Augustine's Press, 2015).

**Stewart Goetz** is the Ross Frederick Wicks Distinguished Professor in philosophy and religion at Ursinus College. He has written extensively on the freedom of choice and the nature of the self. His books include *Naturalism* (co-authored with Charles Taliaferro, Eerdmans, 2008), *Freedom, Teleology, and Evil* (Continuum, 2008), *The Soul Hypothesis* (co-edited with Mark Baker, Continuum, 2011), *A Brief History of the Soul* (co-authored with Charles Taliaferro, Wiley Blackwell, 2011), *The Purpose of Life: A Theistic Perspective* (Continuum, 2012), *The Routledge Companion to Theism* (coedited with Charles Taliaferro and Victoria Harrison, Routledge, 2012), *A Philosophical Walking Tour with C. S. Lewis: Why It Did Not Include Rome* (Bloomsbury, 2014), and *God and Meaning* (co-edited with Joshua Seachris; Bloomsbury, 2016). Goetz is the senior editor of the Bloomsbury book series *Bloomsbury Studies in Philosophy and Religion* and co-senior editor of the forthcoming *Wiley Blackwell Encyclopedia of Philosophy of Religion*.

**Joel B. Green** (PhD, Aberdeen) is Provost and Professor of New Testament interpretation at Fuller Theological Seminary. He has written or edited more than forty-five books, including *Body, Soul, and Human Life: The Nature of Humanity in the Bible* (Baker Academic, 2008) and *What About the Soul? Neuroscience and Christian Anthropology* (Abingdon Press, 2004). He is the editor of the *New International Commentary on the New Testament* series (Eerdmans), editor-in-chief of the *Journal of Theological Interpretation*, and serves on the editorial boards of *Theology and Science* and *Science and Christian Belief*.

**Gary R. Habermas** (PhD, Michigan State University) is the distinguished research professor and chair of the Philosophy Department at Liberty University. He has authored or edited over forty books (more than twenty of these on Jesus's resurrection). Besides this chief area of research, he has published many other items on near-death experiences, the historical Jesus, religious doubt, and personal suffering. He has also published more than seventy chapters or articles in other books, plus approximately 150 articles in journals and periodicals. He has been a visiting or adjunct professor at 15 different graduate schools and seminaries (teaching about fifty courses) in the United States and abroad. See his website at [www.garyhabermas.com](http://www.garyhabermas.com).

**William Hasker** (PhD, Edinburgh) is Professor Emeritus of Philosophy at Huntington University, where he taught from 1966 until 2000. He was the editor of *Christian Scholar's Review* from 1985 to 1994, and the editor of *Faith and Philosophy* from 2000 until 2007. He has contributed numerous articles to journals and reference works, and is the author of *Metaphysics* (InterVarsity Press, 1983), *God, Time, and Knowledge* (Cornell University Press, 1989), *The Emergent Self* (Cornell University Press, 1999), *Providence, Evil, and the Openness of God* (Routledge, 2004), *The Triumph of God Over Evil* (InterVarsity Press, 2008), and *Metaphysics and the Tri-Personal God* (Oxford University Press, 2013).

**Ross Inman** is Assistant Professor of Philosophy at Southwestern Baptist Theological Seminary in Fort Worth, Texas. His research to date has focused primarily on metaphysics

and philosophy of religion and has appeared in *Philosophical Studies*, *Oxford Studies in Philosophy of Religion*, *Metaphysica*, and *Philosophia Christi*. He is a former Templeton research fellow at both the University of Notre Dame's Center for Philosophy of Religion and Saint Louis University. He holds an MA in philosophy, an MA in Theology from Talbot School of Theology (Biola University), and a PhD in philosophy from Trinity College, Dublin.

**Jaegwon Kim** is the William Perry Faunce Professor of Philosophy *Emeritus* at Brown University, where he taught for over 30 years. He also taught at Swarthmore College, Cornell University, and, for many years, the University of Michigan. Kim was educated at Seoul National University, Dartmouth College, and Princeton University. His books include *Supervenience and Mind* (1993), *Philosophy of Mind* (1996), *Mind in a Physical World* (1998), *Physicalism, or Something Near Enough* (2005), and *Essays in the Metaphysics of Mind* (2010).

**Robert C. Koons** is a professor of philosophy at the University of Texas at Austin. He specializes in metaphysics, philosophy of religion, and philosophical logic. He is the author, with Timothy Pickavance, of *The Atlas of Reality* (Wiley Blackwell, 2016) and *Metaphysics: The Fundamentals* (Wiley Blackwell, 2014). He coedited *The Waning of Materialism* (Oxford University Press, 2010) with George Bealer, and he is the author of *Realism Regained* (Oxford University Press, 2000) and *Paradoxes of Belief and Strategic Rationality* (Cambridge University Press, 1992). Koons's most recent work is on neo-Aristotelian metaphysics, especially as a basis for a metaphysics of modern science.

**Jonathan J. Loose** is a senior lecturer in philosophy and psychology and director of learning and teaching at Heythrop College, University of London. He has research interests in philosophy of mind, metaphysics, and philosophy of religion arising from prior work in cognitive science and psychology. Recent work focuses on contemporary dualist and physicalist views of human persons in relation to theological and practical concerns and he is currently working on a monograph exploring these issues.

**E. J. Lowe** (1950–2014), Durham University, United Kingdom, authored, among other books, *Kinds of Being* (Blackwell, 1989), *Subjects of Experience* (Cambridge University Press, 1996), *The Possibility of Metaphysics* (Oxford University Press, 1998), *The Four-Category Ontology* (Oxford University Press, 2006), *Personal Agency* (Oxford University Press, 2008), and *More Kinds of Being* (Blackwell, 2009).

**William G. Lycan** is William Rand Kenan Jr. Professor of Philosophy *Emeritus* at the University of North Carolina, and currently Distinguished Visiting Professor at the University of Connecticut. He is the author of *Logical Form in Natural Language* (1984, MIT Press), *Knowing Who* (with Steven Boër, 1986, MIT Press), *Consciousness* (1987, MIT Press), *Judgement and Justification* (Cambridge University Press, 1988), *Modality and Meaning* (Kluwer Academic, 1994), *Consciousness and Experience* (MIT Press, 1996), *Philosophy of Language: A Contemporary Introduction* (Routledge, 1999), and *Real Conditionals* (Oxford University Press, 2001).

**Michael N. Marsh**'s career was in academic clinical and biomedical research (Department of Medicine, Manchester, and professor of intestinal immunopathology in Ankara), his

major interests being the host's graded intestinal mucosal responses to environmental antigenic challenge, particularly with celiac disease, tropical infectious and parasitic diarrheas, and host–host transplantation histo-incompatibility, resulting in the international “Marsh Classification.” He published three books and over two hundred papers, holding visiting professorships in America (2), New Zealand (Otago) and Australia (Adelaide). On approaching retirement, he read the Oxford theology degree, subsequently writing his DPhil thesis on the neuropathology and theology of near-death and out-of-body experiences (Oxford Theology Monographs Series, 2010). He currently writes on bioethical issues arising from developments in medical science and clinical practice. Now at Wolfson College, he was formerly a fellow at The Oxford Centre for Christianity & Culture, Regent's Park College, University of Oxford.

**Angus J. L. Menuge** is a professor and Chair of Philosophy at Concordia University Wisconsin and president of the Evangelical Philosophical Society. Menuge is author of many journal articles and book chapters on the philosophy of mind and has broad interests in philosophical and theological anthropology. He is editor of several books, including *Reading God's World* (Concordia Academic Press, 2004) and *Legitimizing Human Rights: Secular and Religious Perspectives* (Ashgate, 2013), and author of *Agents Under Fire: Materialism and the Rationality of Science* (Rowman and Littlefield, 2004). His most recent edited book is *Religious Liberty and the Law* (Routledge, 2017).

**Trenton Merricks** is Commonwealth Professor of Philosophy at the University of Virginia. He is the author of *Objects and Persons* (Oxford University Press, 2001), *Truth and Ontology* (Oxford University Press, 2007), and *Propositions* (Oxford University Press, 2015). He has also written many articles in metaphysics, philosophy of religion, epistemology, philosophy of language, and philosophy of mind.

**J. P. Moreland** is Distinguished Professor of Philosophy at Talbot School of Theology, Biola University. He has authored, edited, or contributed papers to ninety-five books, including *Does God Exist?* (Prometheus), *Universals* (McGill-Queen's), *Consciousness and the Existence of God* (Routledge), *Blackwell Companion to Natural Theology* (Blackwell), and *Debating Christian Theism* (Oxford). He has also published over eighty-five articles in journals such as *Philosophy and Phenomenological Research*, *American Philosophical Quarterly*, *Australasian Journal of Philosophy*, *MetaPhilosophy*, *Philosophia Christi*, *Religious Studies*, and *Faith and Philosophy*. In August 2016 he was named by *The Best Schools* as one of the 50 most influential living philosophers.

**Nancey Claire Murphy** is Senior Professor of Christian Philosophy at Fuller Seminary, Pasadena. She received her PhD. from University of California, Berkeley (philosophy of science) and the ThD from the Graduate Theological Union. Her first book, *Theology in the Age of Scientific Reasoning*, won the American Academy of Religion award for excellence. She is author of eight other books and coeditor of twelve. Her most recent is (with Warren Brown) *Did My Neurons Make Me Do It? Philosophical and Neurobiological Perspectives on Moral Responsibility and Free Will*. Her research focuses on the role of modern and postmodern philosophy in shaping Christian theology, on relations between theology and science, and on neuroscience and philosophy of mind.

**Timothy O'Connor** is Professor of Philosophy at Indiana University. He has held visiting research fellowships at the Universities of Notre Dame, St. Andrews, and Oxford and has given 160 academic lectures in 22 countries. His main areas of scholarship are metaphysics, philosophy of mind, and philosophy of religion. He has published over seventy scholarly articles, edited seven books, and written two monographs, *Persons and Causes: The Metaphysics of Free Will* (Oxford University Press, 2000) and *Theism and Ultimate Explanation* (Blackwell, 2008). He is now writing a third book for a broader audience, *Thinking About Faith: Philosophy, Science, and Christian Belief*.

**Eric T. Olson** is Professor of Philosophy at the University of Sheffield. He is the author of *The Human Animal* (Oxford University Press, 1997) and *What Are We?* (Oxford University Press, 2007), and of more than fifty articles on metaphysics, philosophy of mind, philosophy of religion, and ethics.

**Joshua Rasmussen** is Assistant Professor of Philosophy at Azusa Pacific University. He works on topics in metaphysics, with a focus on basic categories, minds, and necessary existence. He is also interested in the intersection between these topics and natural theology. He is author of *Defending the Correspondence Theory of Truth* (Cambridge University Press, 2014) and coauthor of *Necessary Existence* (with Pruss; Oxford University Press, 2017). He has published more than two dozen articles on these and related topics. Rasmussen values collaboration across disciplines and perspectives.

**Ian Ravenscroft** is Associate Professor of Philosophy at Flinders University, South Australia. He works in the philosophy of mind and cognitive science, with occasional sallies into applied social theory. Amongst other publication, he is author of *Philosophy of Mind: A Beginner's Guide* (Oxford University Press, 2005). Strongly influenced by Frank Jackson and the late J. J. C. Smart, it is not surprising that he is an atheist, a physicalist, and a consequentialist. He lives in the Adelaide Hills with his wife and children to whom he is greatly devoted.

**Brandon L. Rickabaugh** is currently a doctoral student in the Department of Philosophy at Baylor University. He has earned degrees in philosophy from University of California, Irvine and Biola University. Rickabaugh's research focuses on the intersection of epistemology and metaphysics regarding issues in the philosophy of mind (consciousness, intentionality, the mind/body problem), the ontology of human persons, and philosophy of religion. His work has been published in a number of academic journals including the *International Journal for Philosophy of Religion*, *Heythrop Journal*, and *Philosophia Christi*.

**Richard Swinburne** is a fellow of the British Academy. He was Professor of the Philosophy of Religion at the University of Oxford from 1985 until his retirement in 2002. Since then he has lectured or taught for extended periods in various countries. He is the author of many books on different philosophical topics, especially books on the nature and existence of God. His most recent new book, *Mind, Brain, and Free Will* (2013), advocates substance dualism and libertarian free will. A largely rewritten second edition of his *The Coherence of Theism* was published in 2016.

**Charles Taliaferro** is Professor of Philosophy and Chair of the Department of Philosophy, St. Olaf College. He is the author, coauthor or editor of over twenty books, including *Consciousness and the Mind of God*, *Evidence and Faith*, *Philosophy and Religion since the 17th Century*, *The Image in Mind* with Jil Evans; and *A Brief History of the Soul* with Stewart Goetz. He is the philosophy of religion editor for *Religious Studies Review*.

**Luke van Horn** is an instructor in philosophy for Ashford University. He has authored or coauthored articles in the journals *Faith and Philosophy*, *Philosophical Studies*, and *Journal of the Evangelical Theological Society*.

**Peter van Inwagen** is the John Cardinal O'Hara Professor of Philosophy at the University of Notre Dame. In addition to numerous articles, some of his most prominent books include *Thinking about Free Will* (Cambridge University Press, 2017), *Existence: Essays in Ontology* (Cambridge University Press, 2014), *The Problem of Evil* (Oxford University Press, 2006), and *Ontology, Identity, and Modality* (Cambridge University Press, 2001).

# Introduction

## *Substance Dualism and Its Physicalist Rivals*

JONATHAN J. LOOSE, ANGUS J. L. MENUGE, AND  
J. P. MORELAND

To say the least, substance dualism has not enjoyed good public relations within academic philosophy, or for that matter, within related disciplines, such as psychology, biology, or neuroscience. So it is natural that some readers will want to know how, and even why, this volume came about. In this introduction – and more fully in the book itself – we hope to show that due to recent developments within the philosophy of mind, a renewed interest in historical and contemporary theories of the soul, and a more careful evaluation of what does and does not follow from neuroscience, substance dualism is back on the table for a serious critical reevaluation.

At the outset, it is important to be clear that, unless otherwise indicated by an individual author, this volume will understand “substance dualism” in a very broad sense that is by no means exhausted by the Cartesian variety. By “substance dualism” we mean the generic view that (1) there is a substantial self, soul, or ego that is immaterial and (2) that self, soul, or ego is not identical to the body and is the bearer of personal identity. Given the variety of theories about what constitutes a substance (or substance-like entity), substance dualism thus defined is compatible not only with Cartesian dualism but also with a number of non-Cartesian alternatives, including several varieties of Thomistic (or neo-Thomistic) dualism, Hasker’s emergent subject dualism, and the holistic anthropology of E. J. Lowe.

We will see that substance dualists and their many critics have been brought together by a shared focus on the nature of mental subjects. And, as much is at stake, including the tenability of the reigning doctrine of naturalism, it is not surprising that the debate is intense. From the beginning, we therefore felt that the only fair way to present this new development – *the return of the subject* to the center stage of philosophy of mind – is to construct a level playing ground of debate for all of the various positions and their critics, in hopes that readers can decide for themselves where the better arguments lie.

We will begin with a brief explanation of why this book is timely (Section 1.1), then review in more detail recent developments in the philosophy of mind (Section 1.2) and in scholarship on the soul (Section 1.3). We conclude by considering the broad implications of the return of the subject for the larger question of the tenability of naturalism (Section 1.4) and give a brief outline of the structure of the book (Section 1.5), followed by summaries of each chapter (Section 1.6).

## 1.1 An Inconceivable Book?

### 1.1.1 *The official doctrine*

Go back a few decades and the idea of a wide-ranging scholarly examination of the merits of substance dualism would have seemed outlandish. Dennett captured the mood at the time when he wrote, “it is widely granted these days that dualism is not a serious view to contend with, but rather a cliff over which to push one’s opponents” (Dennett 1978, 252). While in some cases incredulity about substance dualism has resulted from sophisticated but ultimately resolvable difficulties such as those raised by Wittgenstein’s discussion of private language (2009 [1953]), most professional philosophers are simply inoculated against any version of substance dualism by a seemingly unanswerable objection firmly impressed on their minds during their very first class. Descartes argued that the soul and body are substances of fundamentally different kinds, the one an immaterial, indivisible thinking thing with no spatial extension, the other a material, divisible entity that necessarily occupies space. A standard rhetorical question follows: how can substances of such fundamentally different kinds possibly causally interact, as Descartes maintains that they do? It seems inconceivable that items not sharing a common medium (space) could influence one another, and Descartes’s well-known replies to Princess Elisabeth of Bohemia’s pointed questions on this issue look like the hand-waving of an *ancien régime*, about to be swept aside by a scientific outlook that has no room for the soul.

In his day, Gilbert Ryle (1949) complained that an essentially Cartesian view of the mind was still part of the “official doctrine” about the nature of consciousness. But for most twentieth-century philosophers of mind it was the perceived failure of substance dualism in general that deserves that title. It was assumed that Descartes’s version of substance dualism had been fully understood, found irredeemably flawed, and that other versions, if they were considered at all, were subject to the same fatal defect. The Cartesian vignette that has for decades adorned almost every introductory class in philosophy is one of many reasons that twentieth-century philosophy of mind was dominated by research programs that ignored an understanding of the conscious subject (apparently too much like a Cartesian ego), but instead focused on scientifically tractable aspects of cognition, such as the explanation of behavior and the relation between mental and physical states and events. While behaviorism soon fell, philosophy of mind embraced a physicalist research program, according to which mental states either are, or are entirely determined by, physical states of the organism. For many still today, the triumph of physicalism, as the attempt to integrate human beings into a consistent, scientifically grounded picture of the world, is so complete that the soul must be dismissed, along with epicycles, the humors, and phlogiston, as outmoded and redundant.



### 1.1.2 *Fault lines in physicalism*

How then, could substance dualism have earned the right to a serious, well-rounded, critical examination? To some, this will still appear as unmotivated as a contemporary reevaluation of alchemy. But the truth is far different from the simplistic narrative of Cartesian failure and physicalist triumph. One problem is that, while physicalism has generated an extraordinary variety of theories of the mind, they generally have serious, if not fatal, problems. Not only that, but there is also a recurring pattern of failure, that suggests there is something wrong, not with the specific details of a given account, but with the whole approach. In one way or another, these theories fail to capture basic aspects of the mind, such as phenomenal consciousness, intentionality, and even rational thought. They do not seem to capture accurately what it is like to feel pain, the fact that my thought can be about something beyond itself (including future, fictional, and even necessarily nonexistent entities that cannot causally explain the thought), or the fact that my thought can access and be governed by noncontingent norms of logic.

As a result, there has been a move toward theories of mind that embrace some version of emergentism, a nonreductive version of physicalism which allows that novel mental qualities and powers may emerge from the right physical base. But now some difficult questions arise. How far can emergence go before it abandons core doctrines of physicalism? At what point does emergence become a form of dualism under a different name, if it effectively concedes most of what dualists have maintained about the distinctive characteristics of the mind? That these are serious questions is shown by the fact that there are emergent subject dualists (Hasker 1999, and this volume), as well as emergent physicalists. We will explore the move from standard physicalism to emergentism in more depth in Section 1.2.

Another problem with our opening narrative is that scientifically minded modern people, including analytic philosophers, have often spent very little time investigating the soul. Perhaps a majority are unaware that there are many, quite different views of the nature and function of the soul. And even in Descartes's case, reliance on a brief caricature may have obscured a more accurate and fair understanding of his theory of the soul. In recent years there has been an explosion of research on the soul, mining the historical sources, adopting some of their insights, but also proposing constructive modifications to handle well-known problems and objections. We will survey some of this thinking in Section 1.3.

### 1.1.3 *The return of the subject*

As it happens, these two threads – the fault lines in physicalism and the reconsideration of the soul – draw together in a fascinating sea change in the philosophy of mind. While standard versions of physicalism were largely atomistic, focusing on particular mental states and events, there is an increasing recognition that philosophy of mind must address the nature of mental subjects. One of the most puzzling things about conscious mental states is that they are intrinsically subjective and, of course, subjectivity requires a subject. For many, it is strongly inconceivable that thoughts and experiences be ownerless: there cannot be an experience of a sunset that is no one's experience, or a thought that the sunset is beautiful that is no one's thought. But if that is right, then thoughts and experiences cannot be understood as independent atoms: their nature and existence depends on a unified whole to which they necessarily belong. And, embarrassingly enough, the person who insisted on this point – that thoughts are not detachable from thinkers – was our friend Descartes.

A mental subject, it seems, is a basic precondition of thought, just as Descartes said, so that Ryle's behaviorism was guilty (quite literally) of changing the subject, by refusing to speak of something essential to our mental lives. Attention to this fact has led to a fascinating development in the philosophy of the mind in recent years, what we have called *the return of the subject*. Even those resolutely opposed to the Cartesian paradigm increasingly feel compelled to offer some account of the origin and nature of mental subjects. Hard questions again arise. Can the subject be accommodated without allowing Descartes an unwelcome revenge on his many critics? Jaegwon Kim (1998, 46) had already noted that the same problem of causal interaction raised for Descartes at the level of mental and physical substances reappears for non-Cartesian property dualists at the level of mental and physical properties. But if everyone (beyond those eliminativists who implausibly deny conscious phenomena altogether) must give an account of mental subjects, it is not obvious that an appeal to "emergence" will save these accounts from facing a question uncomfortably like the one posed to Descartes.

This is by no means a counsel of despair, partly because several philosophers have pointed out that there is something wrong with the question. Hume taught us that there is no logical connection between causes and effects, that causes do not even have to be like their effects, and that we often have very good reason to think that two kinds of events are causally related without knowing how. And Hume's point seems correct even if one does not embrace his view of causation.

If this is true in general about causation, then the fact that we lack a fully adequate account of how mind and body interact does nothing to discredit the overwhelming *prima facie* case that they do (Swinburne 2013). And critics of Descartes typically operate from an event-causal paradigm that fails to take seriously the idea of substance causation anyway. As Lowe (2008) and Swinburne (2013) have argued, if there are substances with basic causal powers, it is much less obvious why a mental substance could not have the power to influence (and be influenced by) the physical world.

In any event, there is no doubt that contemporary philosophy of mind has seen a major shift toward an attempt to understand the mental subject, with book-length studies of the unity of consciousness, the self, and the first-person perspective (Searle 2001, 2007; Tye 2003; Bayne 2012; Baker 2013).

Another of the hard questions is whether this return of the subject is really a good fit for naturalism as the dominant approach to philosophy. One might raise the question of whether theism is more plausible than naturalism as an explanation of the existence of mental subjects. We will return to this theme in Section 1.4.

As the subject has moved to center stage, another development has been a broader understanding of what may qualify as "substance dualism." In this sense, not all of today's substance dualists would accept Descartes's view that mental and physical substances are in principle independent of one another. Some take the view that while mental subjects have powers different from physical brains, the mental subjects are still ontologically dependent on those brains. Others, though, side with Descartes or Aquinas, and allow that the existence of disembodied souls is possible.

Whether this possibility is required to make sense of basic Christian doctrines, such as the resurrection, is a major issue that divides Christian dualists (Loose 2012; Moreland 2014; Van Horn 2010) and Christian materialists (Corcoran 2006; Murphy 2006; van Inwagen 2007). More generally, every version of "substance dualism" in the broad sense is also challenged by other philosophers, some more, some less sympathetic to standard physicalism. So this book

follows a debate-style format allowing a fair comparison of each position and its rivals (for a brief outline, see Section 1.5, and for summaries of the debates, see Section 1.6).

## 1.2 From Standard to Emergent Physicalism

It is not an exaggeration to say that the history of “standard” physicalism, that is, the largely reductive physicalism preceding the contemporary emphasis on emergence, has been one of persistent failure. Early attempts to understand mental attributions without postulating an occult ego suggested that, for example, to say someone was in pain was really to say something about certain characteristic body movements, such as wincing, crying out, withdrawing affected limbs, and so on. While behaviorism had several versions, a common complaint was that no alleged pain behavior was in all cases necessary and sufficient for pain (Putnam 1968). Spartans (or super-Spartans) may so train themselves as to emit no pain behavior despite excruciating agony, and method actors (or Italian soccer players) may be utterly convincing in the expression of pains they do not experience. This is not surprising since the most important part of pain is the *qualé* – what it is like to feel pain – not the behaviors it typically produces. The point was made, in hindsight rather obvious, that mental states are causes, and one cannot adequately define causes exclusively in terms of their effects.

If mental substances are assumed to be out of the question, but mental states are causes, the natural next suggestion is that mental states are in some sense identical with physical states (Smart 1959). For type identity theorists, there is an identity between being in pain (as a type of mental state) with some physical type of state (with the firing of “C-fibers” standing in for whatever scientists tell us is the real physical substrate for pain). This idea did not survive long because it was soon realized that there are marked neurophysiological differences between creatures that experience pain, making it unlikely that the substrate for pain is the same in all cases (Lewis 1980). Token identity theorists thus proposed the more modest thesis that each particular pain state was identical with some physical state of the brain, and the identity might also be species-relative. However, the token view failed to explain what it is in virtue of which instances of mental states qualify as tokens of the same type and thus only avoided the problem of multiple realizability of particular mental states by undermining their identity conditions and thus their very existence.

But even if token identity theory were not flawed, the identity theory faced other large and obvious objections. Physical states can be completely described in an impersonal way and do not seem to be about anything. But mental states are inherently personal, having a subjective, what-it-is-like character (Nagel 1974; Jackson 1982, 1986). And they typically are also about something. For example, we may have a thought about Paris or a pain “in the foot,” indicating that the thought and the pain have intentional content. The Leibnizian principle of the indiscernibility of identicals implies that mental states are not physical states because mental states have properties that physical states lack.

At about the same time philosophers of mind became dissatisfied with the identity theory, many of them fell in love with a computational model of the mind. In the 1970s and 1980s it came to seem almost self-evident that the human mind was much like a computer and a confluence of ideas from computer science, psychology, linguistics, and philosophy birthed the new paradigm of cognitive science. One of the attractions of the computational model is that computers can be understood at a variety of levels. It was thought that the main error of the identity theory was to view the brain at too low a level of abstraction. If,

instead of focusing on the physical hardware (or “wetware”), we move up to the level of its functional organization, we will get the high-level, abstract view of the brain appropriate for understanding the mind.

At first sight, this proposal, known as functionalism, seemed highly promising. Functionalism offered a simple solution to the main objection to type physicalism because, on the functionalist view, mental states are not identified with specific types of physical states, but with their functional roles. A common analogy was the mousetrap. The functional role of a mousetrap – a device that traps and kills mice – has multiple physical realizers, ranging from the standard 5-part trap to the most convoluted Rube Goldberg machines. In the same way, the functional role of a pain state – a state mediating impending or actual bodily damage and other appropriate states and behavior – may be realized differently in different species or even within the same species, due to developmental differences, brain damage, and so on. And the functional role is abstract, something different from ordinary physical characteristics, so perhaps this might also explain the failure of the token identity theory.

Unfortunately, it did not take long for skeptics to realize that functionalism was, in a way, a sophisticated, internalized version of behaviorism. Among many problems raised, the most decisive is that a functionalist theory of pain can be realized by a robot that does not feel pain and whose states are not about anything. Early on Ned Block (1978) pointed out that if all that matters for the mind is having the right functional roles, then if a crowd of billions of people emulate every functional role of the neurons in their brain, the crowd must have its own consciousness and mental states, over and above those of each member of the crowd. But we judge this to be obviously false.

A variant of the same problem arises in functionalist attempts to explain away qualia in terms of powers to discriminate. For example, perhaps all we mean by our ability to distinguish green qualia and red qualia is that we have the power to discriminate between two stimuli. But this is again obviously false since a robot can be equipped with a device that distinguishes the different wavelengths of light and which is programmed to behave differently in the two cases, but no one thinks the robot has a conscious experience of red or green, or knows what it is like to be appeared-to-redly or appeared-to-greenly. This problem of accounting for the subjectivity and intentionality of mental states continues to afflict the most sophisticated varieties of functionalism. In 1992, John Searle, himself a well-known critic of functionalism as a theory of artificial intelligence (Searle 1980, 1983) looked back at the dismal state of affairs.

The most striking feature is how much of mainstream philosophy of mind of the past fifty years seems obviously false. I believe there is no other area of contemporary analytic philosophy where so much is said that is so implausible . . . In the philosophy of mind, obvious facts about the mental, such as that we all really do have subjective conscious mental states and that they are not eliminable in favor of anything else, are routinely denied by many, perhaps most, of the advanced thinkers in the subject. (Searle 1992, 3)

The take-away message from the decline of standard physicalism is that we cannot substitute physical or functional states for mental states without losing the very characteristics (phenomenal subjectivity and intentionality) that make the states mental. Fifty years is a rather long time to spend exploring alternatives to this (in hindsight) inevitable conclusion.

Given these persistent difficulties, philosophers have looked for weaker relations between the physical and the mental. The core principle of physicalism is the principle of causal closure (PCC).

Pick any physical event . . . and trace its causal ancestry or posterity as far as you would like; the principle of causal closure says that this will never take you outside the physical domain. Thus, no causal chain involving a physical event ever crosses the boundary of the physical into the nonphysical: If  $x$  is a physical event and  $y$  is a cause or effect of  $x$ , then  $y$  too must be a physical event. (Kim 2011, 214)

But PCC is at least consistent with denying that mental properties (and states) are identical with physical properties (and states). Initially, the most popular proposal was that the mental depends on the physical by way of supervenience (Davidson 1980; Kim 1982, 1984). According to the basic idea of supervenience (developed further in weak or strong forms), there is no mental difference without a physical difference, and so if we fix the physical facts, we thereby fix the mental facts. Supervenience is not identity, since it is an asymmetric relation. Two molecule-by-molecule physical duplicates could not be in different mental states (at least conceived in terms of so-called “individualistic,” or “narrow content”). But it is possible that the same mental state could supervene on different physical base states in different individuals. Hence, supervenience allows for the idea that supervening states can be “multiply realized” by different subvenient states.

It may seem consistent to claim that mental states characterized by subjectivity and intentionality supervene on physical states of the brain, thereby avoiding the problems besetting standard physicalism. But even if that is true, it looks as if supervenience guarantees that the mental states are epiphenomenal, since, as Kim has argued, granted PCC, the physical base states seem to preempt any distinctive causal contribution from the supervenient mental states – the so-called exclusion problem (Kim 2011, 214–220). Among the many apparent absurdities of epiphenomenalism is that were it true, it seems that no one could come to know it or convince anyone else of its truth, since on the standard causal theories of knowledge accepted by most physicalists, both of these achievements presuppose psychophysical causation, that is, that epiphenomenalism is false. And Swinburne (2013) has pointed out that if epiphenomenalism were true, no scientific test could confirm it, since we can test whether an intentional state has effects only if we can identify when someone is in that state, and we can do this only if a subject’s statement that he is in that state is reliably caused by the state, that is, only if epiphenomenalism is false.

Even if Kim’s exclusion argument can be avoided as some maintain, a more general concern is that supervenience is a highly obscure notion. Supervenience asserts, rather than explains, a systematic correlation between mental and physical phenomena, and this does not come for free out of a metaphysical necessity, because the correlation appears to be contingent. There is no conceptual difficulty in the idea that our physical duplicates in other possible worlds are “zombies” with no conscious states or have inverted qualia (feeling pain when we feel pleasure, etc.). Even if, in our world, psycho-physical laws govern the correlations between physical base states and supervening mental states, those laws themselves are metaphysically contingent, and so we need an explanation of why the correlations obtain.

For these and other reasons, it has become more popular among those sympathetic with physicalism to suggest some version of emergentism. Emergentism typically maintains that there is a *causal* relation between physical base states and resulting mental states. So the idea

is that, in some sense, brains can generate mental states. However, not much is gained by the mere assertion of emergence. For one thing, if Kim is right, the exclusion problem can be reapplied to emergent physicalism (Kim 2006). This creates the dilemma that if the emergent physicalist maintains PCC, he must embrace epiphenomenalism, but if instead he asserts that the mind has novel and independent causal powers, he is rejecting PCC and can no longer claim to be a physicalist.

But even if this argument can be avoided – and again, the matter is disputed – the mere assertion of emergence does not really account for the subjectivity of mental states, and this for two reasons. First, there is simply nothing about the physical base states that predicts or even suggests the emergence of states characterized by subjectivity, and, in a dialectical context that includes dualist rivals, the physicalist cannot simply assert that this is a brute fact. It is a contingent fact with multiple, competing explanations, and the physicalist must therefore show that his explanation is the best. Second, and more fundamentally, subjectivity cannot arise unless a subject emerges, and so the real problem is not simply accounting for this or that mental state but for the mental subject to which all those states belong.

At this point, it has become clear, at least to those with some dualist sympathies, that standard physicalism does not supply the resources needed to account for the existence of mental subjects. This is because the physical brain is a complex aggregate of parts in external relations. But subjects seem to have a unity (they are what all of a certain set of thoughts have in common) and the relationship between thoughts and their subject seems to be internal, since it apparently makes no sense to speak of ownerless thoughts or of another person thinking the very same thought that I am thinking, even if she agrees with me, and has a qualitatively identical thought with the same intentional content and associated qualia. Similar problems arise when we consider a mental subject's persistence over time. In a simple action of confirmation, for example, confirming my hypothesis that I left the window open, it seems that the subject that discovers that the window is open at time  $t$  must be the same as (i.e., identical to) the subject that entertained the hypothesis that it was open at time  $t-k$ . Otherwise it would be as if I thought the window was open, then ceased to exist, and another person, Jack, discovered that the window was open. I do not survive long enough to confirm the hypothesis, and Jack cannot confirm a hypothesis he never entertained.

If we are to take seriously the idea of a subject that owns thoughts at and over time, we must therefore have an illuminating account of mental subjects. Physicalists who have realized this have therefore put considerable effort into developing more sophisticated accounts of the relationship between the brain and the mind, in the hope that they can make more plausible the emergence of a first-person perspective from impersonal gray matter.

Several theories have emerged in recent years which take psychological subjects seriously but attempt to locate them in a broadly physical context. Animalists like van Inwagen (2007) and Olson (2007) maintain that persons are human animals, and so for them the unity and identity of mental subjects are rooted in the unity and identity of particular biological organisms. Constitutionists like Baker (2001, 2013) and Corcoran (2006) disagree, arguing that persons and living human bodies have different persistence conditions. Being a person requires having intentional states and a first-person perspective which some living human beings lack, so while persons depend on living human bodies, a human can persist when a person does not. As a statue may be constituted by, yet not identical with, a piece of marble, so a person is constituted by, but not identical with, a living human body. Others, like Nancey Murphy (2006), pursue the basic idea of nonreductive materialism (emergent monism) and try to explain in more detail under what conditions mental subjects

can be expected to emerge. Likewise Tim O'Connor, while a theist, holds out for a view of human persons on which unified mental subjects emerge under the right kind of physical conditions (O'Connor and Jacobs 2003; O'Connor and Wong 2005).

### 1.3 The Soul Reconsidered

As the self has moved to center stage some have wondered if alternatives to Descartes's view of the soul might be worth a second look (see Goetz and Taliaferro 2011 for a historical survey, Baker and Goetz 2011 for a wide-ranging discussion of how the soul might be integrated with modern science, and Moreland 2009 and 2014 for a recent defense of the soul). The idea of a soul is often ridiculed on the grounds that the soul is not located in space, and therefore, even if it has causal power, there is no reasonable explanation of why it directly causally interacts with just one body. Why, when I want to raise my arm does my arm go up, and not Jeff's across the street? Why, when I stub my toe, do I feel pain, but not Jeff? This is one example of the pairing problem frequently posed for substance dualism (Kim 2011, 50–54). But it is unclear that Cartesians have no response: perhaps souls have a primitive particularity or "thisness" that explains their particular causal powers as Swinburne suggests (this volume). And in any case dualists are not limited to the Cartesian option. For one thing, Augustine and Kant both maintained, *contra* Descartes, that the soul is located in space, but that the mode of presence was different than that of a physical object. While physical objects exclude all others from the same space, and only part of a physical object is located in part of that space, it is argued that all of the soul is present in any part of the living body, so that a pain whose physical source is localized in part of the body – say the toe – is nonetheless experienced by the whole soul. It is not so obvious why such a soul could not causally interact with its body.

But some dualists have followed Aristotle and Aquinas, and asked whether the soul should really be conceived of as a mental substance in the terms of Augustine or Descartes. Perhaps the soul is the "substantial form" of a living human person, that which grounds a person's existence as a rational animal of a certain kind. Like the animalist view, the unity and persistence of the person is rooted in a fact about the living organism. But unlike animalism, the Thomist view is open to the possibility that the soul, understood as a substantial form, can survive the death of the body by continuing to subsist.

As these historic alternatives to Cartesian dualism are given renewed attention, some scholars have also developed revised versions. Alongside modified Cartesianism (Swinburne 2013), there are neo- or quasi-Thomistic positions (Moreland 2014), and Goetz and Taliaferro (2008, 64–69) also consider the merits of the Augustinian/scholastic/Kantian view of the soul. Doubtless there are other possibilities, but even if the quick dismissal of Descartes were justified, the important point is that just as physicalists are now exploring a wide variety of alternatives to standard physicalism, dualists have an equally rich reservoir of theories of the soul on which to draw.

### 1.4 The Return of the Subject: Broad Implications for Naturalism

The three developments just described – the decline of standard physicalism, the rise of emergent physicalism, and renewed interest in historic treatments of the soul – can, broadly speaking, be taken to point in the same direction. That is, they all suggest that we cannot

hope to understand our mental life without giving a coherent account of the mental subject. In many ways, this is an exciting and constructive development, because it means that the thinking of dualists and nondualists increasingly overlaps and there should be many opportunities for each perspective to learn from the other.

At the same time, this development also sharpens some of the hard questions mentioned earlier, and leads us to more momentous questions about the status of naturalism in philosophy. On the one hand, it is possible that the less restrictive versions of physicalism now in vogue will begin to solve the problems that beset standard physicalism, and this will strengthen the case for naturalism. But on the other hand, there is a serious risk that these new versions of physicalism, by accepting and accommodating more facts about mental subjects, become indistinguishable from dualism, and also not a good fit for naturalism.

As J. P. Moreland (2008, 2009) has argued, the more that mental characteristics – qualia, the unity of consciousness, intentionality, and rationality – are recognized as *sui generis* emergents, the less acceptable it becomes for naturalism to pose as the sole plausible explanation. If reductive physicalism had been successful, it would be obvious that our mental lives are an unproblematic part of the natural world. But since reductive physicalism has failed, and, according to nonreductive emergent schemes, there is simply a contingent causal relation between physical states of the brain and mental phenomena, emergent physicalism appears to assert a large number of psychophysical correlations as brute facts. And since they are brute facts, the emergent physicalist cannot claim that our mental lives are such as to be expected, given what we know about the physical states of our brains (viewed both synchronically and diachronically). Thus, if theism, as a major competitor to naturalism, can offer a plausible explanation of these correlations, but emergent physicalism fails to do so, theism is, on this score at least, more plausible than naturalism. It is arguable that if reality has always included a conscious mind with intrinsic subjectivity, intentional states, and goal-directed rationality, namely, God, the existence of finite beings with similar minds is much more to be expected than on the basis of the interactions of unconscious, nonintentional, nonrational matter, governed by purely external impersonal relations and undirected causal processes.

One philosopher who is aware of this dialectical situation is Thomas Nagel (2001, 2012). The situation does not merely concern consciousness, but also arises for objective reason and the nature of moral obligations. To the extent that naturalists concede that consciousness, objective reason, and moral obligations cannot be reduced to the world as described by natural science, and therefore must simply emerge, it seems that naturalists will continue to multiply the number of brute facts and correlations, thus creating an opportunity for theism to demonstrate its superiority over naturalism if it can provide plausible explanations of these facts and relations.

Nagel himself has attempted to avoid this (for him) unwelcome outcome by adopting a dismissive strategy. For example, he points out that we cannot find a higher standard outside of objective reason by which to justify it. For if we attempt to explain why our reason has necessary and universal validity, our explanation will itself assume such reason and so cannot provide independent warrant for accepting our reason as objectively valid. The same strategy could be adapted to dismiss the demand for an explanation of moral obligations: any attempt to justify the moral “point of view” will appeal to moral principles that assume that point of view. And, it might be argued that if human consciousness is mysterious, it does not help to diminish the mystery simply to invoke the existence of another being (God) who is conscious.



However, it is arguable that this dismissive strategy does not accurately reflect the dialectical situation. Certainly one cannot justify logic by logic, moral norms by moral norms, or consciousness by consciousness. But the real issue concerns the contingent exemplification of these characteristics. As a matter of contingent fact, there exist finite creatures that can access universal principles of logic and morality and that are conscious. There are plenty of possible worlds in which no such creatures exist. So now the question looms as to whether naturalism or theism is the better explanation of the existence of a world that does contain such creatures. The existence of a rational, moral, conscious being who wishes to create finite beings in His own image then offers an explanation of the existence of such beings that is arguably more plausible than the naturalist's assertion that such creatures simply emerged.

It is plausible that as emergent physicalism draws closer to dualism in its concern to account for mental subjects, one can expect a strenuous debate precisely because many realize (or have sensed) that the credentials of naturalism (and theism) are at stake. If more sophisticated versions of physicalism prevail, naturalism can claim another advance. But if physicalism simply absorbs without explaining all of the data that theists have used to advance their cause, then naturalism will increasingly look ad hoc – a degenerating research program – and may begin to recede (Koons and Bealer 2010).

## **1.5 Structure of the Book**

For the reasons given above, we think that a comprehensive study of the varieties of substance dualism (broadly conceived) is an exciting, timely topic. It gives an opportunity to hear out the full range of options open to substance dualists and alternative views, side by side with their best critics. With a few exceptions, this book generally uses a debate-style structure, pitting representatives of a given view against their critics. Thus, we first present the case for and against: emergent dualism, classical Thomistic dualism, neo-Thomistic dualism, and Cartesian dualism, together with debates on the evidential implications of the unity of consciousness and near-death experiences. Then we present the case for and against various alternatives to substance dualism, including: animalism, nonreductive physicalism, constitutionalism, and emergent individualism. Finally, and because many are interested in the related theological questions, we conclude with a section of debates on the best understanding of biblical anthropology in general, and of the Incarnation and Resurrection in particular.

We hope this book will be a valuable resource for scholars in a variety of disciplines (notably, philosophy of mind, psychology, and theological anthropology) and that it will be a useful reference for those interested in doing further work advancing the case for or against substance dualism.

## **1.6 The Chapters in Brief**

Closing out the introductory section (Chapter 2), William Lycan, a committed materialist, helps to motivate the book by holding his own feet to the fire and admitting that standard arguments typically relied on to dismiss (Cartesian) substance dualism (SD) are not convincing enough either to set it apart from other philosophical theories or to put it at a significant disadvantage in comparison to materialism. An initial exploration of materialism finds no good arguments in its favor and discussion of nine standard

arguments against SD finds each wanting. Property dualism (PD) fails to avoid the majority of these arguments and it suffers from two further problems of its own. Thus PD is neither significantly less problematic than SD nor an acceptable and “less crazy” way to be a dualist. Lycan remains a convinced materialist but states that in doing so he does not proportion his belief to the evidence.

### *1.6.1 Articulating substance dualism*

In the lead article for the next section (Chapter 3), Charles Taliaferro provides a general defense of substance dualism, conceived as the thesis that persons consist of *at least* two kinds of things. He articulates an integrative version of substance dualism that affirms the value of embodiment and argues that, contrary to the assumption of many physicalists, we have a clearer understanding of mental than of purely physical causation. Taliaferro further suggests that physicalist denials of irreducible subjectivity are incoherent as subjectivity is the noneliminable basis of the scientific model of explanation that physicalism privileges, and he develops a modal argument to show that persons are distinct from their bodies.

At the end of the section (Chapter 16), Ian Ravenscroft provides an opposing chapter, describing what he sees as the strongest general case against substance dualism. He begins by setting out some of the commitments of both substance dualism and its physicalist rivals before turning to a range of arguments in favor of substance dualism, all of which are found wanting. The argument from emergentism to dualism is also rejected. Three arguments against substance dualism and for physicalism about the mental are then explored. One argument against substance dualism is found wanting, but two powerful pro-physicalist arguments are explicated and defended. The latter arguments assess the relative explanatory power of substance dualism and physicalism and concludes that physicalism has vastly more explanatory power than substance dualism and is therefore on current evidence by far the preferred option. The design argument for theism forms a useful analogy for the argument from explanatory power against substance dualism, and is briefly explored.

In between these bookends, we hear the case for and against specific versions of substance dualism (in the broad sense).

#### **1.6.1.1 Debating emergent dualism**

William Hasker defends emergent dualism (Chapter 4). He makes the case that the major alternatives for understanding the nature and origin of the human mind are reductionism, creationism, and emergentism. Difficulties are pointed out for both reductionism and creationism, and it is argued that emergentism is the best of the three alternatives. It is then argued that emergent dualism is the most viable form of emergentism from both a philosophical and a theological standpoint.

In response (Chapter 5), Brandon Rickabaugh makes the case that Hasker’s emergent version of substance dualism has no advantage over nonemergent versions. After a careful exposition of Hasker’s views, Rickabaugh develops four main objections. Emergent substance dualism (1) lacks explanatory power, (2) predicts multiple conscious subjects, and implausibly suggests both (3) that there is a specific number of atomic simples in the brain requisite for a soul to emerge, and (4) that the combination of separable physical parts of the brain in external relations accounts for the unity of consciousness with mental faculties internally related to that consciousness as inseparable parts.

### 1.6.1.2 Debating Thomistic dualism

Edward Feser presents the case for Thomistic dualism (Chapter 6). Feser explains Aquinas's understanding of the human being as a single, true substance of the kind "rational animal." An explanation of substance precedes discussion of the complex nature of a rational animal. Feser defends Aquinas's view that human, animal, and vegetative kinds of living thing are irreducibly different and notes that human capacities of intellection and will are immaterial. Given this, human death is "full body amputation," leaving intact a nonfunctioning intellect and will, thus rendering persistence less than mysterious. Talk of the substantial soul existing by itself is seen to be a loose way of describing the truncated human after death. Aquinas's view faithfully reflects our "weird" and complex human nature, thus avoiding problematic consequences such as imbalanced attitudes to sex that follow from alternatives.

In Chapter 7, J. P. Moreland presents a modified, Thomistic-like form of dualism, otherwise known as "Organicism," and argues that it has certain philosophical and scientific advantages over physicalist treatments of the human person, and, to a lesser degree, advantages over alternate versions of substance dualism. Moreland then responds to a set of objections against his position.

William Hasker then offers a critique of both Thomistic and Thomistic-like dualism (Chapter 8). For Hasker, Aquinas's view is attractive because it takes humans to be deeply integrated with their bodies while being both more than mere animals and capable of postmortem existence. However, he believes the view fails to integrate human beings with nature, to give human souls a sufficient role, to justify their existence adequately within the larger system, and to remain consistent with evolutionary theory. While Moreland's modified Thomism addresses some of these points it remains inconsistent with evolutionary theory and must turn to vitalism. Hasker's alternative way forward is to propose that human and nonhuman animals have souls of the same kind, to abandon essentialism about species and to adopt emergentism.

### 1.6.1.3 Debating Cartesian dualism

Richard Swinburne (Chapter 9) presents a sophisticated case for Cartesian dualism. Swinburne notes that, as he stated it, Descartes's argument for the possibility of disembodied persons fails because it conflates apparent logical possibility with metaphysical possibility. Although there is no obvious logical contradiction in "I am thinking without a body," disembodied existence could still be metaphysically impossible because "I" refers to my body. Swinburne sets out to repair Descartes's argument by appealing to a distinction between informative designators (like "H<sub>2</sub>O") and uninformative designators (like "water"), and argues that provided the former are used, logical possibility does reliably signal metaphysical possibility. He further argues that "I" is an informative designator because no one is as well-placed as I am to know how correctly to apply it. Using this distinction, Swinburne contends that it is metaphysically possible that while I am thinking now, my body is completely destroyed, from which it follows that a soul, but not a body, is essential to a person. Swinburne analyzes the nature of the soul and responds to several important objections to substance dualism. He argues that souls have a basic "thisness" which explains the conceivability of a different person having a life qualitatively identical to my own. He finds the interaction objection to substance dualism weak since we often know that very different kinds of things are causally related without knowing how, argues that experiments designed to refute mental causation actually assume it, and offers an account of why a

particular soul is paired with a particular body. Finally, he offers a thought-experiment to support the plausibility of the claim that humans have thisness: it is conceivable that a person should receive incremental brain transplants and at the end have a completely different brain, yet be the same person.

In his critique of Cartesian dualism (Chapter 10), Kim provides a careful exposition of Cartesian dualism and its main supporting arguments. He then considers Princess Elisabeth's famous complaint that Descartes's view makes psychophysical causal interaction unintelligible, and develops a related "pairing problem." Kim argues that dualists cannot explain why the causal influence of a mind is paired with some bodies (and some other minds) but not others. Unlike the physicalist, Descartes cannot appeal to spatial relations to solve the problem, since he maintains that immaterial minds are not in space. But even if they were, it would not help, because spatial location does not individuate minds.

Then (Chapter 11), the late E. J. Lowe offers the alternative of an interesting non-Cartesian variety of substance dualism. He draws together three insights about persons: they are psychological (Locke), substantial (Descartes), and not necessarily immaterial (Aristotle). He emphasizes the simplicity of persons as psychological substances distinct from their bodies but (*pace* Descartes) also possessing physical characteristics consistent with simplicity. A particular body is mine because my physical properties supervene on its, it responds to my will, is known by me in a special way, and is located at the point from which I perceive the world. The simplicity of the self explains its unity at a time and justifies the ungroundedness of identity over time. This view of the self entails that neuropsychology has value as a source of facts about what goes on in nervous systems when they think or feel or act, while not providing an account of what constitutes mental states: "Thought can no more *be*, or be constituted by a brain process than a chair can *be*, or be constituted by, a set of prime numbers."

#### 1.6.1.4 Debating the unity of consciousness

Dualists often appeal to the unity of consciousness as an important datum favoring their position. In this vein, J. P. Moreland (Chapter 12) explores several theses about the nature of a unified consciousness developed by Tim Bayne and David Chalmers, and argues that they are best explained by some version of substance dualism. He defends William Hasker's argument that materialism cannot account for the unity of consciousness and contends that Bayne's "virtual phenomenalism" is inadequate, since if the "self" is only a stream of consciousness, there is no way to explain the fact that the contents of consciousness are inseparable parts of that consciousness. Moreland concludes by suggesting that resistance to substance dualism is fueled by its apparent theistic implications.

However, Tim Bayne (Chapter 13) is not convinced by the argument for substance dualism from the unity of consciousness. After rejecting as inconclusive the classical versions of the argument due to Descartes and Leibniz, Bayne considers the sophisticated contemporary versions of David Barnett, William Hasker, and Richard Swinburne. He contends that Barnett's argument rests on a "mereological illusion," that contra Hasker, a materialist can deny an atomistic view of consciousness, and that contra Swinburne, a materialist should deny that there must be a determinate fact of the matter about which candidate (if any) for a person's future self is identical with that person.

### 1.6.1.5 Debating near-death experiences

Could the widely reported cases of near-death and out-of-body experiences (NDE/OBE) support substance dualism by providing scientific evidence of the possibility of disembodied persons? Gary Habermas (Chapter 14) considers whether there is good evidence that NDEs support the claim that an immaterial human mind/personality/“soul” might function for at least minutes beyond near-death states. After critiquing the recent negative position espoused by physician and scholar Michael Marsh, Habermas turns to descriptions of many recent evidentially verified NDE reports as samples of the 300 plus reported accounts. These include cases from inside the NDEr’s room, from outside the vicinity including long distances away, reported corroborated information from deceased individuals, NDEs witnessed by healthy individuals, and reports from blind NDErs. The essay closes with a consideration of potential challenges. Habermas concludes that the available evidence certainly appears to indicate that these experiences occur even after the cessation of measurable heart and brain function.

To the contrary, Mike Marsh (Chapter 15) critically reviews assertions that NDE/OBE offer proof of extra-corporeal existence when the brain is supposedly “dead” or “clinically dead.” Marsh argues that studies have failed to produce corroborative empirical evidence for these assertions and that it is unclear how the memory required for recall could be set down with a properly dead brain at that critical time-point. He suggests that NDE/OBE occur as subjects are regaining full conscious-awareness and are analogous to hypnopompic dream awakenings. He points out that most recollections are intensely geo-physical, anthropomorphic, banal, and illogical: they provide nothing revelatory about life without a brain, or importantly, about other supposed cosmic contexts. There is also a marked chasm, Marsh argues, dividing NDE and the associated conceptualizations of “heaven” from true, classical spiritual encounters with the divine: the former are inconsistent with dogmatic (Christian) understandings of the afterlife and are decidedly not excursions of “souls” to some “heavenly” abode. Since prevalence rates are extremely low (< 1% globally), Marsh suggests that those undergoing NDE/OBE may have predisposed brains, genetically, structurally, or resulting from previous psychological stress.

### 1.6.2 *Alternatives to substance dualism*

In the bookend chapters of this section, Kevin Corcoran and Angus Menuge consider whether Christians should embrace some form of materialism about human persons. This view is often called “Christian physicalism,” although Corcoran prefers to use “physicalism” to describe more reductive views, while Menuge allows the term to include nonreductive varieties.

Corcoran (Chapter 17) argues that since Christians should embrace the truth, and materialism about human persons is true, Christians should be materialists. He then offers three main arguments in support of materialism. The aesthetic argument points out that materialism avoids a cleavage in nature that applies only to human beings. The biological argument holds that materialism gives a better account than dualism of the gradual development of consciousness. And the neuropsychological argument points to phenomena such as blindsight and phantom limb syndrome as revelatory of both the fine-grained dependence of conscious experience on the brain and the complex structure of what otherwise seems to be unified conscious experience. He turns to the hard problem of

consciousness, arguing that since phenomenal consciousness is fundamental and irreducible, certain kinds of explanation are unavailable in principle and it is likely that any explanation of how neural functioning gives rise to it may be forever beyond our grasp.

In his critique of Christian physicalism (Chapter 26), Menuge defines “Christian physicalism” very broadly, as the thesis that a human person is either identical to or constituted by a physical object possessing mental features that are either emergent from or reducible to physical properties of the brain. Menuge sees as the primary appeal of Christian physicalism (CP) its promise to reconcile Christian anthropology with a modern scientific worldview. However, Menuge argues that, being unable to account adequately for the first-person perspective or knowledge of the natural world, CP fails to ground human capacities to distinguish the self from the rest of creation, to grasp moral obligations, and to carry out plans to take care of the world; capacities required for stewardship of the natural world and thus presupposed by Scripture. Menuge also argues that these capacities are presupposed by science and that CP may not qualify as a physicalist view. If that is right, then Christian physicalism is insufficiently Christian and insufficiently physicalist and it cannot fail to disappoint those who rely on its promised reconciliation.

In between these bookends, we hear the case for and against specific versions of materialism/physicalism.

### **1.6.2.1 Debating animalism**

Eric Olson defends animalism (Chapter 18), the thesis that human persons are animals of a particular kind. He begins by clarifying to what animalism is and is not committed. He asserts that animalism neither assumes nor entails a particular metaphysical theory about the nature of animals, and does not claim to be the whole truth about human persons. Olson’s central argument for animalism parallels a common argument for substance dualism. Substance dualists often defend their position by pointing out that in introspection, we appear to ourselves to be immaterial simples, and that even the strongest physicalist arguments are not strong enough to unseat this appearance. Similarly, Olson argues that we appear to be animals, that alternative views involve much more surprising claims, and that even the strongest objections to animalism are not strong enough to overcome it. In particular, Olson argues that dualist arguments from introspection only show that we sometimes do not appear to be animals, not that we appear *not* to be animals. Olson considers several other objections, including life after death, and brain transplants, and argues that none is sufficiently compelling to unseat animalism.

In his critique of animalism (Chapter 19), Stewart Goetz does not claim to provide a knock-down refutation, but argues instead that we have stronger intuitive grounds for thinking that we are souls essentially, and only animals accidentally. Even psychologists who are critics of substance dualism agree that it reflects the normal, default beliefs of human beings. We seem to be aware in introspection that, unlike bodies, we lack separable parts, and while there are well-known objections to Descartes’s version of dualism, they can be mitigated by supposing that souls are located in space where their living bodies are, but that they occupy space in a different way than those bodies (the whole soul is present in every part of the body). Goetz argues that a major problem for animalism is that it grounds the persistence of an animal in its life, defined as a complex event. If the subevents of this life are always changing, how exactly can one claim that it is the same life that persists, and how can one ground personal identity across time?

### 1.6.2.2 Debating nonreductive physicalism

“Reduction” is a notoriously slippery word, admitting of ontological, epistemic, and linguistic interpretations, and while it may be that all physicalists are “reductionist” in some sense, many would affirm that there are important ways in which a physicalist can be a non- or antireductionist. As a leading exponent of nonreductive physicalism, Nancey Murphy (Chapter 20) sets out to defend the view from common philosophical and theological objections. While physicalism is widely believed to be the only metaphysical account of human nature compatible with developments in neuroscience, it is a matter of serious debate in the philosophy of mind whether physicalism can avoid being reductive. On the other hand, while most lay Christians hold dualist accounts of human nature, Christian scholarship over the past century has increasingly called for the acceptance of physicalism. But if physicalism is to work theologically, then a successful argument against reductionism is required. Murphy first argues that biblical and theological objections to physicalism can be countered. She then illustrates how physicalism converges with cognitive neuroscience, and finally considers in some depth a new scientific paradigm, complex dynamic systems, that undermines the claim that reductionism is synonymous with the “scientific” approach.

In response, Joshua Rasmussen (Chapter 21) provides what he sees as an in-principle refutation of any form of physicalism, including nonreductive forms. While Murphy’s defense of physicalism is largely empirical, Rasmussen’s critique is more conceptual. He develops a counting argument, inspired by Cantor’s diagonalization proofs, to show that there are more mental properties than physical properties, and therefore that *some* mental properties are not physical properties. On grounds of uniformity, he then argues that there must be a categorical difference between mental and physical properties, and concludes that *all* mental properties are nonphysical. The argument is extended to show that mental properties cannot even be grounded in physical properties, in which case all standard versions of physicalism (which, at a minimum, assume psychophysical supervenience) are false. In the last part of the chapter, Rasmussen outlines how dualists can give at least as good an explanation of the primary data used to motivate physicalism. Taken together, Rasmussen’s arguments support a “basic mentality thesis,” which asserts that at the foundation of our nature is a mental substance.

### 1.6.2.3 Debating constitutionalism

Another sophisticated alternative to reductive physicalism is constitutionalism, ably articulated and defended by Lynne Baker (Chapter 22). Baker argues that it is not inconsistent with Christian Scripture or doctrine to hold to person–body constitutionalism: that what is essential to one’s existence as a human person is the possession of a second-order capacity for a robust first-person perspective; a second-order capacity to think of oneself as oneself, “from the inside.” On this view, a human person has different modal properties than a human body and is thus not identical with it, but rather is a distinct material object that is constituted by it. Baker claims that constitutionalism offers an account of resurrection that is consistent with Christian Scripture while avoiding problems associated with Thomistic and mind-body dualist alternatives. She also holds that it offers a better account of the Incarnation than mind-body dualism by maintaining the materiality of Christ’s human nature.

In his response (Chapter 23), Ross Inman develops several challenges for constitutionalism. If persons and bodies are atom-for-atom physical duplicates then four problems

emerge. (1) How are the modal differences of persons and bodies to be grounded to establish that they are distinct objects? An appeal to relational properties will not do. (2) How many thinkers are there? If human organisms *cannot* think then, implausibly, zombies are real; if they *can* (as Baker argues) then there are two thinkers for every human person. In arguing that we can nevertheless count them as one, Baker further increases the view's unacceptable metaphysical price tag. (3) On constitutionalism, it is necessary to reject Andrew Bailey's independently plausible priority principle – that human persons possess all of their mental properties in the primary and nonderivative sense. (4) Baker cannot maintain the ontological uniqueness of human persons given that both higher nonhuman animals and human infants possess first-person perspectives. Her attempt to do so only serves to undercut her account of personhood.

#### 1.6.2.4 Debating emergent individualism

Yet another view is emergent individualism, which holds that individual subjects are radically emergent phenomena. Timothy O'Connor (Chapter 24) maintains that persons, who have the natural potential for subjective awareness, intrinsic intentionality, and intentional action, are emergent individuals. On this view, people are wholly physically composed and yet they exhibit higher order properties that, though not basic, are fundamental: persons have "new" causal powers, powers not possessed by the physical systems that compose them, including free will in a robust, libertarian sense. To make sense of this, O'Connor develops a particular version of the substratum-attribute theory of objects, defends it against objections, and applies it to human persons. He ends the chapter by recognizing that for persons, the substratum theory seems inadequate, since that theory is designed to attach substrata to universals, not particular objects. And yet, without substrata, O'Connor worries that (emergent) substance dualism may be the only reasonable alternative.

In a friendly critique (Chapter 25), Robert Koons argues for the superiority of Thomistic hylomorphism over emergent individualism. Koons provides a helpful conceptual map to show the main differences between nonreductive accounts of human persons. He then develops three reasons for preferring hylomorphism to emergent individualism. First, he argues that a top-down model of de-escalation (division of substances) is more plausible than the emergentist idea that coherent substances emerge bottom-up from many independent parts, and that de-escalation is a good fit with the hylomorphic view that substantial forms make a difference to what the parts of a composite substance do. Against many critics, Koons contends that hylomorphism gives a superior account of the possibility of disembodied existence between death and resurrection. Finally, he argues that appeal to formal causes operating through bodily instruments provides a better account of mental causation than emergent individualism's downward causation.

#### 1.6.3 Substance dualism, theology, and the Bible

The editors realize that for readers primarily interested in mainstream philosophy of mind, theological concerns about the compatibility of various theories with Scripture may not be of much interest and in deference to such readers have placed this shorter section at the end of the book. Yet, for very many theologians, Bible scholars, Christian apologists, and philosophers – including many contributors to this volume – these concerns are of the utmost importance. So we round out our *Companion* with due consideration for biblical issues.



### 1.6.3.1 Debating biblical anthropology

John Cooper defends a holistic and dualist biblical anthropology (Chapter 27). He challenges the contemporary claim that Plato's influence caused the Church Fathers mistakenly to adopt body-soul dualism rather than monism. Exploring in detail the influential work of eminent scholars Joel Green and N. T. Wright, Cooper affirms biblical holism with them but challenges their claim that holism is inconsistent with dualism. While Green affirms essential embodiment and thus a view inconsistent with postmortem existence, Wright affirms the two-stage biblical eschatology that grounds dualism but holds to ontological holism instead. However, their key arguments against dualism are "compromised by problematic hermeneutics, conceptual confusions and faulty reasoning." Furthermore, Scriptural accounts of both humanity's creation and a relationship with God between death and bodily resurrection are inconsistent with monism and establish a dualism compatible with holism.

In response, Joel Green (Chapter 28) wonders why "soul" is vanishing from English translations of the New Testament (NT). He argues that the term reflects an anthropological dualism influenced by Plato that is mistakenly read into the text. Historical inquiry has uncovered both the important *monistic* influence of the content and theological trajectories of Israel's Scriptures on NT authors and the diverse conceptions of "soul" operative in the NT world. Other sociocultural forms of inquiry also suggest *holistic* anthropology and there are now better ways to understand anthropological terminology that were previously taken as an obvious indicator of a partitive view. The vanishing of soul enables readers of the NT to engage in a richer theological exploration of the nature of embodied, situated human life.

### 1.6.3.2 Debating the incarnation

Luke van Horn (Chapter 29) contends that, in debating philosophical anthropology and mind, Christian philosophers have not yet paid enough attention to Christ's Incarnation. Van Horn contends that the Incarnation is inconsistent with materialism about human beings. While several materialist accounts of the Incarnation have been proposed, Van Horn argues that they face serious metaphysical objections and tend to revive ancient heresies like Nestorianism, according to which Christ consists of two distinct persons. He goes on to argue that substance dualism has the resources to avoid these (and other) objections, and hence that Christians should prefer substance dualism to materialism.

To the contrary, Trenton Merricks (Chapter 30) claims that dualism cannot give a credible account of the Second Person of the Trinity being embodied in Jesus of Nazareth. He argues that dualist accounts of what it means for an immaterial being to control a body lead to absurd conclusions: we are disembodied if our soul is not exercising control over the body; all three Persons of the Trinity are embodied in every human being; and becoming human does not imply having a human body. Merricks suggests that a physicalist view of the Incarnation gives a simpler account of embodiment that solves several problems facing dualists. He is aware of critics who argue that it is metaphysically impossible for an immaterial being to become a material being, but counters that this assumes kind essentialism – if something is an object of a particular kind, it is essentially of that kind – and gives reasons to resist this thesis.

### 1.6.3.3 Debating the general resurrection

At the end of the book, we consider the end times and whether dualism or materialism offers the best account of the general resurrection. Jonathan Loose (Chapter 31) begins by

pointing out the advantages of dualism in accounting for personal identity across the bridge of death to the resurrection. An immaterial soul accounts for the highly intuitive “simple” view of personal identity – there is always a determinate fact of the matter about whether two individuals are the same person – and makes it easy to see how the very same person who died can live again. But materialism typically endorses a complex view of personal identity – which suggests identity can be indeterminate – and seems to entail a gap between death and new life that no person can cross. However, several materialist models of the resurrection have been proposed. Loose critiques Peter van Inwagen’s “simulacrum” model and Dean Zimmerman’s “falling elevator” model, and argues that Lynne Baker’s constitutional theory is also unable to meet the challenge. He therefore concludes that if there are no better materialist proposals, believers in the Resurrection should be dualists.

The last word on the end times goes to Peter van Inwagen (Chapter 32). He begins by noting that our popular culture encourages unbiblical ideas of the afterlife, and questions whether dualism is required to make sense of the Resurrection. Certainly, the biblical perspective is not the Platonic one that the body is a prison-house and that disembodied existence is ideal. On the other hand, van Inwagen concedes that there are serious difficulties in the idea that God can simply reassemble a person who was made of different particles at different times. Van Inwagen does not claim to offer a definitive justification of a materialist resurrection – the equivalent of a theodicy for the problem of evil – but instead offers a defense to show that materialism is consistent with the Resurrection. While admitting that his “simulacrum” model has won few adherents, he defends it against objections as a “just-so-story” that shows that a materialist resurrection is conceivable, and argues that it is less problematic than the alternatives proposed by Baker and Zimmerman.

We hope that you will benefit from the rich interchange that follows!

## References

- Baker, Lynne Rudder. 2001. “Materialism with a Human Face.” In *Soul, Body and Survival: Essays on the Metaphysics of Human Persons*, edited by Kevin Corcoran, 159–180. Ithaca, NY: Cornell University Press.
- Baker, Lynne Rudder. 2013. *Naturalism and the First-Person Perspective*. Oxford: Oxford University Press.
- Baker, Mark C., and Stewart Goetz, eds. 2011. *The Soul Hypothesis: Investigations into the Existence of the Soul*. London: Continuum.
- Bayne, Tim. 2012. *The Unity of Consciousness*. Oxford: Oxford University Press.
- Block, Ned. 1978. “Troubles with Functionalism.” In *Minnesota Studies in the Philosophy of Science IX*, edited by C. Wade Savage, 261–325. Minneapolis: University of Minnesota Press.
- Corcoran, Kevin. 2006. *Rethinking Human Nature: A Christian Materialist Alternative to the Soul*. Grand Rapids, MI: Baker Academic.
- Davidson, Donald. 1980. “Mental Events.” In *Essays on Actions and Events*. Oxford: Clarendon Press.
- Dennett, Daniel C. 1978. “Current Issues in the Philosophy of Mind.” *American Philosophical Quarterly*, 15(4): 249–261.
- Goetz, Stewart, and Charles Taliaferro. 2008. *Naturalism*. Grand Rapids, MI: Eerdmans.
- Goetz, Stewart, and Charles Taliaferro. 2011. *A Brief History of the Soul*. Malden, MA: Wiley Blackwell.
- Hasker, William. 1999. *The Emergent Self*. Ithaca, NY: Cornell University Press.
- Jackson, Frank. 1982. “Epiphenomenal Qualia.” *Philosophical Quarterly*, 32: 127–136.
- Jackson, Frank. 1986. “What Mary Didn’t Know.” *Journal of Philosophy*, 83: 291–295.

- Kim, Jaegwon. 1982. "Psychophysical Supervenience." *Philosophical Studies*, 41(1): 51–70.
- Kim, Jaegwon. 1984. "Concepts of Supervenience." *Philosophy and Phenomenological Research*, 45(2): 153–176.
- Kim, Jaegwon. 1998. *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation*. Cambridge, MA: MIT Press.
- Kim, Jaegwon. 2006. "Emergence: Core Ideas and Issues." *Synthese*, 151: 547–559.
- Kim, Jaegwon. 2011. *Philosophy of Mind*, 3rd edn. Boulder, CO: Westview Press.
- Koons, Robert, and George C. Bealer, eds. 2010. *The Waning of Materialism*. New York: Oxford University Press.
- Lewis, David. 1980. "Mad Pain and Martian Pain." In *Readings in the Philosophy of Psychology*, edited by Ned Block, vol. 1, 216–222. Cambridge, MA: Harvard University Press.
- Loose, Jonathan. 2012. "Constitution and the Falling Elevator." *Philosophia Christi*, 14(2): 439–449.
- Lowe, E. J. 2008. *Personal Agency: The Metaphysics of Mind and Action*. Oxford: Oxford University Press.
- Moreland, J. P. 2008. *Consciousness and the Existence of God: A Theistic Argument*. New York: Routledge.
- Moreland, J. P. 2009. *The Recalcitrant Imago Dei*. London: SCM Press.
- Moreland, J. P. 2014. *The Soul*. Chicago: Moody Press.
- Murphy, Nancey. 2006. *Bodies and Souls, or Spirited Bodies?* New York: Cambridge University Press.
- Nagel, Thomas. 1974. "What Is It Like to Be a Bat?" *Philosophical Review*, 83: 435–450.
- Nagel, Thomas. 2001. *The Last Word*. New York: Oxford University Press.
- Nagel, Thomas. 2012. *Mind and Cosmos: Why the Materialist Neo-Darwinian Conception of Nature is Almost Certainly False*. New York: Oxford University Press.
- O'Connor, Timothy, and Jonathan D. Jacobs. 2003. "Emergent Individuals." *Philosophical Quarterly*, 53(213): 540–555.
- O'Connor, Timothy, and Hong Yu Wong. 2005. "The metaphysics of emergence." *Noûs*, 39(4): 658–678.
- Olson, Eric T. 2007. *What are We? A Study in Personal Ontology*. Oxford: Oxford University Press.
- Putnam, Hilary. 1968. "Brains and Behavior." In *Analytical Philosophy*, edited by R. J. Butler, 1–19. Malden, MA: Wiley Blackwell.
- Ryle, Gilbert. 1949. *The Concept of Mind*. London: Hutchinson.
- Searle, John. 1980. "Minds, Brains and Programs." *Behavioral and Brain Sciences*, 3: 417–457.
- Searle, John. 1983. *Minds, Brains, and Science*. Cambridge, MA: Harvard University Press.
- Searle, John. 1992. *The Rediscovery of the Mind*. Cambridge, MA: MIT Press.
- Searle, John. 2001. *Rationality in Action*. Cambridge, MA: MIT Press.
- Searle, John. 2007. *Freedom and Neurobiology*. New York: Columbia University Press.
- Smart, J. J. C. 1959. "Sensations and Brain Processes." *Philosophical Review*, 68: 141–156.
- Swinburne, Richard. 2013. *Mind, Brain, and Free Will*. Oxford: Oxford University Press.
- Tye, Michael. 2003. *Consciousness and Persons: Unity and Identity*. Cambridge, MA: MIT Press.
- Van Horn, Luke. 2010. "Merricks's Soulless Savior." *Faith and Philosophy*, 27(3): 330–341.
- van Inwagen, Peter. 2007. "A Materialist Ontology of the Human Person." In *Persons: Human and Divine*, edited by Peter van Inwagen and Dean Zimmerman, 199–215. New York: Oxford University Press.
- Wittgenstein, Ludwig. 2009 [1953]. *Philosophical Investigations*, translated by G. E. Anscombe. Wiley Blackwell.

# Redressing Substance Dualism

WILLIAM G. LYCAN

I have been a materialist about the mind for forty years, since first I considered the mind-body issue. In all that time I have seen exactly one argument for mind-body dualism that I thought even *prima facie* convincing.<sup>1</sup> And like many other materialists, I have often quickly cited standard objections to dualism that are widely taken to be fatal, for example, Lycan (1987, 2–3) – notoriously the dread interaction problem. My materialism has never wavered. Nor is it about to waver now; I cannot take dualism very seriously.

Being a philosopher, of course I would like to think that my stance is rational, held not just instinctively and scientistically and in the mainstream but because the arguments do indeed favor materialism over dualism. But I do not think that, though I used to. My position may be rational, broadly speaking, but not because the arguments favor it: though the arguments for dualism do (indeed) fail, so do the arguments for materialism. And the standard objections to dualism are not very convincing; if one really manages to be a dualist in the first place, one should not be much impressed by them. My purpose in this chapter is to hold my own feet to the fire and admit that I do not proportion my belief to the evidence.<sup>2</sup>

The dualism I shall defend is Cartesian, “substance” dualism. “Property” dualism is more popular nowadays, but it is logically weaker than substance dualism, assuming that a Cartesian ego would necessarily have some irreducibly mental properties, and so, you would think, must be more defensible. (But actually the retreat avails little; see Section 2.7 and Section 2.8.)

## 2.1 Arguments for Materialism

Arguments for materialism are few. Tyler Burge and others have maintained that the naturalistic picture of the world is more like a political or religious ideology than like a position well supported by evidence, and that materialism is an article of faith based on the worship of science.<sup>3</sup> That is an overstatement. But it is true that the original twentieth-century materialists felt no need to defend materialism itself. Ryle gave no such argument that I can recall; he only inveighed against the particularly Cartesian “dogma of the Ghost in the Machine.” Ullin Place, founder of the identity theory, gave none; he was originally a

behaviorist who bravely and honestly acknowledged that introspectible occurrent sensations were a problem for behaviorism and, while making an exception for them, tried to account for them within the materialist framework, but without defending the need to do so.<sup>4</sup>

J. J. C. Smart was perhaps the first to offer reasons.<sup>5</sup> First, he expressly appealed to the scientific view of the world:

[S]ensations, states of consciousness . . . seem to be the one sort of thing left outside the physicalist picture, and for various reasons I just cannot believe that this can be so . . . That everything should be explicable in terms of physics . . . except the occurrence of sensations seems to me frankly unbelievable . . . The above is largely a confession of faith. (Smart 1959, 142–143)

Just so, and just so. I too simply refuse to believe in spookstuff or surds in nature. But this *argumentum ad recuso credere* is no *argument* at all; it is at best, in David Lewis's famous phrase, an incredulous stare.

But then Smart did advance a real argument; he appealed to mind–brain correlations: It is reasonable to think that every mental state or event at least has a *corresponding* type of brain state or event. The best, because most parsimonious, explanation of those correlations is that the mental states/events *just are* the “corresponding” brain states/events. (In general: When Xs are invariably accompanied by Ys and you can find nothing to distinguish Xs from Ys, the best explanation is that Xs just are Ys.)

I firmly agree that parsimony or simplicity is a reason for preferring one hypothesis to another.<sup>6</sup> But it is a very posterior reason. Not only does it always carry the qualification “other things being equal,” but *many*, nearly all, other things must be equal before parsimony is called in to break the tie. And no party to the mind–body dispute will deny that dualists have found plenty of features that seem to distinguish mental states/events from neurophysiological ones – even if, as materialists contend, all those differences are ultimately specious. To anyone uncontaminated by neuroscience or materialist philosophizing, the mental does not *seem* physical in any way at all, much less neurophysiological. The parsimony argument does not even come in the door until it is agreed that we can find nothing to distinguish mental states from neurophysiological ones. And the latter will not be agreed any time soon.

More decisively, Smart's alleged correlations have never materialized. Notice that he certainly meant *type*-correlations; unless one were already presuming token identity, it would have been otiose to say that for every mental token, there is a “corresponding” neurophysiological token. There may be a few type-correlations holding within particular species, but if so they are very few. Whatever is in common as between all human beings who believe that a Frenchman has been assassinated in Trafalgar Square (to take an old example of Dennett's), that feature could not possibly be characterized in neuroscientific terms; there are no “Frenchman” neurons, nor “assassination” areas of the cerebral cortex; at best the feature would be a complicated set of external psychosemantic relations to Frenchmen, to assassinations, and to Trafalgar Square. (And good luck to the psychosemanticist.)<sup>7</sup><sup>8</sup>

Matters improved when, independently of each other, David Lewis and D. M. Armstrong offered their respective causal arguments for identifying mental states and events with neurophysiological states and events (Lewis 1966, 1972; Armstrong 1968, 89–90).

Their common idea was that mental concepts are causal role concepts, and so they afford role-occupant identifications (as in the case of genes and segments of DNA molecules). For example:

- 1 Pain = Whatever state of a person plays role P (being typically caused by tissue damage, and in turn causing wincing, crying out, withdrawal, favoring, etc.) [We know this a priori; we have all got the concept of pain.]
- 2 The occupant of role P = the firing of c-fibers<sup>9</sup> (i.e., it is c-fiber firings that are typically caused by tissue damage, etc.). [Discovered empirically by neuroscientists.]

---

∴ 3 Pain = the firing of c-fibers. QED

This was an important development, because the argument was deductive and obviously valid. But is either premise true? Premise 1 was counter-examined early on by Keith Campbell (1970, 100–109), Ned Block (1978, 277–282) and others. A state of a creature, or for that matter of an assembly of Tinkertoys or beer cans, could occupy the commonsense role of pain but without being mental at all, much less feeling like a pain.<sup>10</sup> Remember, premise 1 is a conceptual or at least a priori claim; fantastical imaginary cases are fair play. And remember how little information there is in a commonsense causal analysis of pain; see, for example, Armstrong's analysis (Armstrong 1968, 310–316).

Also, premise 1 is a culpably good premise for materialists. *Obviously*, if the very concept of pain is a causal concept like "poison" or "sunburn" or "footprint," and what gets caused is physical motion in the form of behavior, it would be hard to resist the inference that pain is physical. Premise 1 does not formally beg the question, but it comes close. And I shall argue shortly that a dualist can quite reasonably resist it. The dualist should never and would never accept premise 1 in the first place. Pain is first and foremost what presents itself to consciousness as pain, what feels like pain. That sort of sensation is indeed caused by tissue damage and does cause the customary behavior, but those are plainly a posteriori facts. (For the dualist to insist that they are *contingent* would beg the question, but the present materialist claim is that they are not just necessary but a priori.<sup>11</sup>)

More generally: The materialist of course takes the third-person perspective; s/he scientifically thinks in terms of looking at other people, or rather at various humanoid bags of protoplasm, and explaining their behavior. But the dualist is back with Descartes in the first-person perspective, acquainted with the contents of her own consciousness, aware of them as such. Notice carefully that we need not endorse many of Descartes's own antique and weird views about the mind (that it is entirely nonspatial, that it has no parts, that mentality requires language). The point is only that we know the mind primarily through introspection. Duh! That idea *may*, very surprisingly, be wrong; it has been attacked by Ryle, by Wittgenstein, and by Sellars, among others.<sup>12</sup> But it is obviously common sense, and to deny it is a radical move. Note that it does not entail or even strongly suggest that the mind is *better* known than the body or the rest of the physical world.

Turning to the Lewis-Armstrong premise 2, it seems fine until one realizes that its first word is "the." Premise 2 begs the question against the dualist view that role P is causally overdetermined: the typical causes cause both neural events and immaterial pain events, and pain-behavior is doubly caused by the neural events aforementioned and the

immaterial pain events. (One may feel – as I certainly do – that this overdetermination view is silly and unmotivated.<sup>13</sup> But on what evidence? Of course, the view offends against parsimony, but as before, parsimony must wait till all substance has been adjudicated.)

In “Naturalism, Materialism and First Philosophy,” Armstrong gives a general argument for the thesis that we should count a thing as real and admit it to our ontology only if we can identify it by its causal powers, for: “if a thing lacks any power, if it has no possible effects, then, although it may exist, we can never have any *good reason* to believe that it exists” (Armstrong 1981, 156). That claim leads directly to materialism, Armstrong contends, because we know of no physical effects produced by supposedly immaterial occurrences; “[m]ost neurophysiologists would be astounded to hear that what happens to the brain has any other cause except earlier states of the brain and its physical environment” (154).

Of course, the causal criterion is controversial, because numbers and sets seem to be exceptions. And the argument for it is flawed, because as we know from epistemology, knowledge does not require that one’s belief has been caused by the fact known.<sup>14</sup> But in any case the inference to materialism rests on remorselessly third-person scientism and (again) on the tacit assumption that the physical effects are not overdetermined.

David Papineau (2002) offers a simple deductive argument for materialism, based on the causal completeness of physics:<sup>15</sup> Conscious events have physical effects; all physical effects have sufficient physical causes; the physical effects of conscious causes are not, or not always, overdetermined by physical causes; therefore conscious events are physical events.

This too is an argument rather than merely a confession of faith. But Papineau admits that there is nothing to support the first premise against epiphenomenalism, preestablished harmony, and other noninteractive dualisms save appeal to “standard principles of theory choice” (Papineau 2002, 23). And he does in fact appeal to parsimony: “If both views can accommodate the empirical data equally well, then ordinary scientific methodology will advise us to accept the simple view that unifies mind and brain, rather than the ontologically more profligate story which has the conscious states dangling impotently from the brain states” (Papineau 2002, 23). Of course I agree, but this argument is hardly deductive, and without it the first premise begs the question.

And on behalf of good old Cartesian interactive dualism, the same point can be made against the third premise that I have made against Lewis and Armstrong, that so far as has been shown, physical events *are* systematically overdetermined by physical and nonphysical causes. As before, there is no evidence against that view; it only offends parsimony.

Besides, given quantum indeterminism, it is open to the dualist to deny that all physical effects have sufficient physical causes, as Descartes certainly did for the case of human actions. The argument for the completeness of physics itself has to be compelling enough to convince the dualist. I know of no other arguments for materialism.

## 2.2 Objections to Substance Dualism

Here, very briefly, are the four standard objections I highlighted in *Consciousness* (1987). (This was the usual perfunctory throat-clearing; we all know why Cartesian dualism was rejected.) (1) The interaction problem of course. (2) Cartesian egos are excrescences, queer, and obscure, and they are not needed for the explanation of any publicly known fact. (3) Even if conceptually intelligible, Cartesian interaction violates known laws of physics, particularly the conservation of matter-energy (Cornman 1978, 274). (4) Evolutionary

theory embarrasses dualism, since we have no idea how natural selection could have produced Cartesian egos; an immaterial substance could not possibly be adaptive.

In his well-regarded textbook *Matter and Consciousness*, Paul Churchland too has rehearsed objections (1)–(4) (Churchland 1988 [1984], 18–21), and like Smart he appeals to simplicity. He adds two further criticisms: (5) In comparison to neuroscience, dualism is explanatorily impotent (pp. 18–19). (This is not a repetition of (2). The point is neither that Cartesian egos are *entia non grata* nor that they are not *needed* for explanation. It is that the dualist theory itself explains nothing.) (6) All known mental phenomena are highly dependent on detailed brain function (p. 20); Churchland says this “comes close to being an outright refutation of (substance) dualism.”

There are even more objections, not mentioned by Churchland or me: (7) Ryle argued that Descartes got the epistemology radically wrong. If substance dualism were true, we could not possibly ever know what was going on in someone else’s mind; yet we have such knowledge very easily.<sup>16</sup> (8) There are problems of unity and individuation. In virtue of what are the contents of a Cartesian mind contents of that mind rather than another one? We might answer that by reference to the uniquely associated body, but then what accounts for the unique relation between the mind and that body? (9) There is the more specific “causal pairing problem” formulated by John Foster (1991) and pursued by Jaegwon Kim (2005): Why does one immaterial ego rather than another count as causally interacting with a given body? It seems there could be two Cartesian minds running exactly in parallel but having different sets of physical effects if any.

The case sounds overwhelming. But now suppose, if you can, that you are a substance dualist.

Would you be cowed? No. There are nine objections to your view. Of course there are; any interesting philosophical view faces at least nine objections. The question is, how well you can answer them? And I contend that the dualist can answer them fairly respectably. I shall start with the *interaction problem* because I think it is by far the most damaging.

## 2.3 The Interaction Problem

Entirely nonspatial mental events could not possibly cause physical motion in the way that billiard balls cause physical motion; that is nearly tautologous. But (to my knowledge) no one has ever believed that mental events do cause physical motion in the way that billiard balls do.<sup>17</sup> What, then, is the problem?

I believe it is that, as Robinson (1982) puts it, even now we have *no good model at all* for Cartesian interaction. Descartes tried the analogy of gravitational attraction, which was promptly blasted by Elisabeth. No one has done much better since.<sup>18</sup>

I agree that the lack of a good model is a trenchant objection and not just a prejudice. But it is hardly fatal as yet. For one thing, the lack results at least partly from the fact that we have no good theory of causality itself. The theories that have been called theories “of causality” often seem to have been theories of different things, not of a single phenomenon with agreed-upon clear cases.<sup>19</sup>

More to the point, causal realism itself has not been popular until pretty recently.<sup>20</sup> Twentieth-century theories of causality were predominantly Humean, though of course there were exceptions. The more recently prevalent counterfactual theories such as David Lewis’s (1973) are not antirealist, but they are semi-Humean, requiring only specific forms



of counterfactual dependence; and no reason has been given why physical events could not depend counterfactually on Cartesian mental events. (Note that if one says that the relevant counterfactuals need actual categorical truth makers, one thereby gives up the counterfactual theory in question. Lewis himself held that the counterfactuals' truth makers were facts about other possible worlds and relations between them, but that would not per se embarrass Cartesian interaction.)

Now, further: Give up any tacit assumption of physical determinism. I believe that will help reduce the sense of outrage, and even hint at a model: perhaps mind-body interaction is only probabilistic, as purely physical causation is.<sup>21</sup>

And now acknowledge the prevalence of weird quantum phenomena. Though there is as yet no model for Cartesian interaction, microphysics gets more and more bizarre, and indeed itself resorts (on some interpretations of quantum mechanics) to quasi-mental vocabulary.<sup>22</sup> We cannot possibly be sure that no model for Cartesian interaction will emerge.

Finally, I have a revisionist suggestion. The big problem for interaction is and remains the utter nonspatiality of Cartesian egos. (By now we can all tolerate action at a distance. But action at a distance is at least at a *distance*.) My suggestion is that the dualist give up nonspatiality. Descartes had his own seventeenth-century metaphysical reasons for insisting that minds are entirely nonspatial, but we need not accept those. Why not suppose that minds are located where it feels as if they are located, in the head behind the eyes?<sup>23</sup> If it be protested that our heads are already entirely full of physical stuff and that two things cannot occupy the same region of space at the same time: (1) Immaterial minds are not physical. And what is true is only that two *physical* things cannot occupy the same region of space at the same time. For that matter, (2) our heads are *not* entirely full of physical stuff. Physically, they are mostly empty space, with minuscule particles zipping through them at very high speeds.<sup>24</sup> Notice further that ghosts and disembodied spirits supposedly move about in space, and that does not cause readers/audiences any conceptual dissonance.

For the rest of this chapter, I shall assume that minds, though immaterial, have locations in physical space. (It may be wondered wherein, then, minds are *immaterial*.<sup>25</sup> In at least two ways: They do not have other physical properties such as mass or charge; and unlike brain matter, they are not made of atoms or subatomic particles.)

## 2.4 Objections (2)–(4)

(2) *Excrescencehood*. In complaining that Cartesian egos are ill-behaved entities that fail to earn their keep, the materialist here lodges firmly in the third-person perspective and assumes a very strong form of the "Theory" theory, that the sole job of mental ascription is to explain facts about the physical world. But as before, the dualist cannot be expected to grant any such assumption in the first place. Cartesian minds are not explanatory posits at all, much less posited to explain physical facts. They are known from the inside, and there is nothing queer or obscure about that. (And arguments are given for the view that they are Cartesian rather than physical.) Nor is the strong "Theory" theory tenable: As Kathleen Wilkes (1993) has argued, mental ascriptions have all sorts of uses other than explanatory ones.

(3) *Laws of physics*. Here I am not qualified to adjudicate. But it has been argued by E. Averill and B. Keating (1981) that the conservation laws regarding linear momentum and

matter-energy come in weak versions and stronger versions.<sup>26</sup> The weak versions are what actually figure in physics. But they are logically compatible with Cartesian causation. The stronger versions have been adumbrated by some philosophers, but are not required for physics and also would beg the question against the dualist.

Classical Cartesian egos do have one property that is flatly incompatible with modern physics' conception of spacetime: Cartesian mental events occur in real time, but not in space; that is impossible if time is only one of the four dimensions of spacetime.<sup>27</sup> Fortunately, we have abandoned Descartes's nonspatiality assumption.

(4) *Evolutionary theory*. At least as stated, the objection is that natural selection could not have produced Cartesian egos because they could not be adaptive. But that assumes an extreme Panglossianism: that a trait or entity could not emerge in the course of evolution by natural selection unless *it*, itself, were adaptive. No evolutionary biologist believes that. Frank Jackson (1982) points out on behalf of epiphenomenalism that many un- or maladaptive traits are concomitants or by-products of other traits that were adaptive. Swinburne (1986) offers some tentative suggestions as to how an immaterial soul might have evolved.

More to the point, why could the egos not be adaptive, given that they causally interact with the physical? (We have already addressed the interaction problem, and are entitled to assume on the dualist's behalf that minds and bodies interact.) The objector may appeal to the causal completeness of physics, even granting the possibility of overdetermination noted in our discussion of Papineau's argument: It is never solely because of a Cartesian ego that a creature did well in the struggle for resources and safety, and indeed the creature's physical characteristics would have taken care of that on their own. But on the overdetermination view, it was not, in fact, solely because of the physical characteristics either.

Unlike mine, Churchland's version of the evolution objection does not specifically appeal to adaptiveness. What he says is, rather:

For purposes of our discussion, the important point about the standard evolutionary story is that the human species and all of its features are the wholly physical outcome of a purely physical process . . . We are notable in that our nervous system is more complex and powerful than those of our fellow creatures. Our inner nature differs from that of simpler creatures in degree, but not in kind. (Churchland 1988 [1984], 21)

Which simply and blatantly begs the question.<sup>28</sup>

## 2.5 Churchland's Added Objections

Each of Churchland's two new objections is a bit odd. (Which is itself odd, because his book is a textbook.)

(5) *Explanatory impotence*. The premises are true; neuroscience explains a great deal and dualism explains hardly anything. But the comparison is misplaced. Dualism competes, not with neuroscience (a science), but with materialism, an opposing philosophical theory. Materialism per se does not explain much either. (It would have explained Smart's mind-brain correlations, had they existed.)

Materialism does have one explanatory advantage: Obviously it explains why brain facts are highly relevant to mental facts, better than dualism does. But the dualist does have an explanation. Though many physical stimuli affect the mind, those that do are meager in

their information content. Even patterned retinal hits greatly underdetermine the incredibly rich visual experiences that result, and the immediate perceptual beliefs that the subject will form as a result of those. Prodigious transducing is needed in order to send the required gigantic mass of hyper-finely structured information to and through the pineal gland. And that is what the brain is for. (Plausible? Of course not. But I think only because dualism itself is not plausible. *If* one actually is a dualist and holds fixed the assumption of Cartesian interaction, the transducer explanation is pretty good.)

(6) “*Neural dependence.*” Here I must quote:

If there really is a distinct entity in which reasoning, emotion, and consciousness take place, and if that entity is dependent on the brain for nothing more than sensory experiences as input and volitional executions as output, *then one would expect reason, emotion, and consciousness to be relatively invulnerable to direct control or pathology by manipulation or damage to the brain.* But in fact the exact opposite is true. (Churchland 1988 [1984], 20; emphasis in original)

Of course the opposite is true. But why would any dualist accept the premise’s second conjoined antecedent? What dualist ever said or even implied that the mind is dependent on the brain for nothing more than sensory experiences as input and volitional executions as output? Descartes himself knew very well that the mental depended in a detailed way upon the brain. And the transducer explanation applies here as well. We may even add that cognition may interdepend in a close way with brain activity. There is no reason to suppose that the mind can do complicated reasoning without the aid of a physical calculator; in the real world, most people cannot do complicated reasoning without the aid of a physical calculator. Mind–brain interaction may be constant and very intimate. (Here again, the picture is implausible, but only because dualism and Cartesian interaction are implausible in the first place. Subtract those two implausibilities, and the rest of the picture is not bad at all.)

## 2.6 The Remaining Objections

(7) *Epistemology of other minds.* Cartesian egos were nonspatial, which made their epistemology seem utterly hopeless. But remember that Cartesian dualism is interactionist. Mental events (now occurring inside our heads) cause behavior. And so, for all that has been shown, we know that our own mental events cause behavior and we infer like causes from like effects. This is a far from satisfactory solution, but except for analytical behaviorism, no other is less problematic. The present objection adds nothing to the interaction problem itself.

Ryle thought that you can just see (some of) other people’s mental states and events, and do not even unconsciously have to infer them. I think that view contains a very large grain of truth, even though I also think that the mental states and events themselves are neuro-physiological states and events inside our skulls. But this is an issue in the philosophy of perception, not for philosophers of mind.

(8) *Unity and individuation.* Again, Cartesian dualism is interactionist. The contents of a Cartesian mind are contents of that mind (rather than another) in virtue of its exclusive causal connection to the relevant human body.

But then what explains the unique relation between the mind and that body? This is indeed an embarrassing question, but the answer is to be found in whatever would explain the appearance of minds in the evolutionary process. The objection collapses into objection (4).

(9) *The “pairing problem.”* Observe that this is not cured by noting that Cartesian dualism is interactionist. But all parties agree that the problem would be solved if Cartesian minds were located in space. So it is not a problem for my version of Cartesianism.

Even without my spatializing move, there are options. For one thing, we need not grant that such differences in causal efficacy need explaining; causal relations may be brute (Robb and Heil 2003). For another, as Karen Bennett (2007) points out, there are comparable differences in purely physical scenarios, so the dualist is not distinctively afflicted.<sup>29</sup>

## 2.7 Property Dualism

Would property dualism be even better off? If so, that would be excellent news for some. But, actually, I believe not.

During the last thirty or so years, property dualism has been doing surprisingly well: Campbell (1970), Madell (1988), Robinson (1988, 2004), Seager (1991), Forrest (1993), Strawson (1994), Chalmers (1996), Taliaferro (1996), Bealer (1994, 1997, 2010), Stubenberg (1998), Griffin (1998), Siewert (1998), Hasker (1999), Rosenberg (2004), Zimmerman (2010), and others (and see especially Koons and Bealer 2010). The general idea is that property dualism is tenable (or even demonstrated), but we are not crazy.

I reject this disparity. I think that most of the standard objections to substance dualism (SD) count as effectively against property dualism (PD), and that PD is hardly more plausible, or less implausible, than SD. Granted, assuming that a Cartesian ego would *eo ipso* have some immaterial mental properties, SD is logically stronger than PD; so one would need a reason for accepting SD over and above PD, and there must be at least one objection that applies to SD but not to PD. However, as we shall see, nonsubstance property dualism (PD and  $\sim$ SD) faces at least two objections that SD does not.<sup>30</sup>

Let us revisit our nine objections to SD, starting with the four that I agree do *not* afflict PD quite as severely if at all. In each of those first four cases I shall argue that the difference is small, if only because the original objection to SD was overblown to begin with.

(2) *Excrescencehood.* If we look at the issue from just the third-person explanatory point of view, the PDist must think that PD explains more or better or both than does SD. What, then, would nonphysical properties explain that an ego would not? Well, we do seem to be aware of properties that are problematic for materialism: intentional properties, qualia, “what it’s like” and such. And if Hume is right, we are not aware of our selves, and certainly not of their being immaterial substances.

But the objection had little bite to begin with. We do *seem to be aware of* the properties, from within, from the first-person perspective. As I have emphasized in Section 2.1 above, dualists do not think of either Cartesian egos or immaterial properties as explanatory posits. I know of no dualist who grants the “Theory” theory, nor do I know of any who holds either SD or PD on explanatory grounds; dualist arguments are generally deductive.<sup>31</sup>

There is one explanatory role that SD might play *if* PD is assumed, for a “how-possibly” question would arise: Why or how on earth would a merely physical object, even one as complex as a brain, give rise to immaterial properties? We do not see how it could. If persons have immaterial mental properties, then most likely the persons themselves are or incorporate immaterial things. The idea would be that while there is nothing puzzling about an immaterial substance’s having immaterial properties, it is extremely strange to think that an otherwise purely physical object might have them. (Of course, we may wonder where the

immaterial substance came from in the first place, especially if it is to emerge when a functioning brain does. But that is a different issue; on emergence, see Section 2.8.)

Still, it is possible that immaterial properties might play an explanatory role even though Cartesian egos play none. Here is one:<sup>32</sup> PD can explain psychological continuity, by appealing to continuing dependence of mental states on one and the same physical brain, while SD can only refer tautologically to continuing dependence on one and the same immaterial ego. So the objection may give PD a slight edge over SD.

(4) *Evolution.* Evolutionary theory is supposed to embarrass dualism generally; but it does seem easier at least to imagine weird properties emerging from brain complexity than whole individual substances doing so.<sup>33</sup>

(6) *Neural dependence.* To his argument quoted in Section 2.5 above, Churchland adds, “Property dualism, note, is not threatened by this argument, since, like materialism, property dualism reckons the brain as the seat of all mental activity” (Churchland 1988 [1984], 20). I have already argued that finely tuned brain processing may be constantly and absolutely necessary for activity in a Cartesian mind, but that is an extra commitment.

There is also a more abstract ontological issue that does give PD a further slight advantage over SD: that of disembodied existence. PD is not per se committed to the possibility of such, but, arguably, SD is. Being attributes of or inhering in brain states, PD’s immaterial properties must vanish when the brain ceases to function, but an entirely immaterial Cartesian ego might persist. Churchland and others may see this as a large difference in plausibility.

But, first, remember that the *conceptual* possibility of disembodied existence is granted by nearly everyone, the only exceptions being analytical behaviorists and (if any) analytical eliminativists; Armstrong (1968) uses this as one of his basic desiderata for theories of the mind, and uses it to rule out analytical behaviorism in short order. The identity theorists and the functionalists have both vigorously insisted on it. And so should the PDist; the dependence of her/his immaterial mental properties on brains is hardly a conceptual truth.

The PDist is free to maintain, and almost surely will maintain, that the immaterial properties are metaphysically dependent on brains or brain analogs. But is not SD committed to the metaphysical independence of Cartesian egos from brains and from all else that is physical? Now the question is that of metaphysical possibility, and the PDist again sees a difference.

But it is not obvious that SD is committed to the metaphysical possibility of disembodied existence. There is no entailment (in the strict sense of logical or even conceptual entailment); at least one additional premise would be needed – perhaps a Descartes-Hume principle regarding distinct existences, or the doctrine(s) involved in an attempt to show that in some suitably constrained cases, metaphysical possibility can be inferred from conceptual possibility, as in Chalmers (2002). No such thesis is analytic or even uncontroversial, even if it is in fact a metaphysical truth.

So the relevant difference between PD and SD is that from SD but not from PD, *it is possible to argue* that minds (metaphysically) might exist quite independently of brains. Perhaps an advantage for PD, but I do not see it as large.

(9) *Causal pairing.* We found this objection to be weak, but so far it does not even apply to PD.

So, now, on to the other five objections.

(1) *Interaction.* As always, the problem is widely thought to be fatal to SD. But there is no obvious difference in the case of PD. How could a nonphysical *property* causally affect

physical ones, any more than could a nonphysical substance? Perhaps by way of a brute or primitive psychophysical law (Chalmers 1996); but equally such laws could hold between immaterial properties of Cartesian egos and physical properties.

(To return for a moment to nonspatiality, which was probably the nastiest characteristic of a strictly Cartesian ego, and not just because it is what makes the interaction problem so bad. Is it worse for an ego-substance to be nonspatial than for a property to be? Perhaps a bit, since we do not much think of properties as being spatially located in the first place, while we do usually think of *things* as located. But we could just agree to locate immaterial properties in brains. And, of course, we have departed from Descartes and spatialized the egos themselves. Thus, were we to locate immaterial properties in brains, we could and should also agree to locate immaterial egos in heads.<sup>34</sup> There is here no advantage for PD.)

(3) *Laws of physics.* As with interaction, the nonphysical properties are supposed to have causal powers. Whatever problems SD has with the conservation laws or with general relativity, PD should have too (bar those which may arise *solely* from traditional Cartesian egos being nonspatial).

(5) *Explanatory impotence.* This has been seen to be a nonissue in the first place.

(7) *Epistemology.* SD's ontology, especially nonspatially interpreted, makes the epistemological problem at first look more grotesque than for PD. But as before, if the interaction problem were solved, that appearance would cease, and since PD has the interaction problem too, the epistemological objection is in fact no worse for SD.

(8) *Unity and individuation.* Here too, if the interaction problem were solved, so too would be the individuation problem (also, if we do spatialize the egos, there is simply no problem in the first place).

## 2.8 Property Dualism's Special Problems

Precisely for the reason Churchland gives, that PD "reckons the brain as the seat of all mental activity," non-Cartesian PD incurs at least two objections that SD does not.

First objection: strong emergence.<sup>35</sup> If a brain state or event *itself* has immaterial properties, that is, that are not properties of an immaterial thing, the properties are emergent in an objectionably strong sense: namely, for the state/event to have the property is not just constituted by the subject's entirely material parts being arranged in a particular way. Either that, or as Sellars (1962, 1965) and Galen Strawson (2006) have reminded us, the brain state or event itself has some immaterial component.<sup>36</sup> Note that the sense of emergence here is a matter of wholes and their component individual parts, as discussed in Meehl and Sellars (1956); that is why the objection does not apply to immaterial properties of immaterial substances.<sup>37</sup> Sense may be made of such emergence, but it is considered a serious liability of a view, on whatever topic, that entails it.<sup>38</sup>

But an important distinction must be made.<sup>39</sup> I have been speaking of "Cartesian minds," but they are not the only sorts of individual substances that could free dualism from the emergence problem. First, an immaterial substance need not be a continuant mind; it could be a momentary mind. But more interestingly and pertinently, it could be a noncontinuant immaterial individual that is not a mind at all: say, a sense datum. The sense-datum theorist too avoids the emergence problem. Of course, there is little point in appealing to sense data unless there is a mind that is acquainted with them; and the point remains that Cartesian minds do afford SD the present advantage over PD.

Second objection: If a sentient subject is not a Cartesian substance but an animal with a brain, and what are immaterial about that brain are only properties of otherwise physical states and events of/in it, how does the subject obtain knowledge of those of her own mental states that have the immaterial properties?<sup>40</sup> A Cartesian mind knows its own immaterial properties either because they are merely modifications of it or because they are directly presented to it in acquaintance, but a brute physical brain seems the wrong sort of thing to be acquainted with immaterial properties, even if the property instances have spatial location.

I suppose the best way of answering this would be (i) to maintain that PD is not true of beliefs and other cognitive states, so that they can be identified with brain states, (ii) adopt a higher-order-perception view of introspective knowledge, and then (iii) argue as I did versus Ryle that if the interaction problem has been solved for properties, the epistemological problem is not additional. But (i) and (ii), at least, will be contested by many dualists.

## 2.9 Property Dualism Scorecard

(1) Each of the first four objections has a little differential force. But I have argued that they have little if any more than that. (2) I do not think the remaining five have any differential force at all. And (3) non-Cartesian PD faces two objections that SD does not, of which the first, at least, is serious. I myself think (3) outweighs (1). But I shall be magnanimous. Verdict: PD is perhaps a little better off than SD. But not so much better off that property dualists should go on boasting that they are not so crazy as to be Cartesians.

## 2.10 Conclusion

I mean to have shown here that although Cartesian dualism faces some serious objections, that does not distinguish it from other philosophical theories, and the objections are not an order of magnitude worse than those confronting materialism in particular. There remain the implausibilities required by the Cartesian view; but bare claim of implausibility is not argument. Nor have we seen any good argument for materialism. The dialectical upshot is that, on points, and going just by actual arguments as opposed to appeals to decency and what good guys believe, materialism is not significantly better supported than dualism.

Yet, I am inclined to believe, the charge of implausibility is not irrational or arational either, and I would not want this chapter to turn anyone dualist. Have a nice day.<sup>41,42</sup>

## Notes

1. It is the argument from qualia stated in my book *Consciousness* (Lycan 1987, 84–85). But it is countered by the representational theory of qualia, defended by me there and in Lycan (1996, 1998, 2001). For the record, I now believe that there is a more powerful argument for dualism based on intentionality itself: from the dismal failure of all materialist psychosemantics; see Note 7.
2. In mitigation, I would note that no philosopher has ever proportioned her/his belief to the evidence; see van Inwagen (1996). Note that we also always hold our opponents to higher standards of argumentation than we obey ourselves. I have always felt entitled to thumb my nose at dualism so long as no valid deductive argument has been presented for it, each of whose premises I *must* accept. My admirers (however many or few those may be) need not worry about my allegiance: I have no sympathy with any dualist view, and never will. This chapter is only an uncharacteristic exercise in intellectual honesty. It grew out of a seminar in which for

methodological purposes I played the role of a committed dualist as energetically as I could. That was a strange feeling, something like being a cat burglar for a few months. You could see there was a *modus vivendi* here, however uncongenial.

3. Burge (1993). I believe my own faith in materialism is based on science worship. For a compelling justification of the latter attitude (as opposed to an actual argument for materialism), see Armstrong (1973); but Armstrong also does argue there for materialism, as I shall address below.
4. Functionalism, the reigning materialist view of the past 40 years or so, does not strictly entail materialism, but has been held largely because it is the least bad way of remaining a materialist. The only functionalist dualist I have ever known or heard of was the late Roland Puccetti (1981); note that "functionalism" in his title meant, functionalist *materialism*.
5. Of course there had previously been the damningly quick positivist argument from the verification theory of meaning to analytical behaviorism, but: (i) That was no argument for materialism *per se*. And (ii) so much the worse for the verification theory; in any number of cases, it led too quickly to bad metaphysics, such as the view that there are not really any little subvisible particles such as electrons.
6. That is, as more likely to be true. Despite obvious examples of curve-fitting and the like, not everyone grants this; for example, eloquently, Van Fraassen (1980). In his contribution to the 1967 Presley volume, the late Don Gunner (1967) asked even more eloquently why simplicity should be a reason for belief: "[A] question should be raised as to whether the principles of parsimony and simplicity have not become restrictive principles of stinginess and over-simplification. (Nature is lush, prodigal, messy, wasteful, sexy, etc.)" (pp. 4–5). ("Etc.?" )
7. For the record, I think intentionality is a much greater obstacle to materialism than is anything to do with consciousness, qualia, phenomenal character, subjectivity, and so forth. If intentionality itself is naturalized, those other things are pretty easily explicated in terms of it (Lycan 1996). But in my view, current psychosemantics is feeble: it treats only concepts tied closely to the thinker's physical environment; it addresses only thoughts and beliefs, and not more exotic propositional attitudes whose functions are not to be correct representations; and it does not apply to any thought that is even partly metaphorical.
8. Not that Smart ever believed in type correlations between *beliefs* and anything neurophysiological; he remained a behaviorist about beliefs. Also, as a referee has pointed out, there may still be species-specific *one-way* correlations that need explaining. I shall take up this point and offer a dualist alternative in Section 2.5, in response to objection (5).
9. To correct a common misconception: neither Place, Smart, nor Armstrong mentioned c-fibers. C-fibers were introduced to the identity theory literature by Putnam in "Minds and Machines" (1960). Thanks to Jack Smart for the reference.
10. Note that this sort of counterexample should not be taken to encourage "zombie" objections to materialism generally. It refutes only claims that (i) are conceptual and (ii) explicate mental expressions in terms of circumstantial and behavioral events described in ordinary English, as in Lewis's and Armstrong's commonsense causal theory of the mind (sometimes misleadingly called "analytical functionalism"). It has no force against a posteriori versions of functionalism – at least, not on its own, without supplementation by cumbersome and contested apparatus such as that of David Chalmers (1996).
11. As a referee noted, not many materialists themselves accept premise 1. I am not sure whether even Armstrong did for much longer.
12. Ryle's material on this point was pretty desperate. Wittgenstein's private language argument has never been well understood, much less generally accepted. Sellars (1956) argued more clearly for the publicity of mental terms' linguistic meaning, but even Sellars's detailed and ingenious account is contested at many points. The first-person perspective is emphasized by Searle (1992) and by Georgalis (2006).
13. Its falsity is simply assumed by William S. Robinson (1982) – himself a dualist – and by David Papineau. But for defense, see Mills (1996).
14. At least two reasons support the latter point. First, we have at least some knowledge of the future. Second, although *perception* requires causation by the state of affairs perceived, not all knowledge is perceptual.
15. See also his excellent paper (2001) on the history of that idea.
16. In making this observation, Ryle meant to be calling attention to a familiar phenomenon, not offering a solution to the philosophical problem of other minds (though of course he did go on to furnish a roughly behaviorist solution to that problem). The point here is, rather, that *if* traditional substance dualism were correct, the familiar phenomenon would be a *complete* mystery and not just the source of a standard philosophical problem.
17. It is just possible that Davidson's (1970) early view of events, causation, and laws entails this.



18. Of course, one can give up mind-to-body causation and go to epiphenomenalism, but that view retains body-to-mind causation and so does not claim to solve the interaction problem. It is motivated mainly by the causal closure of physics and the implausibility of overdetermination. (Being no scholar, I do not know whether these are what motivated Malebranche. But see again Robinson 1982.) It faces extra objections of its own, but those too can be answered: Jackson (1982), Robinson (1982, 2003, 2007). Robinson is probably America's most committed and ingenious defender of epiphenomenalism. For myself, I do not see its advantages over Cartesian dualism as weighty.
19. A similar but distinct point is made by Karen Bennett (2007).
20. Notice a general tendency in philosophy: when working in one area, we feel free to presuppose positions in other areas that are (at best) highly controversial among practitioners in those areas. To take a limiting example, philosophers nearly everywhere outside epistemology presuppose that we have reason to believe in the external world. If we do have it – as I too presume we do – epistemology has delivered not one tenable account of how that can be so. (Except *possibly* my own)
21. There is the causal closure of physics. But the causal closure principle is an empirical thesis, and a recent one (Papineau 2009, sec. 2.2). Lowe (2008) accepts the closure principle but explores ways of making an emergentist substance dualism compatible with closure.
22. Note that I am very far from joining in the suggestion made by some that quantum mechanics can explain important facts about consciousness (e.g., Lockwood 1989; Marshall and Zohar 1990; Hodgson 1993; Smith 2003; Stapp 2004). I do not believe that quantum mechanics could explain anything at all about consciousness *per se*; see Lycan (2011). The present point is only about models for Cartesian interaction.
23. After drafting this, I learned that my bold move was anticipated by no less a figure than Isaac Newton (thanks to Hylarie Kochiras for the references):

That substances of different kinds do not penetrate each other [i.e., co-occupy space] does not at all appear from the phenomena. And we ought not rashly to assert that which cannot be inferred from the phenomena. (Newton 1962b [1713], 360–361)]

No being can exist which is not in some way related to space. God is everywhere, created minds are somewhere, and body is in the space it occupies. Whatever is neither everywhere nor somewhere does not exist. (Newton 1962a [before 1670], 141)

Following Penelhum (1970, ch. 2), dualist W. D. Hart (1988) tries to make room for the idea that an entirely disembodied person might be able to see real things as well as merely have visual experience, and he argues that that would require the person to be located in physical space. (Real seeing requires causal contact, and, according to Hart, causal interaction is a matter of energy flow within spacetime. Hart posits a nonphysical “psychic energy” that will satisfy the conservation laws.) Colin McGinn (1995) argues that although mental states and events have some space-related properties, they are fundamentally nonspatial. Therefore, he infers, “we need, at a minimum, a new conception of space. We need a conceptual breakthrough in the way we think about the medium in which material objects exist, and hence in our conception of material objects themselves” (p. 226). That is, space has to be futuristically reconceived, in order to allow for the existence in the real world of items that are at best quasi-spatial.

24. There seem to be dualists who accept mental substances but are emergentists, and for whom the substances are not immaterial egos entirely distinct from our bodies; possibly Taliaferro (1996), Hasker (1999), Lowe (2008), and Zimmerman (2010). More on this below; but how does a spatialized substance dualism differ from the emergentist variety? The emergentists too locate soul-substances in space and make them causally dependent on brains – I think the chief difference is in the nature or at least the degree of causal dependence. Emergence is an intimate relation; an emergent entity depends on its base or substrate for its very existence and for at least some of its nature. Spatialized substance dualism makes no such assumption; even spatialized Cartesian souls may have been, for example, created independently by God, and *may* be able to float free of brains and bodies.
25. Meehl and Sellars (1956) distinguished two senses of “physical,” one of which was just *defined* as, spatial. It was the other that mattered to Sellars (1971), roughly, “figures in the explanation and description of ordinary matter,” which presumably does not apply to Cartesian egos.
26. My philosophy of physics colleagues John Roberts and Marc Lange have at least cautiously concurred. See also Larmer (1986).
27. Actually, William Lane Craig (2001) has suggested that it is only “coordinate time” that requires spatial coordinates; if time functions rather as a parameter, it is independent of space. I have no idea what that means, or whether it is true, or whether if true it would save Descartes. But thanks to Ken Perszyk for the reference.

28. Since question begging is such an elementary and easily identifiable fallacy, why do we seasoned professionals commit it as often as we do? (I am no exception.) I believe the answer is a more general fact: that we accept deductive arguments mainly when we already believe their conclusions.
29. Foster (1991) offers an ingenious solution to the problem, but it is too complicated to expound here.
30. Francescotti (2001) suggests that actually PD necessitates SD. Having considered a number of possible ways of defining "physical particular," he concludes that "on any plausible definition of a physical particular, it is unclear how one can consistently endorse property dualism . . . while rejecting substance dualism" (p. 114). If PD does secretly entail SD, then of course my comparative view is true quite regardless of my arguments for it. Another defense of the entailment claim is given by Sharpe (2011). Zimmerman (2010) argues that if one holds PD but not SD, one probably must also be an adverbialist regarding sensory qualities, and that given adverbialism, a weak form of substance dualism is more plausible than any strict materialism.
31. In teaching the mind-body problem, we tend to treat Cartesian dualism as a *theory*, competing with behaviorism, the identity theory, and so forth. But remember that for Descartes it was not a theory, but the conclusion of each of several deductive arguments that he thought were sound. He did not *like* the view; he agreed with Elisabeth that it was extremely problematic. But what about those deductive arguments – Leibniz'-Law objections, the knowledge argument, conceivability arguments, and so forth? Some of them have been thought to be sound, and would establish PD but not SD. Moreover, that is one obvious reason why their proponents hold PD but not SD. Of course: if any such argument is sound, then PD is true, period. Yet, note that any such would need to be checked, to make sure that a parallel argument would *not* establish SD. But if there is no sound argument for SD, the received view is correct and we should accept PD but not SD. I am assuming at least for the sake of discussion that no argument for PD is deductively sound; for my question is that of whether, *considered as theories only*, SD is less plausible than mere PD.
32. Suggested by a reviewer.
33. As was observed to me by Bryce Huebner (but see again the "emergentist" substance dualists in Note 24).
34. Choice: Would an ego be merely a spatial point (*nonmass*), or would it have extension within the brain? I am thinking of it in the latter way, since at least my own mind seems fairly large and intersects with various sense modalities, but Chisholm (1976) seems to go for the point view.
35. Dave Chalmers has made this point in conversation.
36. And that way panpsychism lies, as Strawson happily points out; for discussion, see Lycan (2006, 2011). I shall not here try to make the case that SD is less implausible than panpsychism.
37. Presumably it does not apply to emergentist substance dualism either, though the existence of an immaterial substance within the physical universe would itself remain extremely problematic. SD does not imply the strong emergence of properties. As in our earlier discussion of excrescencehood, we may wonder where an immaterial ego might come from, and/or why the development of a brain might be accompanied by an ego's coming into being, but these questions do not elicit the immediate intuition of impossibility that attends strong emergence of properties. Moreover, SD per se is not committed to any emergence doctrine.
38. Churchland (1988 [1984]) notes that the property dualist can get around the emergence objection by claiming that the immaterial properties are fundamental, on the model of electromagnetic properties:

Such a view may be called *elemental property dualism* . . . Unfortunately, the parallel with electromagnetic phenomena has one very obvious failure. Unlike electromagnetic properties, which are displayed at all levels of reality from the subatomic level on up, mental properties are displayed only in large physical systems that have evolved a very complex internal organization. . . . They do not appear to be basic or elemental at all. (p. 21)

Churchland is assuming, reasonably, that any fundamental property will be found throughout the universe, and in particular existed long before the descent of living creatures.

39. Thanks to a reviewer for pointing this out.
40. Amber Ross has raised this question to me.
41. Thanks to Robert Francescotti, to Laurie Paul, and to each of several anonymous reviewers for very helpful comments on previous drafts.
42. This chapter is a lightly edited amalgam of my articles. "Giving Dualism its Due," *Australasian Journal of Philosophy*, 87 (2009): 551–563, and "Is Property Dualism Better Off than Substance Dualism?" *Philosophical Studies*, 164 (2013): 533–542. It has not been updated. Material from these papers is reprinted here by permission of Taylor & Francis and Springer respectively.

## References

- Armstrong, D. M. 1968. *A Materialist Theory of the Mind*. London: Routledge and Kegan Paul.
- Armstrong, D. M. 1973. "Epistemological Foundations for a Materialist Theory of the Mind." *Philosophy of Science*, 40: 178–193.
- Armstrong, D. M. 1981. "Naturalism, Materialism and First Philosophy." *Philosophia*, 8: 261–276. Reprinted in *The Nature of Mind and Other Essays*, Ithaca, NY: Cornell University Press; page references are to the latter.
- Averill, E., and B. Keating. 1981. "Does Interactionism Violate a Law of Classical Physics?" *Mind*, 90: 102–107.
- Bealer, G. 1994. "Mental Properties." *Journal of Philosophy*, 91: 185–208.
- Bealer, G. 1997. "Self-Consciousness." *Philosophical Review*, 106: 69–117.
- Bealer, G. 2010. "The Self-Consciousness Argument: Functionalism and the Corruption of Content." In *The Waning of Materialism*, edited by Robert Koons and George Bealer, 137–158. Oxford: Oxford University Press.
- Bennett, K. 2007. "Mental Causation." *Philosophy Compass*, 2: 316–337.
- Block, Ned. 1978. "Troubles with Functionalism." In *Minnesota Studies in the Philosophy of Science: Perception and Cognition*, edited by C. Wade Savage, 261–325. Minneapolis: University of Minnesota Press.
- Burge, Tyler. 1993. "Mind–Body Causation and Explanatory Practice." In *Mental Causation*, edited by J. Heil and A. Mele, 97–120. Oxford: Clarendon Press.
- Campbell, Keith. 1970. *Body and Mind*. New York: Anchor Books.
- Chalmers, D. 1996. *The Conscious Mind*. Oxford: Oxford University Press.
- Chalmers, D. 2002. "Does Conceivability Entail Possibility?" In *Conceivability and Possibility*, edited by T. Gendler and J. Hawthorne, 145–200. Oxford: Oxford University Press.
- Chisholm, R. M. 1976. *Person and Object*. La Salle, IL: Open Court.
- Churchland, P. M. 1988 [1984]. *Matter and Consciousness*. Cambridge, MA: Bradford Books/MIT Press.
- Cornman, J. W. 1978. "A Nonreductive Identity Thesis About Mind and Body." In *Reason and Responsibility*, edited by J. Feinberg, 272–283. Encino, CA: Dickenson Publishing.
- Craig, W. L. 2001. *God, Time, and Eternity*. Berlin: Springer.
- Davidson, Donald. 1970. "Mental Events." In *Experience and Theory*, edited by L. Foster and J. W. Swanson, 79–101. Amherst: University of Massachusetts Press.
- Forrest, P. 1993. "Difficulties with Physicalism, and a Programme for Dualists." In *Objections to Physicalism*, edited by H. Robinson, 251–270. Oxford: Clarendon Press.
- Foster, J. 1991. *The Immaterial Self*. London: Routledge.
- Francescotti, R. 2001. "Property Dualism without Substance Dualism?" *Philosophical Papers*, 30: 93–116.
- Georgalis, N. 2006. *The Primacy of the Subjective*. Cambridge, MA: MIT Press.
- Griffin, D. R. 1998. *Unsnarling the World-Knot*. Berkeley: University of California Press.
- Gunner, D. L. 1967. "Professor Smart's 'Sensations and Brain Processes.'" In *The Identity Theory of Mind*, edited by C. F. Presley, 1–20. St. Lucia, Australia: University of Queensland Press.
- Hart, W. D. 1988. *The Engines of the Soul*. Cambridge: Cambridge University Press.
- Hasker, W. 1999. *The Emergent Self*. Ithaca, NY: Cornell University Press.
- Hodgson, David. 1993. *The Mind Matters: Consciousness and Choice in a Quantum World*. Oxford: Clarendon Press.
- Jackson, Frank. 1982. "Epiphenomenal Qualia." *Philosophical Quarterly*, 32: 127–136.
- Kim, J. 2005. *Physicalism, or Something Near Enough*. Princeton: Princeton University Press.
- Koons, R. C., and G. Bealer, eds. 2010. *The Waning of Materialism*. Oxford: Oxford University Press.
- Larmer, R. 1986. "Mind–body Interactionism and the Conservation of Energy." *International Philosophical Quarterly*, 26: 277–285.

- Lewis, D. K. 1966. "An Argument for the Identity Theory." *Journal of Philosophy*, 63: 17–25.
- Lewis, D. K. 1972. "Psychophysical and Theoretical Identifications." *Australasian Journal of Philosophy*, 50: 249–258.
- Lewis, D. K. 1973. "Causation." *Journal of Philosophy*, 70: 556–567.
- Lockwood, Michael. 1989. *Mind, Brain, and the Quantum*. Oxford: Basil Blackwell.
- Lowe, E. J. 2008. *Personal Agency: The Metaphysics of Mind and Action*. Oxford: Oxford University Press.
- Lycan, William G. 1987. *Consciousness*. Cambridge, MA: MIT/Bradford Books.
- Lycan, William G. 1996. *Consciousness and Experience*. Cambridge, MA: MIT/Bradford Books.
- Lycan, William G. 1998. "In Defense of the Representational Theory of Qualia." In *Philosophical Perspectives. Volume 12: Language, Mind and Ontology*. Atascadero, CA: Ridgeview Publishing.
- Lycan, William G. 2001. "The Case for Phenomenal Externalism." In *Philosophical Perspectives. Volume 15: Metaphysics*. Atascadero, CA: Ridgeview Publishing.
- Lycan, William G. 2006. "Resisting ? -ism." *Journal of Consciousness Studies*, 13: 65–71.
- Lycan, William G. 2011. "Recent Naturalistic Dualisms." In *Light Against Darkness: Dualism in Ancient Mediterranean Religions and the Contemporary World*, edited by A. Lange, E. Meyers, and R. Styers, 348–363. Göttingen, Germany: Vandenhoeck and Ruprecht.
- Madell, G. 1988. *Mind and Materialism*. Edinburgh: University of Edinburgh Press.
- Marshall, I., and D. Zohar. 1990. *The Quantum Self: Human Nature and Consciousness Defined by the New Physics*. New York: Morrow.
- McGinn, Colin. 1995. "Consciousness and Space." *Journal of Consciousness Studies*, 2: 220–230.
- Meehl, P., and W. Sellars. 1956. "The Concept of Emergence." In *Minnesota Studies in the Philosophy of Science. Volume I: The Foundations of Science and the Concepts of Psychoanalysis*, edited by H. Feigl and M. Scriven, 239–252. Minneapolis: University of Minnesota Press.
- Mills, Eugene O. 1996. "Interactionism and Overdetermination." *American Philosophical Quarterly*, 33: 105–115.
- Newton, Isaac. 1962a [before 1670]. *De Gravitatione et aequipondio fluidorum*. In *Unpublished Scientific Writings of Isaac Newton*, edited and translated by A. R. Hall and M. B. Hall, 89–156. Cambridge: Cambridge University Press.
- Newton, Isaac 1962b [1713]. Draft Conclusion for General Scholium, MS. C (MS. Add 3965 pp. 360–362), in *Unpublished Scientific Writings of Isaac Newton*, edited and translated by A. R. Hall and M. B. Hall. Cambridge: Cambridge University Press.
- Papineau, David. 2001. "The Rise of Physicalism." In *Physicalism and its Discontents*, edited by C. Gillett and B. M. Loewer, 3–36. Cambridge: Cambridge University Press.
- Papineau, David. 2002. *Thinking About Consciousness*. Oxford: Oxford University Press.
- Papineau, David. 2009. "The Causal Closure of the Physical and Naturalism." In *The Oxford Handbook of Philosophy of Mind*, edited by B. McLaughlin, A. Beckermann, and S. Walter 53–65. Oxford: Oxford University Press.
- Penelhum, T. 1970. *Survival and Disembodied Existence*. New York: Humanities Press.
- Puccetti, Roland. 1981. "Why Functionalism Fails." Presented at the American Philosophical Association (Eastern Division) meeting.
- Putnam, Hilary. 1960. "Minds and Machines." In *Dimensions of Mind*, edited by Sidney Hook, 148–179. New York: Collier Books.
- Robb, D., and J. Heil. 2003. "Mental Causation." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. Accessed June 28, 2017. <http://plato.stanford.edu/entries/mental-causation/>.
- Robinson, H. 2003. "Dualism." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. Accessed June 28, 2017. <http://plato.stanford.edu/entries/dualism/#SubDua/>.
- Robinson, William S. 1982. "Causation, Sensations and Knowledge." *Mind*, 91: 524–540.
- Robinson, William S. 1988. *Brains and People*. Philadelphia, PA: Temple University Press.

- Robinson, William S. 2004. *Understanding Phenomenal Consciousness*. Cambridge: Cambridge University Press.
- Robinson, William S. 2007. "Epiphenomenalism." *The Stanford Encyclopedia of Philosophy* (Spring 2007 edition), edited by Edward N. Zalta. Accessed June 28, 2017. <http://plato.stanford.edu/entries/epiphenomenalism/>.
- Rosenberg, G. 2004. *A Place for Consciousness*. Oxford: Oxford University Press.
- Seager, W. 1991. *Metaphysics of Consciousness*. London: Routledge.
- Searle, John R. 1992. *The Rediscovery of the Mind*. Cambridge, MA: MIT Press.
- Sellars, Wilfrid. 1956. "Empiricism and the Philosophy of Mind." In *Minnesota Studies in the Philosophy of Science. Volume I: The Foundations of Science and the Concepts of Psychoanalysis*, edited by H. Feigl and M. Scriven, 253–329. Minneapolis: University of Minnesota Press.
- Sellars, Wilfrid. 1962. "Philosophy and the Scientific Image of Man." In *Frontiers of Science and Philosophy*, edited by R. Colodny, 35–78. Pittsburgh, PA: University of Pittsburgh Press.
- Sellars, Wilfrid. 1965. "The Identity Approach to the Mind–Body Problem." *Review of Metaphysics*, 18: 430–451.
- Sellars, Wilfrid. 1971. "Seeing, Sense Impressions, and Sensa: A Reply to Cornman." *Review of Metaphysics*, 24: 391–447.
- Sharpe, K. 2011. "Property Dualism Entails Substance Dualism." Presented at the APA (Central Division) meetings, Minneapolis, MN, April.
- Siewert, C. 1998. *The Significance of Consciousness*. Princeton: Princeton University Press.
- Smart, J. J. C. 1959. "Sensations and Brain Processes." *Philosophical Review*, 68: 141–156.
- Smith, Quentin. 2003. "Why Cognitive Scientists Cannot Ignore Quantum Mechanics." In *Consciousness: New Philosophical Perspectives*, edited by Q. Smith and A. Jokic, 409–446. Oxford: Oxford University Press.
- Stapp, Henry. 2004. *Mind, Matter, and Quantum Mechanics*. Berlin: Springer.
- Strawson, G. 1994. *Mental Reality*. Cambridge, MA: Bradford Books/MIT Press.
- Strawson, G. 2006. "Realistic Monism – Why Physicalism Entails Panpsychism." *Journal of Consciousness Studies*, 13: 3–31. Reprinted in *Consciousness and its Place in Nature*, edited by A. Freeman. Exeter: Imprint Academic.
- Stubenbergh, L. 1998. *Consciousness and Qualia*. Amsterdam: John Benjamins.
- Swinburne, R. 1986. *The Evolution of the Soul*. Oxford: Oxford University Press.
- Taliaferro, C. 1996. *Consciousness and the Mind of God*. Cambridge: Cambridge University Press.
- Van Fraassen, Bas. 1980. *The Scientific Image*. Oxford: Oxford University Press.
- van Inwagen, Peter. 1996. "It Is Wrong, Everywhere, Always, and for Anyone, to Believe Anything upon Insufficient Evidence." In *Faith, Freedom, and Rationality*, edited by J. Jordan and D. Howard-Snyder, 137–154. London: Rowman & Littlefield.
- Wilkes, Kathleen. 1993. "The Relationship between Scientific Psychology and Common Sense Psychology." In *Folk Psychology and the Philosophy of Mind*, edited by S. Christensen and D. Turner, 144–187. Hillsdale, NJ: Lawrence Erlbaum.
- Zimmerman, D. 2010. "From Property Dualism to Substance Dualism," *Aristotelian Society*, 84 (Suppl.): 119–150.

## Further Reading

- Kripke, S. 1972. "Naming and Necessity." In *Semantics of Natural Language*, edited by D. Davidson and G. Harman, 253–355. Dordrecht: D. Reidel.

## PART I

# Articulating Substance Dualism



# Substance Dualism: A Defense

CHARLES TALIAFERRO

“Substance dualism” is sometimes caricatured as introducing an ad hoc, mysterious bifurcation between the evident, physical body of persons and an immaterial, non-spatial, immortal, invisible soul. Mind-body dualism is routinely dismissed in introductory philosophy texts, and in the philosophy of mind literature, as thoroughly undermined by our everyday experience of ourselves and as utterly antiscientific. To make things even worse for substance dualists, some theologians charge that dualism denigrates the body, linking it with death and sin, versus the soul that must seek to transcend or, better, escape its body if it is to be saved. While it might appear that substance dualism would enjoy some support from religious traditions, for some Christians, in particular, dualism is seen to be antithetical to the central role of the body in the incarnation and resurrection. Some Christian philosophers have therefore joined in the virtual juggernaut secular movement of materialism or physicalism (I use these last two terms interchangeably). So, substance dualism is often thought to face an up-hill battle for both secular and religious reasons.

There are three sections in this chapter. Section 3.1 sets out to separate the caricatures of dualism from a serious philosophical and theological view of human, and nonhuman animal nature. Section 3.2 addresses one of the key sources for discontent with substance dualism: the assumption that we have a clear, problem-free understanding of what it is to be physical. I contend that we do not have a clear concept of mind-independent physical objects and events. We have instead (and necessarily so) a clearer understanding of what is mental (ideas, concepts, thoughts, thinking, feeling, reasoning) and what may be called *mental causation* than we do of what is physical, including physical causation. Historically, the position I adopt on this has its roots in the work of John Locke and Thomas Reid. I then use this affirmation of the primacy of the mental to articulate reasons why we should not identify what we can recognize as mental with what physicalists (or materialists) identify as physical. In Section 3.3 I develop a further argument for why we should believe that human persons are not numerically identical with their bodies.



### 3.1 Substance Dualism

In accepting my assignment from the editors to defend *substance dualism*, I accepted the task using their terms; after all, using a standard philosophical lexicon, I would be labeled a substance dualist. However, it should be duly noted that “dualism” is a term that only came into play late in the history of ideas, and then at first to describe the great Persian religion Zoroastrianism with its cosmic division between good and evil. Perhaps its Zoroastrian past plays a role in the thinking that we so-called dualists separate the soul and body as good and evil. Moreover, the term suggests that those of us who are dualists, believe that there are ultimately only two kinds of things. Both matters are regrettable. First, as I hope to convince you, there is no essential denigration of the body in a plausible version of dualism. Neither dualism nor materialism, by themselves, imply an illicit disdain of the body, and the form of dualism I hope to convince you of has a built-in commitment to what I refer to as *the virtues* (or goods) of *embodiment*.

Second, there is no reason to assume that those called dualists insist there are only two kinds of things. This is perhaps why none of the so-called seminal dualists historically – from Pythagoras and Plato to Augustine and Descartes – ever used the term “dualism” (or its cognates). Given complete free rein over editorial matters, I would prefer arguing that *monism is false* (the theory that there is only one kind of thing, for example, that which is physical or material, is false), whereas *pluralism is true* (the theory that there are many different kinds of things is true). In light of that approach, please read what I am referring to as *substance dualism* as holding that *there are at least two kinds of things*, not that there are *only* two kinds of things.

Let us now turn to a positive concept of substance dualism, a position I have elsewhere defended using the term *integrative dualism*. According to substance (or integrative) dualism, a healthy, functioning human person lives and acts as a functional unity. Persons are substantial, concrete subjects who endure over time and (when mature and healthy) have the powers to think, to remember, to reason, to sense (feeling a myriad of bodily states including proprioception; they see, hear, taste, smell), to have emotions, to speak, and to act in or as fully embodied beings. Under healthy, ideal conditions, to engage with me in conversation is for us to be fully available to each other. It would be very odd for us to think dualists are committed to thinking that you and I are not so available, but are instead (primarily) invisible souls that are using our bodies like puppeteers. This more bifurcated caricature of dualism is preeminently advanced by Gilbert Ryle in *The Concept of Mind*, but it is also found more recently in work by Trenton Merricks who thinks that, if substance dualism is true, one cannot kiss one’s spouse but only his or her body (Ryle 1949; Merricks 2007). While a properly integrated substance dualism implies no such bifurcation, it should be noted that there are unhealthy, nonideal conditions when this integration falls apart. From an ethical point of view, I might be involved in masking my true intentions and goals and be only going through the motions of appearing to be your friend. My intentions and goals might even be (without using a metaphor) invisible or undetectable. And to take up Merricks’s case, you may believe that your spouse is lovingly offering you a kiss, whereas the spouse is faithless and plotting to poison you later that night. His outward appearances are merely that, a kind of masquerade. And in cases of serious physical (or psycho-physical) impairment, there might be respects in which my embodiment is compromised. Imagine I lose my powers of proprioception – the ostensibly direct awareness I have of my spatial location, for example, I can sense my legs are in the position they are in from “inside,” and

without having to make further, visual observation or tactile confirmation. In such a case, we might well come to think that my control over my limbs requires the kind of volitional effort involved in moving a puppet.

Elsewhere I have filled out my preferred integrative form of dualism by highlighting what I refer to as *the virtues of embodiment* (Taliaferro 2001a). From this enhanced perspective, I do not seek to identify embodiment only in terms of the functional unity of persons who have the powers of sensation, agency, feelings, and so on. I, rather, propose that we recognize these powers as *virtues*, not moral virtues, but virtues in the sense of nonmoral excellences. So, the power to see and to think are *good-making powers*, while they are not like the moral virtues (such as courage), they are essential in our fostering and pursuing moral virtues. I believe that this expanded understanding of embodiment is supported by our deep appreciation (and implicit recognition) of the good of embodiment. Although substance dualists might differ on this point, there is no reason why substance dualists cannot affirm that embodiment is such a basic good, that embodiment is itself part of the *telos* or purpose of finite personal, human life. (On this point, a substance dualist may be seen as adopting a similar position to the one we find in the work of Thomas Aquinas.) Our implicit recognition of the goodness of embodiment may be readily illustrated: if you learned that I fell down some stairs and lost my powers to think, feel, reason, and so on, but you went on to ask: "Is Taliaferro hurt?" you would either not fully believe the report of my falling or you would be making a joke or revealing that you are linguistically impaired. In contemporary philosophy of mind, there is often a failure to take note of the values that are involved in the mind-body relationship. I offer an historical account of how a value-laden view of embodiment is evident in the first works of philosophy in English, though, sadly, this was largely neglected in the evolution of modern philosophy (Taliaferro 2005).

If substance dualists are able to view human persons as unified (while allowing that vice or damage can bring about terrible dysfunctions), what does substance dualism offer us that we do not get with some forms of materialism? In the next section, I offer reasons why materialism is unacceptable in terms of mental-physical identity, but here I note that one can appreciate the significance of substance dualism when we reflect on one of the most important facts about human persons: we are all going to die. Is our death – which we may refer to as *biological death* – a substantial change? Arguably, at a person's biological death, the body is a corpse or "the remains," but the person as an individual, substantial being has ceased to be, from a biological or physical point of view. If a person were the very same thing as its body, we would be inclined to think "the person" was not so much an individual, but a mode or a phase that the body went through. Modes or phases are periods or ways a thing is, as opposed to being concrete individual things (see Hoffman and Rosenkrantz 1997). Being a student or a runner are modes or phases you go through. I suggest that we have very strong reasons for thinking that we persons are not modes or phases of some other thing. Modes do not think, feel, act, sense, remember, talk, and so on. (You as a student, study, and as a runner, run, but being a student or being a runner does not study or run.) Instead, thinking, feeling, acting, sensing, remembering, talking are modes or ways that persons are and act in the world. If substance dualism is true, we have a clear way of understanding that at biological death, the person qua individual is no more present, and the body is a corpse or remains. On some forms of materialism in which the person is his or her body, matters are less clear. In fact, it is more natural to think that insofar as after biological death, the body is still there, then you are still there even though your body is no longer functioning as you. You wind up as a passing mode of a body, something that I suggest is deeply counter-

intuitive. I will not be hanging my case of substance dualism on this distinction between a mode and a substance, but I will go on in Section 3.3 to defend the possibility of persons ceasing to be while their bodies survive, and persons surviving death despite the annihilation of their bodies. Let us take note here, however, of an important principle when it comes to recognizing the identity and difference of things: the principle of *the indiscernibility of identicals*, according to which if two (perhaps ostensibly) different things are the same, then whatever is true of one, is true of the other.

Take identity statements in which we affirm (for example) that *Samuel Clemens is Mark Twain* or *Water is H<sub>2</sub>O*. In these two identity claims, while a person may not know their truth and believe that when he is shaking hands with Samuel Clemens he is not shaking hands with Mark Twain, he is in fact shaking the same hand; similarly, even if someone does not know the atomic theory of matter, whenever she drinks water, she is drinking H<sub>2</sub>O. This principle will come into play in both Section 3.2 and Section 3.3, when I shall propose that there are truths about thinking, feelings, having emotions, sensations, and so on, that are not true of what physicalists identify as physical things and events such as brain processes or the brain itself.

### 3.2 The Primacy of the Mental

Many philosophers today assume that we have a clearheaded, problem-free concept of what is physical, and they assume that we do not have a very clear idea of what is the mental. I propose that matters are in fact otherwise, and necessarily so.

Consider two statements in which Daniel Dennett advances the dominant claim about the primacy of our grasp of what is physical.

There is only one sort of stuff, namely *matter* – the physical stuff of physics, chemistry, and physiology – and the mind is somehow nothing but a physical phenomenon. In short the mind is the brain . . . We can (in principle!) account for every mental phenomenon using the same physical principles, laws, and materials that suffice to explain radioactivity, continental drift, photosynthesis, reproduction, nutrition, and growth. (Dennett 1991, 33)

Dennett claims that, in contrast with the lucid, monolithic, scientific understanding of a mind-independent world, dualism is hopelessly antiscientific.

Dualism (the view that minds are composed of some nonphysical and utterly mysterious stuff) . . . [has] been relegated to the trash heap of history, along with alchemy and astrology. Unless you are also prepared to declare that the world is flat and the sun is a fiery chariot pulled by winged horses – unless, in other words, your defiance of modern science is quite complete – you won't find any place to stand and fight for those obsolete ideas. (Dennett 1996, 24)

What to make of these claims?

I propose that Dennett's position is self-undermining and confused. It is self-undermining to the extent that Dennett cannot presume to have any clearer understanding of nonmental physical phenomena than he does of concepts, reasons and reasoning, grasping entailment relations, reliance on experience and observations that go into the practice of the sciences, and the kind of reasoning that goes on at philosophy conferences. When he begins his credo, he writes as though we have a clear idea about

physical principles and the physical sciences, when in fact what he is appealing to is our *ideas (theories, thoughts, concepts)* of what is physical. To appeal to *physics, chemistry, and physiology* is (if it means anything at all) to appeal to what persons as scientists practice with their theories and observations, their conceiving of and intentionally undertaking experiments and making predictions, their recording data and drawing inferences. Nutrition and reproduction as events only come to be understood by us when we engage in understanding, thinking, reasoning, and so on. Stan Klein rightly observes the absurdity of treating the mental or experience as of secondary intelligibility compared to a mind-independent concept of the physical.

According subjectivity, at best, “second class citizenship” in the study of mind is particularly ironic in virtue of the fact that subjectivity is the very thing that makes the scientific pursuit of such knowledge (actually any knowledge) possible. Timing devices, neuroimaging technologies, electroencephalographs, and a host of modern means of obtaining objective knowledge of minds are useless absent an experiencing subject . . . To believe otherwise has the absurd consequence of rendering our knowledge of mind (or more generally, of reality) dependent, in its entirety, on the provisions of an experiential conduit stipulated either to be unworthy of study or essentially nonexistent. (Klein 2015, 43–44)

I suggest that Dennett’s view of selves is far more preposterous than the examples he gives in mordant depiction of dualists (his blustering reference to dualism being on a par with astrology, alchemy, and the belief that the sun is a chariot) because Dennett famously denies that selves are real; we are instead fictive, narrative centers of gravity (Dennett 1986). I have developed this critique elsewhere (Taliaferro and Evans 2011 and Taliaferro 1994) and will summarize my proposal: It is Dennett who is profoundly antiscientific because he denies the belief that scientists (as persons who are real and nonfictive) exist. To be intelligible, Dennett must treat as nonfictive (i.e., treat as real) that he and we and scientists truly exist as thinking, feeling, reasoning beings, and yet, in the final analysis, according to Dennett, none of us is fundamentally real. The following observation by S. Gallagher and D. Zahavi may seem so obvious, it is regrettable that it is necessary to put in print, but it is a truth that Dennett’s form of materialism drives him to deny or obscure: “Science is performed by somebody; it is a specific theoretical stance toward the world . . . scientific objectivity is something we strive for but it rests on the observations of individuals” (Gallagher and Zahavi 2008, 41).

Dennett complains about dualists’ positing nonphysical stuff. But consider first whether we have a clear idea of *physical stuff*? Does physical stuff have color, taste, and smell? Does it make sounds? These questions take us to the classic, modern problem of conceiving of physical objects that are truly independent of minds. I do not embrace the all-out-subjectivist idealist claim that it is impossible to conceive of something that is ontologically independent of (or not constituted by) minds (though on theistic grounds, I am convinced that no physical object can exist without dependence on the causal power of the divine mind). But we should appreciate that *all our appeals to physical objects are conceptually mediated* (your thinking about any physical object necessarily involves the exercise of your conceptual powers) and as we examine the concepts themselves we face a host of difficulties in terms of identifying the necessary and sufficient conditions for being physical. A commonplace proposal, for example, is that *an object is physical if it is spatiotemporal*, but there are many philosophers from the Cambridge Platonists to G. E.

Moore, H. H. Price, and Howard Robinson who recognize sensory images, after-images, and dream-images as spatiotemporal but they are not in the physical space(s) as described in the physical sciences. For reasons outlined above (and so to repeat), identifying the physical in terms of successful or ideal results of the physical sciences does not help us get to a mind-independent notion of the physical because the physical sciences involve the mental.

Stepping back a bit, I suggest we are in an epistemic position the very opposite of Dennett's and those like him who contend that we have a clear concept of what is physical but not of what is mental, and this includes matters of causation. In Dennett's examples, causation occurs without mental phenomena (radioactivity, continental drift, etc., involves no beliefs, ideas, thoughts, etc.), but we necessarily have a clearer concept of mental causation than we have of mind-independent physical causation. You can form no *understanding* of radioactivity unless you understand *ideas* of particles and what is involved with particles being emitted from nuclei, *ideas* about many nuclear isotopes and how *those ideas* give rise to *ideas* about different forces. In reasoning about radioactivity, and in fact in reasoning of any sort (deductive, inductive, in abduction or basic perception), we readily grasp the entailments and inferences in which *we believe some things in virtue of (because of, in a causal sense) believing others*. I believe that six is the smallest perfect number because I grasp that it is the smallest number equal to the sum of its divisors including one, but not including itself ( $6 = 1 + 2 + 3$ ). Mathematical reasoning would not make any sense without such mental causation as well as our reasoning about virtually indefinitely many subjects: it is because I believe you are drinking coffee and discussing philosophy that I conclude it is false to deny you are drinking coffee and discussing philosophy, *ad infinitum*.

In light of the above, let us consider what some philosophers claim is a fatal problem for substance dualism. Philosophers going back to Descartes's lifetime thought it highly problematic to explain how the soul (or mind or person as a nonphysical thing) causally interacts with what is physical. I suggest matters are different. We have a clear grasp of mental causation and in exercising our mental powers we may readily grasp that there is causal interaction between our bodily states and mental lives. We do not, however, have a lucid understanding of physical causation, and we have only a wobbly concept of it; rather, we have a clearer understanding of mental causation and such mental causation is essential for us to even begin to understand what is involved in physical causation. Consider, for example, the extant theories of causation on offer today: Humean association, counterfactual accounts, causation as a matter of timeless laws involving states of affairs, and (my preferred account) causation as dependent on the concrete, basic powers and liabilities of fundamental objects. None of these are understandable or usable unless we have confidence in mental causation. To understand a thrown baseball breaking a window, you need to employ (or not deny) first-order logic, to grasp basic entailments (if the baseball is smaller than the window, the window is larger than the baseball), and to engage in basic matters of identity and difference. One might object that all kinds of animals can grasp causal relations without such mental apparatus. I believe we can be agnostic about the extensiveness of mental life of nonhuman animals and propose that the primacy of the mental is evident in cases where persons engage in explicit reflection on what there is; when persons engage in thinking about why things occur and explanations of radioactivity, for example, and of ourselves (what accounts for our thinking, and so on).

The position advanced here is in the tradition of John Locke and Thomas Reid. Consider this passage from chapter 21 (“Of Power”), section 4 of Locke’s *An Essay Concerning Human Understanding*:

Bodies by our senses do not afford us so clear and distinct an idea of *active power*, as we have from reflection on the operations of our minds . . . Neither have we from body any idea of the beginning of motion . . . The idea of the beginning of motion, we have only from reflection on what passes in ourselves, where we find by experience, that barely by willing it, barely by a thought of the mind, we can move the parts of our bodies, which were before at rest. (Locke 2004, 221)

In terms of other historical precedents, the primacy of the mental was robustly defended by Bertrand Russell in his *The Analysis of Matter*: “As regards the world in general, both physical and mental, everything that we know of its intrinsic character is derived from the mental” (Russell 1954, 402). For further reflections on rationality and mental causation, see Angus Menuge’s *Agents Under Fire* (2004) and Peter Unger’s *All The Power In The World* (2006).

So far, we have not secured dualism or the nonidentity of the mental and physical. Some philosophers who insist on the open-ended, indeterminate nature of the physical world, such as Galen Strawson and Noam Chomsky are not friends of dualism, but their insistence on our lack of a clear concept of what is physical should give pause to those represented by Dennett who assume that we have a problem-free concept of the physical. Chomsky writes that “The notion of ‘the physical world’ is open and evolving” (Chomsky 1980, 5). Much earlier, Bertrand Russell observed that contemporary physics has blown away a “common sense” grasp of matter: “Matter has become as ghostly as anything in a spiritualist séance” (Russell 1927, 78). I propose that the way forward is to consider whether there is reason to believe that what we know in our experience (our thoughts, feelings, sensing, perceiving, remembering, reasoning, etc.) is the very same thing as what most physicalists claim is physical: our brains and brain processes or our bodies as a whole. Setting to one side the difficulties of distinguishing the mental and nonmental properties of the physical (is it a physical property of my body that it has a certain color and smell, and so on?), are there reasons for identifying our thinking with such material states and things? In what follows, I develop what is called *the knowledge argument* with assistance from the principle of the indiscernibility of identicals.

The knowledge argument has a long history, going back at least to Goethe, but in recent times it is principally associated with T. L. S. Sprigge, Thomas Nagel, and Frank Jackson. Here is Sprigge’s succinct development of the position:

The main reason for holding [that there is a distinction between the mental and the physical] is that it seems entirely possible that a scientist should have complete knowledge of a human organism as a physical system and yet be ignorant of the special character of that individual’s consciousness. (Sprigge 2011, 9)

Sprigge presses his point further in terms of our experience of other persons:

For that matter, there is nothing physical about another person, which absolutely proves that he is conscious. His consciousness is not something which could be located in his brain for everything about the brain could be as it is without the individual being conscious. (Sprigge 2011, 9)

I think Sprigge is correct. On this point, there is an interesting accord between some radical physicalists and dualists. For example, Dennett observes that “the problem with brains is that when you look in them, you discover that there is nobody home” (Dennett 1991, 29). Paul Churchland even sets up what he regards as a problem for dualism by asking us to imagine a neuroscientist who has fully explained the brain without bringing in the mental, and is then confused about how the mental might come into play (Churchland 1988). In light of the earlier case I developed above for the primacy of the mental, I suggest that the last thing Dennett and Churchland should question is the existence of the self and the mental. If a person does not see the person in the physical world (or the brain), the most reasonable inference is that the person is not physical or identical with what is assumed to be physical. In Churchland’s case, it seems that for the neuroscientist to be puzzled is itself for her to be engaged in mental reflection. If she is convinced she is puzzled and is puzzled by not finding the puzzlement in her view of the brain, shouldn’t she be convinced that the puzzlement is not identical with the part or whole of the brain? Banishing the existence of puzzlement and other mental states because of their not fitting into one’s thinking about the mind and brain seems too high a cost. The absurdity of denying the existence of our thinking, feelings, sensing, and so on, led A. J. Ayer to comment that such philosophers must pretend to be anesthetized to take their views seriously (cited in Lund 2005, 11). Though the situation is even worse: pretending requires using the imagination, thoughts and concepts. I will not speculate how one might pretend to not pretend.

Let us consider the most common objection to the knowledge argument. It has been objected that the argument only establishes a kind of conceptual or epistemic dualism, but not one that provides evidence that there are (in reality) two nonidentical properties or things – being mental and being physical or, more precisely, being a brain state or process or being a brain or a material body. Could it be that the knowledge argument only gives us reason to believe that there are two different ways of knowing (or being aware of or conceiving of) something? Perhaps the way I know my brain states is by two routes: by having sensations, thoughts, and so on (on the one hand) or through observing brain states or understanding the body in terms of anatomy or through brain scans (on the other hand)? After all, you could know the same person as Samuel Clemens or Mark Twain or know water in different ways and yet you would not thereby be entitled to conclude there are two authors or two elements.

The problem with this objection is that it denies or obscures the very nature of experience and the mental. When we are thinking, feeling, and sensing, we are not simply taking up a point of view on something or other; the thinking, feeling, and sensing is the point of view. Thinking, and so on, is an activity of subjects and the activity and content of the thinking (ideas, concepts, and so on) is something intrinsic to the mental process. Thinking cannot take place without thinking, but (back to Sprigge’s version of the knowledge argument) we need reasons to infer that when we observe bodily events (brains or brain processes) we are observing the physical correlates of mental experiences. The importance of this point can be underscored by noting that in the case of Mark Twain, we can readily reconcile how different modes of access to the author can give rise to mistakenly thinking Twain is distinct from Clemens. We can see that the claim “Water is H<sub>2</sub>O” is a matter of material composition (the colorless, odorless liquid we call “water” is composed of hydrogen and oxygen). But there is no such reconciliation when it comes to identifying mental, experiential states and physical, bodily states. For Churchland’s neuroscientist, to

observe her patient's brain is not to observe mental states, which is why (in Churchland's narrative) she is puzzled.

The clinical neuroscientist and philosopher, Raymond Tallis, summarizes his frustration with identity materialism:

The most obvious trouble with the view that neural activity on the one hand, and experience on the other, are the same thing is that they should appear like one another. But nothing could be further from the truth. The colour yellow or more precisely the experience of the colour yellow, and the neural activity in the relevant part of the visual cortex, however it is presented, look not in the slightest bit similar. There is nothing yellow about the nerve impulses and nothing nerve-impulse-like about yellow. If however, they were the same thing, the least one might expect is that it would appear as if they were the same thing. Surely it is not too much to expect that something should look like itself. (Tallis 2011, 85)

In my view, Tallis rightly points out how those materialists who charge that the mental is only the physical as viewed from a certain angle do not at all make it plausible that there is no observing subject because the very idea of viewing the physical from a certain angle only makes sense if there are conscious, observing subjects. In the following passage Tallis observes the lack of any apparent identity between the mental and physical and then goes on to make a broader point about the stubborn, ineliminable role of the conscious subject.

The most fundamental and obvious problem is one that we have touched on already: namely, that nerve impulses *are not at all like qualia*. Those impulses in the visual cortex do not look like, say, the colour or shape or size of my red hat. We have seen how some philosophers have tried to deal with this suggestion that what we see on a brain scan or an EEG is only one aspect of the neural activity and that consciousness is another aspect. This does not make the identity between neural activity and conscious experience any more plausible because the very notion of "aspects" presupposes consciousness: an observer looking at something from a particular angle or in a particular way (as when it is examined through the lens of instruments, concepts and theories). (Tallis 2011, 95)

Added to this, note the oddity of a materialism which seeks to eliminate the mental but retain what it calls "aspects." Mind-independent physical objects have sides; think, for example of a barn having two sides. But to suppose that in addition to *sides* there are *aspects*, seems to involve positing points of view, and it is hard to conceive of a point of view without positing a subject or sensing thing that takes up (or can take up) that point of view.

The objection we are considering here (the denial of subjective experience) is sometimes articulated in a way that I suggest is implicitly self-refuting. It has sometimes been argued that it only appears that there are appearances. Or, alternatively, it has been suggested we wrongly experience (or we experience an illusion) that there are experiences. But if there appear to be appearances, there are appearances, and if we have wrongful or illusory experiences, then there are experiences. The absurdity of Dennett's position may be likened to someone going into a house and, not seeing anyone (else) in the house, concluding that he is not in the house.

Because of the importance of recognizing the primacy of the mental for my project, let us linger to consider just how far some materialists will go to deny what seems experientially evident. Paul Churchland likens those who, like me, argue against eliminating the mental



through appealing to what seems experientially evident in our subjective thinking, feeling, and so on, to those employing faulty, outdated conceptual frameworks. Churchland writes:

The eliminative materialist will reply that that argument makes the same mistake that an ancient or medieval person would be making if he insisted that he could just see with his own eyes that the heavens form a turning sphere, or that witches exist. The fact is, all observation occurs within some system of concepts, and our observation judgments are only as good as the conceptual framework in which they are expressed. In all three cases – the starry sphere, witches, and the familiar mental states – precisely what is challenged is the integrity of the background conceptual frameworks in which the observation judgments are expressed. To insist on the validity of one's experiences, traditionally interpreted, is therefore to beg the very question at issue. For in all three cases, the question is whether we should *reconceive* the nature of some familiar observational domain. (Churchland 1988, 47–48)

First, note that insofar as Churchland acknowledges the very existence of conceptual frameworks, he is implicitly acknowledging the importance, and perhaps even the primacy, of the mental. But second, and more importantly, consider whether it is plausible to liken our experiences of sensory states, such as searing pain, to our observation of remote objects like the sun or someone's judging that they are seeing a witch. Arguably, these are utterly different. In the case of experiencing pain, we are experiencing something we can conceptually reflect on, but the pain itself is a datum, something that is not itself a concept. Feeling pain is not feeling a concept. William Vallicella (2013) offers this response to Churchland:

Suppose I feel a pain. I might conceptualize it as tooth-ache pain in which case I assign it some such cause as a process of decay in a tooth. But I can "bracket" or suspend that conceptualization and consider the pain in its purely qualitative, felt, character. It is then nothing more than a sensory quale. I might even go so far as to abstract from its painfulness. This quale, precisely as I experience it, is nothing like a distant object that I conceptualize as this or that.

Now the existence of this rock-bottom sensory datum is indubitable and refutes the eliminativist claim. For this datum is not a product of conceptualization, but is something that is the "raw material" of conceptualization. The felt pain *qua* felt is not an object of observation, something external to the observer, but an *Erlebnis*, something I live-through (*erleben*). It is not something outside of me that I subsume under a concept, but a content . . . of my consciousness. I live my pain, I don't observe it. It is not a *product* of conceptualization – in the way a distant light in the sky can be variously conceptualized as a planet, natural satellite, artificial satellite, star, double-star, UFO, etc. – but a *matter for* conceptualization. So the answer to Churchland is as follows. There can be no question of reconceptualizing fundamental sensory data since there was no conceptualization to start with. So I am not begging the question against Churchland when I insist that pains exist: I am not assuming that the "traditional conceptualization" is the correct one. I am denying his presupposition, namely, that there is conceptualization in a case like this.

I would only add that I suggest that the radical move to eliminate the evident reality of the mental is even more radical than overturning someone's claim to see that the sun circles the earth or there are witches. It is more akin to challenging someone's claim to have evident experiences of thinking, observing, feeling, and so on. As Richard Fumerton notes in his brilliant book *Knowledge, Thought, and the Case for Dualism*: "Acquaintance is a real relation and its obtaining is contemporaneous with the existence of its object. One can't use

a razor [as in Ockham's razor] to slice away from one's ontology one's searing pain" (Fumerton 2013, 247).<sup>1</sup>

Let us now consider a more substantial form of dualism. Someone might grant that there are good reasons to recognize the reality of the mental and to not identify the mental and the physical and yet to believe that persons themselves are physical or material bodies. Why go further to believe that persons are themselves not identical with their bodies?

### 3.3 A Modal Argument for Substance Dualism

Most substance dualists are committed to believing that the person is a substantial individual who endures over time as the self-same subject. Most of us (who are substance dualists) contend that we are aware of ourselves as subjects over time, and it is this self-awareness that grounds (or forms the foundation) for our identifying the objects around us. Think of the apparent difficulty of denying this. How might you identify that you have heard Big Ben ring three times unless you were aware of yourself being the same individual over time who heard Big Ben ring twice earlier? Even to point to your body ("Look at this body" you say, pointing) would be puzzling without presupposing your self-awareness of someone doing the pointing. (For an extensive, excellent defense of the primacy of our self-awareness, see David Lund's *The Conscious Self*.)

Some philosophers have built a case for substance dualism on the grounds of our realization of our endurance as the self-same person over time. Our physical bodies are changing continuously with the loss and renewing of cells. Given the indiscernibility of identicals, it may be argued that while it is strictly speaking true that you are the same person today who, decades ago, attended the first grade, it would be false to claim that the body you have now (or the body you are) is, strictly speaking, the same as the body you had then.

I actually think the above reasoning is sound, but it rests on a controversial account of identity over time, *mereological essentialism*, according to which every whole thing consisting of parts, has those parts essentially. In light of that severe account, my physical body late at night is not, strictly speaking, the same thing as my body earlier today because of its lacking and gaining parts. I find it reasonable to adhere to such a strict view of identity, while allowing for a "loose and popular" sense of identity, and to deny accounts of "relative identity" that might make it (more or less) the case that you are kind of identical both personally and physically with you as a first grader and kind of not. But because of the murkiness of these philosophical waters, I develop here a different reason for adopting substance dualism. What follows is called a modal argument, as "modal" refers to what is possible.

The argument has three premises. Premise one: if "A" is "B," then whatever is true of "A," is true of "B." This is our "friend," the principle of the indiscernibility of identicals. In order to avoid some possible counter-examples, in this context, let us add that we are employing the principle according to which "A" and "B" do not refer to things in an abstract fashion using titles or general descriptions, but rather "A" and "B" strictly refer to individuals: they are what some philosophers call "rigid designators." Rigid designators are like the names "Mark Twain" and "Samuel Clemens" as opposed to the general title "The author of Tom Sawyer and Huckleberry Finn." The reason for this distinction is because we can imagine that, while Mark Twain did write that book, the book might not have been written or was

written by someone else. Under those conditions we face the muddle of thinking that there could be something true of Mark Twain that is not true of “The author of Tom Sawyer and Huckleberry Finn.” The latter could have been someone else, but presumably Mark Twain could not have been someone other than Mark Twain. Let us move to the second premise, ensuring we are using terms that are strictly referential.

Premise two: let “A” refer to myself, and “B” refer to my body as a whole or some part of me, like my brain. Given premise one, whatever is true of myself, is true of my body as a whole or some part of me, like my brain.

Premise three: but it is reasonable to believe that there are truths about me, not true of my body or a part of my body like my brain. I can exist without the later, and the later can exist without me. Why think this is true? One response is to appeal to what we think it reasonable to believe actually occurs. A plausible case might be our experience (noted earlier) of how a person can cease to be present and yet their body is present. If we are prepared to accept reports of out of the body experiences this might be another means of support (see the discussion in Hasker and Taliaferro 2014). Alternatively, we may appeal to thought experiments in which it appears that we can consistently imagine and describe a person surviving without their body and vice versa. More on this below.

Conclusion: if the premises are right, then the conclusion that the person is distinct from the body is reasonable.

I offer further reasons for accepting this argument in the course of replying to the following, three objections: support for premise three of the thought experiments is illicit; the argument is circular; the conclusion is subject to a host of independent objections that makes it unacceptable. The latter objection is the most general and can amount to a partial concession and then an independent objection: someone might grant that the premises offer some reason for accepting the conclusion, but contend that the conclusion is defeated because we have independent reasons for thinking the conclusion is false.

*Objection 1:* The imagination is not a plausible guide to recognizing what is possible. Philosophers have objected that imagining persons without their bodies or by switching bodies is no reason to think that such states of affairs are actually possible. Arguably, we can imagine all kinds of things (time travel) that turn out not to be *bona fide* possibilities.

*Reply:* I concede that sometimes we may appear to imagine a state of affairs obtaining that then turns out not to be possible. However, I have argued extensively in many places that our seeming to conceive of a state of affairs obtaining does provide *prima facie* reasons for thinking that the state of affairs is actually possible (see Taliaferro 1994, 2001b, 2012a, 2012b). Here I would like to modify such a principle in accord with David Lund’s notion of *secunda facie* (second appearance) reasons derived by conceiving of states of affairs with great scrutiny, considering possible defeaters, and so on. Lund offers these observations in his development of a dualist modal argument.

Though ideal conceivability would yield conclusive knowledge of metaphysical possibility, the ideal form is apparently beyond our reach, at least in the case of any matter as complex as one’s possible disembodiment. On the other hand, we should strive to attain more than *prima facie* conceivability, for it may be quite vulnerable to defeaters revealed by a more detailed conception and better reasoning. *Secunda facie* conceivability, by contrast, survives an informed and painstaking search for possible defeaters. Though the likelihood of defeaters is not eliminated, it is greatly reduced. Thus we must acknowledge *secunda facie* conceivability to be an extremely reliable guide to nonactual possibility unless we are willing to deny that any

form or level of conceivability provides any epistemic access to such possibility. But the implications of such a denial would be intolerable if, as seems clear, conceivability is our only basic access to nonactual possibility. (Lund 2005, 264–265)

While Lund's modal argument appeals to disembodiment (the existence of persons without bodies) as nonactual, we may well take note that probably the majority of the world population believes in an actual afterlife in which persons survive the death of their bodies. Allowing that such a belief is false or probably false, is it reasonable to think that such large numbers of persons are entertaining something metaphysically impossible or utterly inconceivable?

Because of the importance of this premise, the conceivability of person and body separation, let us consider further what might be offered to someone who claims that they simply cannot form any concept of person–body separation. I offer two additional observations.

First, as pointed out earlier, we currently lack a clear understanding of what it is to be a physical body or event. Those who embrace what they think of as a common sense understanding of physical bodies (and profess puzzlement over what is not physical), need to take seriously the challenge to their position from the standpoint of contemporary physics. Anthony Kenny, probably the greatest living historian of philosophy, observes:

At one time it seemed as if a robust and substantive naturalism could be easily stated. This was a conception that thought of the world as being made up of solid, inert, impenetrable and conserved matter – a matter that interacts deterministically and through contact. But twentieth-century physics posited entities and interactions that did not fit the materialist characterization of reality, and which took science far away from a world of solid, inert, massy material atoms. (Kenny 2013)

I cite two more authorities on this. Consider Michel Bitbol's claim: "*material bodies are no longer the basic objects of physics . . .* Ironically, the notion of material body motivated the very research that eventually dissolved it" (Bitbol 2007, 243; emphasis in original). Tim Crane and D. H. Mellor write: "The 'matter' of modern physics is not at all solid, or inert, or impenetrable, or conserved, and it interacts indeterministically and arguably sometimes at a distance. Faced with these discoveries, modern materialism's modern descendants have understandably lost their metaphysical nerve" (Crane and Mellor 1990, 186). Philosophers who claim to have a clear grasp of the physical world are indeed open to challenge.

Second, the modal argument only needs there to be truths about persons that are not truths about their bodies. In reply to someone who claims not to be able to conceive of persons without bodies, the modal argument can be redeployed with the modal claim that a person's relationship with their body is contingent – the claim, for example, that while a person is (for example) embodied as a white male, living in the United States, that person could have been embodied as black female, living in Africa. Arguably, in an ethical application of the Golden Rule, we routinely seek to conceive of what the world would be like if we had had different parentage. I suggest that for a white male to protest that he cannot imagine being black or being of a different gender, there may be a failure of moral imagination at work.

In concluding these replies, I propose that in the absence of a positive grasp of a person's necessary relationship to (or essential identity with) his body, the apparent conceivability of

having a different body is strong evidence that the person–body relationship is contingent. I suggest that in work by Colin McGinn we find a representative case of when there is (what I believe to be) an undue resistance to the modal argument. McGinn acknowledges the apparent contingency of the relation between persons and their bodies, but he nonetheless persists in believing there is a necessary tie between persons and bodies even though that tie seems mysterious and inconceivable:

The brain has physical properties we cannot grasp, and variations in these correlated with changes in consciousness, but we cannot draw the veil that conceals the manner of their connection. Not grasping the nature of the connection, it strikes us as deeply contingent; we cannot make the assertion of a necessary connection intelligible to ourselves. There may then be a real necessary connection; it is just that it will always strike us curiously brute and unspicuous. (McGinn 1991, 20)

I believe David Lund is absolutely right in this reply to McGinn:

The mere possibility of there being a necessary connection that is unknown to us provides no genuine reason to mistrust our intuition of contingency, especially if we are willing to grant that our modal intuitions are presumptively reliable. My being told that my intuition of contingency might be due to my ignorance of what might in fact be, for all I know, a noncontingent connection is not sufficient grounds for mistrusting my intuition, even though it is, of course, defeasible. (Lund 2005, 301)

For an excellent, further, critical treatment of how to deploy thought experiments, see Loose (2012).

*Objection II:* Some philosophers object to the modal argument on the grounds that it is circular. Unless one already believes that persons are not identical with their bodies, one would not accept the premise that persons can exist without their bodies and vice versa. In other words, the argument might effectively illustrate the implications of substance dualism, but it would not provide independent reasons for a nonsubstance dualist to accept the conclusion.

*Reply:* One effective reply to this objection is that there are philosophers who actually do recognize the possibility of persons existing without their bodies and yet reject substance dualism. Lynne Baker, David Lewis, and D. M. Armstrong (among others) have accepted as a possibility that persons can exist without their bodies but they provide reasons for why this would not entail substance dualism. So, Baker thinks that persons are neither identical with nor substantially other than their bodies, but persons are, rather, composed of their bodies. Lewis has proposed that when we think we are imagining a person existing without their bodies, we are imagining, not the person himself existing without his body, but his counterpart in another possible world. Armstrong contends that the ostensible imagining of persons disembodied is imagining a general, what philosophers call *de dicto*, state of affairs, and not a matter of imagining in a direct fashion you or I disembodied. It would take a book to fully explore these possible countermoves, but I hope that simply sketching the way some philosophers grant the datum (it appears that we can imagine persons without their bodies and vice versa) is some reason to think the modal argument is not circular. I argue extensively in *Consciousness and the Mind of God* that these alternatives are not as plausible as substance dualism (see also Lund 2005, for an outstanding analysis of this objection and a reply).

A different reply can be mounted by taking into account how the modal argument is akin to many similar thought experiments employed in other domains of philosophy without facing deep problems of circularity.

*Objection III:* Could it be that the modal argument has some plausibility but it leads to a conclusion that we have independent reasons for thinking is false? This predicament is not especially peculiar. In political theory, one might concede that political liberalism is supported by the kind of methodology that John Rawls introduces (appeals to a “veil of ignorance”), but contend that there are individual rights and entitlements of an historical nature that undermine that methodology.

In replying to this objection I report on the breadth and depth of the growing number of works defending substance dualism. All the major independent objections to dualism have been addressed by a growing number of philosophers including (in a partial list) by Mark Baker, George Bealer, Robin Collins, Frank Dilley, C. Stephen Evans, Joshua Farris, John Foster, Stewart Goetz, Rich Halverson, W. D. Hart, William Hasker, E. J. Lowe, John Lucas, Geoffrey Maddel, J. P. Moreland, Alvin Plantinga, Daniel Robinson, Howard Robinson, Richard Swinburne, Peter Unger, Keith Yandell, Dean Zimmerman, and others. For the record, this list of defenders of substance dualism should probably be extended to earlier mid-twentieth-century immanent defenders which include some of my all-time favorite philosophers: C. A. Campbell, J. Eccles, A. C. Ewing, H. D. Lewis, Karl Popper, and A. E. Taylor. In *Consciousness and the Mind of God*, I both defend a version of the modal argument as well as address a host of independent objections to dualism, including the private language argument, the problems of individuation, and the causal pairing argument. In my view, the problems that are advanced against substance dualism are based on caricatures of dualism (as found in the supremely misleading work *The Concept of Mind* by Gilbert Ryle) or the objection stems from illicit or question-begging principles (objecting to mental–physical causation on the grounds that it is different from exclusively physical causation) or the problems dualism faces are problems that are faced by the most plausible alternatives to dualism. The latter is borne out in the title of a paper which I highly recommend: “What is a Problem for All Is a Problem for None: Substance Dualism, Physicalism, and the Mind–Body Problem” (Himma 2005).

In concluding, I highlight what I find to be the least impressive, independent objections to substance dualism. First, consider the claim that if substance dualism is true we should not expect there to be pervasive, wide respects in which our mental life is profoundly causally dependent on brain states and activity. I have even attended a presentation at New York University where the professor claimed in a lecture to undergraduates that substance dualism cannot account for why it is that alcohol is intoxicating. Substance dualism, as articulated here, is committed to a full view of embodiment in which the well-being of persons is profoundly impacted by the body. Profound correlation of the mental and physical is not the same as identity and, indeed, if the knowledge argument is successful, all the correlations imaginable would not add the slightest reasons to compromise one’s confidence in substance dualism.

Second, the claim that substance dualism might impede scientific progress in our understanding of human life or neurology is baseless. There is not, in my view, a single experiment in neuroscience that gains in any scientific prowess if one claims the mental is numerically identical with the physical versus making the claim that the mental is functionally identical with the physical (this is essentially, the position outlined above

about how, in a healthy human embodiment the person is functionally a unity of person and body).

Fortunately, from the standpoint of substance dualism, some contemporary materialists have come to see that the objections to it are not forceful. In an important paper, "Giving Dualism Its Due," William Lycan, a thorough going materialist, makes this concession:

My position (materialism) may be rational, broadly speaking, but not because the arguments favor it: Though the arguments for dualism do (indeed) fail, so do the arguments for materialism. And the standard objections to dualism are not very convincing; if one really manages to be a dualist in the first place, one should not be much impressed by them. My purpose in this paper is to hold my own feet to the fire and admit that I do not proportion my belief to the evidence. (Lycan 2009, 551)

No religious reasons for adopting substance dualism have been introduced in making my case for substance dualism in this chapter. But in closing, I note how some religious considerations may have a positive role. Philosophers who are theists and yet who object to substance dualism on the grounds that it is incoherent to believe that there can be a nonphysical or incorporeal soul, face the embarrassment that in theism they accept the coherence of a nonphysical or incorporeal reality. And Christians who find a physicalist view of human beings attractive will find it difficult to embrace a traditional account of the incarnation. I suggest that it is metaphysically incoherent to believe that a nonphysical Person (the second Person of the Trinity) would come to be numerically identical with an exclusively physical body. Identity relations are transitive, and supposing that the second Person of the Trinity is a physical body would be to suppose that this body has all the divine attributes (being everlasting, existing necessarily, and so on), something that appears conceptually (and religiously) absurd. Once one realizes that substance dualism can fully acknowledge *the goodness of embodiment*, a theological appreciation of *the goodness of the incarnation* is secured without the risk of extravagant, and deeply problematic identity claims.

## Note

1. I add a further point that is argued for in detail by Scott Smith in his excellent book *Naturalism and Our Knowledge of Reality* (2012). Unless there is some access to nonconceptual content, there is no way to show that our concepts are grounded in the nature of things, and some form of antirealism (conventionalism, fictionalism, or postmodern constructivism) is going to be difficult to avoid. If all we have are conceptual frameworks and no way to independently test them, then we end up losing the kind of objectivity that naturalists prize. Smith plausibly argues that for naturalists to secure the objectivity of science, they need to acknowledge our nonconceptual experiences.

## References

- Bitbol, M. 2007. "Materialism, Stances, and Open-Mindedness." In *Images of Empiricism: Essays on Science and Stances, with a Reply from Bas C. van Fraassen*, edited by Bradley Monton, 229–270. Oxford: Oxford University Press.
- Chomsky, N. 1980. *Rules and Representations*. New York: Columbia University Press.
- Churchland, P. 1988. *Matter and Consciousness*. Cambridge, MA: MIT Press.
- Crane, T., and D. H. Mellor. 1990. "There is no Question of Physicalism." *Mind*, 99(394): 185–206.

- Dennett, D. 1986. "The Self as a Center of Narrative Gravity." *Philosophia*, 15: 275–288.
- Dennett, D. 1991. *Consciousness Explained*. New York: Little, Brown.
- Dennett, D. 1996. *Kinds of Minds*. New York: Basic Books.
- Fumerton, R. 2013. *Knowledge, Thought, and the Case for Dualism*. Cambridge: Cambridge University Press.
- Gallagher, S., and D. Zahavi. 2008. *The Phenomenological Mind*. London: Routledge.
- Hasker, W., and C. Taliaferro. 2014. "Afterlife." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. Accessed June 6, 2017. <https://plato.stanford.edu/entries/afterlife/>.
- Himma, K. E. 2005. "What Is a Problem for All Is a Problem for None: Substance Dualism, Physicalism, and the Mind–Body Problem." *American Philosophical Quarterly*, 42(2): 81–92.
- Hoffman, J., and G. Rosenkrantz. 1997. *Substance: Its Nature and Existence*. London: Routledge.
- Kenny, A. 2013. "Mere C. S. Lewis." *Times Literary Supplement*, June 19. Accessed June 6, 2017. <http://www.the-tls.co.uk/articles/public/mere-c-s-lewis/>.
- Klein, S. 2015. "A Defense of Experiential Realism: The Need to Take Phenomenological Reality on its Own Terms in the Study of Mind." *Psychology of Consciousness: Theory, Research, and Practice*, 2 (1): 41–56.
- Locke, J. 2004. *An Essay Concerning Human Understanding*, edited by Roger Woolhouse. New York: Penguin Books.
- Loose, J. 2012. "Constitution and the Falling Elevator: The Continuing Incompatibility of Materialism and Resurrection Belief." *Philosophia Christi*, 14(2): 439–450.
- Lund, D. 2005. *The Conscious Self; The Immaterial Center of Subjective States*. New York: Humanity Books.
- Lycan, W. 2009. "Giving Dualism Its Due." *Australasian Journal of Philosophy*, 87(4): 551–563.
- McGinn, C. 1991. *The Problem of Consciousness*. Oxford: Blackwell.
- Menuge, A. 2004. *Agents under Fire: Materialism and the Rationality of Science*. Lanham, MD: Rowan & Littlefield.
- Merricks, T. 2007. "The Word Made Flesh: Dualism, Physicalism and the Incarnation." In *Persons: Human and Divine*, edited by Peter van Inwagen and Dean Zimmerman, 281–300. Oxford: Oxford University Press.
- Russell, B. 1927. *An Outline of Philosophy*. London: Routledge.
- Russell, B. 1954. *The Analysis of Matter*. New York: Dover Publications.
- Ryle, G. 1949. *The Concept of Mind*. Chicago: University of Chicago Press.
- Smith, S. 2012. *Naturalism and Our Knowledge of Reality*. Farnham, UK: Ashgate.
- Sprigge, T. L. S. 2011. *The Importance of Subjectivity*. Oxford: Clarendon Press.
- Taliaferro, C. 1994. *Consciousness and the Mind of God*. Cambridge: Cambridge University Press.
- Taliaferro, C. 2001a. "The Virtues of Embodiment." *Philosophy*, 76(1): 111–125.
- Taliaferro, C. 2001b. "Sensibility and Possibilia: A Defense of Thought Experiments." *Philosophia Christi*, 3(2): 403–420.
- Taliaferro, C. 2005. *Evidence and Faith: Philosophy and Religion Since the Seventeenth Century*. Cambridge: Cambridge University Press.
- Taliaferro, C. 2012a. *The Golden Cord: A Short Book on Eternity*. Notre Dame, IN: University of Notre Dame Press.
- Taliaferro, C. 2012b. "Experimental Thoughts and Thought Experiments; Reflections on What Matters in Recent Works by Derek Parfit." *Philosophia Christi*, 14(1): 193–208.
- Taliaferro, C., and Jil Evans. 2011. *The Image in Mind*. London: Continuum Press.
- Tallis, R. 2011. *Aping Mankind*. Durham, UK: Acumen.
- Unger, P. 2006. *All the Power In The World*. Oxford: Oxford University Press.
- Vallicella, W. 2013. "Paul Churchland on Eliminative Materialism." *The Maverick Philosopher* [blog]. Accessed June 15, 2016. [http://maverickphilosopher.typepad.com/maverick\\_philosopher/2013/09/paul-churchland-on-eliminative-materialism.html](http://maverickphilosopher.typepad.com/maverick_philosopher/2013/09/paul-churchland-on-eliminative-materialism.html).



## Further Reading

- Goetz, S., and C. Taliaferro. 2011. *A Brief History of the Soul*. Oxford: Blackwell.
- Goetz, S., and C. Taliaferro. 2008. *Naturalism*. Grand Rapids, MI: Eerdmans.
- Moreland, J. P. 2008. *Consciousness and the Existence of God*. London: Routledge.
- Taliaferro, C. 2008. *Dialogues About God*. Lanham, MD: Rowman & Littlefield.
- Taliaferro, C. 2009. *Philosophy of Religion: A Beginners Guide*. Oxford: Oneworld Press.

# Debating Emergent Dualism

# The Case for Emergent Dualism

WILLIAM HASKER

In one sense, we all know what a human mind is. The mind is that by which we think, feel, make decisions, relate to other persons, and so on. But what is the *nature* of the mind? And how does the mind originate? These are important questions, questions that have given rise to impassioned disagreement. The three basic sorts of answers to our two questions go under the labels of *reductionism*, *creationism*, and *emergentism*. The core idea of reductionism is that more complex entities are built up, according to understandable laws and principles, from simpler entities combined and arranged in the right way. A key phrase for understanding reductionism is “nothing but”; the complex entities are “nothing but” the simple entities, properly arranged and interacting. There are many examples of reduction in the sciences. For example, much has been learned about the way in which the many different kinds of material in the world are built up out of 90 or so chemical elements, and these in turn from an even smaller number of elementary particles. As applied to the human mind or soul, the idea is that the mind is “nothing but” the complicated functioning of the neurons of which the brain is composed, operating according to the standard laws of physics and chemistry. Reductionism concerning the mind encounters many problems, as evidenced by a number of the chapters in this collection. In spite of strenuous efforts, no one has produced a convincing explanation of the way in which conscious experiences supposedly consist of arrangements of, and interactions between, the material particles in our brains and nervous systems. (This is called *ontological reductionism*, but it seems unlikely that it can be made to work.) *Causal reductionism*, in which the operations of our minds have their full and complete causal explanation in the physical interaction of neurons in our brains, is more popular, but it also encounters severe difficulties. For example, reductionism faces a fundamental difficulty in accounting for the evolution of the conscious mind. Reductionist views insist that the physical world cannot be affected by anything nonphysical, such as a thought or an emotion. All the causal work is done, according to these views, by the basic physical entities; this is known as the “causal closure of the physical domain.” Evolutionary selection, however, can operate only on physical structures and physical behavior, so if conscious thought has no effects in the physical realm it becomes

invisible to evolutionary selection, and cannot (except by accident) be improved by selection. To give a graphic example, materialist evolutionary theory may be able to give a plausible explanation of why, when a group of primitive hominids are threatened by a predator, they will remove to a safer location – say, by climbing a tree. But materialist evolution has no explanation whatever for the fact that their conscious mental state corresponds to “Let’s get out of here before that saber-tooth cat arrives,” rather than “Isn’t this a delicious meal of baboon meat?” An evolutionary theory that can give no account of the evolution of consciousness is defective at a fundamental level (see Hasker 2013, 1999).

Creationism, in contrast to reductionism, has proved to be quite appealing to religious believers seeking to understand the human mind or soul. Creationists concerning the soul hold that each individual human soul is directly and individually created by God; created *ex nihilo*, “out of nothing,” and not through any natural process. It seems natural to see this view as reflecting the individual, personal care which God manifests toward each and every human being; it also emphasizes the fundamental difference between humans and animals – or, we should more properly say, between humans and *other* animals. The common versions of creationism are generally modeled on the dualism of René Descartes, according to which body and mind are two radically different kinds of substances.<sup>1</sup> The body is physical through and through; it has all the natural, physical properties we are familiar with through the physical sciences, but in itself it has no trace of mental properties such as thought or even sensation. The mind is the “thinking thing”; its essential characteristic is consciousness and conscious thought, with everything that goes along with that. The mind, however, has no physical properties whatsoever: no electric charge, no mass, not even (in most versions) a location in space. But in spite of the radical difference between mind and body, the two do interact: the mind receives information from the body and brain, and the body in turn carries out the decisions made by the mind. The mind, according to creationism, is infused by God early in the development of the organism; it is united with the body throughout the course of a person’s life, but in no way does it owe its origin or its continuance to the body.

Creationism concerning the soul faces some fundamental difficulties. However, one widely publicized objection to this view is extremely overrated: I refer to the notorious “problem of mind-body interaction.” The objector asks, “How can the mind, which has no physical attributes whatsoever, make a difference to the functioning of the physical brain and body?” When no convincing reply to this “how” question is forthcoming, it is assumed that the credibility of dualism has been dealt a decisive blow. This, however, covertly assumes that we *do* understand causation in the body-to-body cases, as in the classic example of two billiard balls colliding and rebounding in different directions. But as David Hume showed long ago, this simply is not the case. We know that the balls *do in fact* act like this, rather than exploding, coming immediately to a dead stop, or passing through one another without resistance. And because we have witnessed this many times, it has come to seem “natural” to us, and we have the illusion that we understand what is going on. At bottom, however, we cannot claim to have rational insight into the *why*; we simply have to say, “that is just what happens.”<sup>2</sup> But surely, there are few other things that “just happen” as often (and without striking us as unusual or surprising) as the fact that our conscious experiences influence, and are influenced by, the state of our physical bodies. This objection, then, can be simply dismissed, in spite of the fact that many still take it seriously, even sophisticated philosophers who ought to know better.

But while the bare fact of mind-body interaction does not pose a serious problem for creationism, the *intimate and pervasive dependence* of mental processes on one’s bodily

condition is harder to reconcile with that view. It is easy to see that, if the body or nervous system are impaired, the mind might be lacking in sensory input, or might not be able effectively to control the muscles that produce movement. By why should *consciousness itself* be interrupted by a blow on the head, or by the action of an anesthetic? And why should damage to the brain produce major alterations (as sometimes happens) to a person's character and personality? Creationists have yet to produce a convincing explanation for this pervasive dependence of mind upon body, a dependence which from their point of view was hardly to be expected.

Further difficulties arise when we consider the relationship of humans to nonhuman animals. If human bodies, as such, are unable to have thoughts, sensations, or any other kinds of conscious experiences, the same must be true of animals. Descartes himself had a simple solution to this problem: animals in fact *do not have* sensations or any other kind of conscious experiences; they are mere automata, and it is simply an illusion that, when you come home and your dog jumps up and wags his tail, he is happy about your return. As I tell my students, if you can believe this you can believe anything! Clearly we cannot follow Descartes in this, so the creationist is forced to attribute divinely created souls to animals – not necessarily souls in every respect equivalent to human souls, but souls that are adequate to the sort of thoughts and experiences the animals do seem to have. But now a problem arises: how far do we carry this? How far down the “scale of life” do we find these divinely created souls? Of course, it is difficult for anyone to say with confidence just which simpler life-forms have conscious experience, and which do not. The problem, however, is that any answer the creationist can give tends to be embarrassing to his view. If the bar for consciousness is set high (one well-known philosopher thought it absurd to suppose that fish have experiences), this looks like a questionable prejudice against “lower” life-forms. Not as bad as Descartes's prejudice, perhaps, but still bad enough! But if we are generous in assigning souls to lower life-forms (another philosopher referred, only partly in jest, to “brother worm”!), we have the unappealing notion of God's creating souls by the billions for spiders, mosquitoes, and intestinal parasites. And then there is the problem of what becomes of all those souls when the creatures perish: since the souls don't depend for their existence on anything physical, it is hard to see how the death of a mosquito can extinguish its immaterial soul.

Yet another (and to my mind, powerful) objection appears when we try to connect creationism to the process of biological evolution. I believe the evidence overwhelmingly supports the claim that such a process occurred, but we obviously can't enter into that debate here.<sup>3</sup> It is clear in any case that anyone who accepts a view of the mind/soul that requires one to reject an evolutionary view of life's history incurs a heavy burden in the debate about the nature of persons. The problem is not merely that one is coming into conflict with a view that is widely accepted in our society; we should not, after all, be overly intimidated by prevailing intellectual fashions. The real burden lies in the need to provide a convincing alternative explanation for the immense, and ever-growing, body of data that supports belief in the evolution of life on the earth.

Unfortunately, creationism concerning the soul is difficult to reconcile with any plausible version of evolutionary biology. Evolution presents us with a very long and complex story in which, over hundreds of millions of years, the most primitive life-forms gave rise through a natural process to the complex kinds of life we observe today, including human life. The question is, how do the divinely created souls fit into this story? More complex animals, with greater cognitive powers, require “higher,” more powerful souls; a highly developed brain is useless to an animal if the animal's soul is unable to utilize it

effectively. And on the other hand, even a high-level soul is of little use if the required cerebral machinery is not in place. We cannot suppose that God waits until the brains have evolved through natural selection, and then supplies the requisite soul with its advanced capabilities: lacking the right sort of soul, the advanced brain would be nonfunctional and would not be conserved through evolutionary selection. The truth is that creationism concerning the soul just does not fit at all comfortably with an evolutionary account of life. This fact creates a significant burden for creationism, though one of which its proponents have often seemed to be unaware.<sup>4</sup>

The third kind of view concerning the nature and origin of souls is *emergentism*. The basic idea of emergence is that, when certain elements are assembled and related to each other in a certain way, something new and surprising can appear – something we would not have anticipated, merely on the basis of what we knew beforehand about the elements. Yet the new thing is not “added from the outside,” as is the case with creationism; rather, it appears as a natural consequence of the elements in their combination and relationship. Map out on a graph the geometric points that satisfy a certain equation, and something new appears: a fractal pattern, sometimes elegant and surprisingly beautiful. There are many candidates for emergence in the natural sciences: examples include the emergence of complex crystalline structures out of material that formerly had no such structure, and the emergence of elaborate group behavior from swarms of insects that individually give no evidence of such behaviors. Applied to the human mind, emergentism implies that consciousness, thought, rational volition, and so on make their appearance naturally as a result of the structure and functioning of the human brain and nervous system. It can be seen that in a sense emergentism is a mediating view between reductionism and creationism. Unlike reductionists, emergentists do not seek to “reduce” mental phenomena to their material base, in the process depriving them of much of the significance we ordinarily take them to possess. But unlike creationists, emergentists do not view the mind and its powers as being, as it were, injected from outside into the human biological system. Instead, the soul appears naturally, given the appropriate physical organization and function of the body and brain.

Emergentism is completely free from the objections we have noted against creationism. It is true that we lack deep insight into the way in which the mind is produced by the body and brain. But given that it is so produced, causal interaction is built in right from the beginning; it does not need to be added later as a separate assumption. Furthermore, the generation of the mind by the brain makes it much easier to understand the intimate, fine-grained dependence of mind on brain that we find to be the case, as compared with the situation for creationism. Emergentism has no problem with the souls of animals: nonhuman animals (those that enjoy any sort of conscious experience at all) have souls appropriate to the kinds of being they are, souls that are generated by their own nervous systems. Furthermore, emergentism provides a remarkably appropriate fit with an evolutionary account of the history of life on earth. Emergentism presents us with a compelling picture of the co-evolution of mind and brain. Genetic changes which lead to a more highly developed brain lead in turn to the emergence of a more sophisticated mind, a mind which has a more accurate grasp of its environment and responds in ways that enhance survival and reproduction. These improved responses in turn lead to the conservation of the genetic changes and to their becoming established in the population. Indeed, this feature of emergentism amounts to powerful argument in favor of emergentism over reductionism – an ironic result, in view of the materialist and reductionist bias of many evolutionary thinkers.

We need, however, to say something more about the kind of emergence that is involved here. A question that needs to be answered is: When we say that the mind is emergent, what precisely is it that is said to emerge? There are three different answers that can be given to that question, answers that yield a series of progressively stronger emergentist positions. The first answer to that question is that what emerges is *mental properties* and *mental events* – in particular, *conscious experiences*. It seems evident that these sorts of items are essential to any possible account of the mind/soul; without them, there is no mind to discuss. These things are not always viewed as emergent, but that is what they must be, given that they cannot be reduced to anything that is purely physical. Reductionists, to be sure, have invested an enormous amount of effort in the attempt to show that mental properties are reducible, either to behavioral properties, or to neurophysiological properties, or (the current favorite) to functional properties. These efforts at reduction have not been particularly successful, as we might expect. It just seems evident on the face of it that, for instance, the mental property of rejoicing in the fact that the Chicago Cubs have at last won the World Series (Hooray! They did it!) is not the same as any physical property, however complex, of the neural assemblages in a person's brain. The frustration that has attended these efforts at reductionism has led some philosophers to embrace *eliminative materialism*, the view that desires, purposes, intentions, and conscious experiences in general as they are ordinarily conceived, simply *do not exist* – that our belief in these things is merely evidence of our adherence to a primitive theory, called “folk psychology,” which like other primitive theories is destined to be replaced, in this case by a more adequate, and thoroughly physical, theory of the mind. (This, unfortunately, is a theory which does not yet exist.) Most of us, however (including most philosophers), find this incredible, which means that mental properties and mental events *do exist* – and if they cannot be reduced to physical properties, they must be considered as being emergent. The view that the *only* thing that emerges in the case of the mind is mental properties is often termed *nonreductive physicalism*; the view is “physicalist” with the sole exception of those properties. This label, and the view itself, have been criticized by Jaegwon Kim, who has argued forcefully that any physicalist view worth its salt needs to be reductionist (Kim 1993). In spite of this criticism, however, the name is still in common use, and we can accept it as a label, while recognizing that views so called may still be reductionist in other ways.

A second answer to our question is that what emerges, in addition to the mental properties, are *novel causal powers* that are associated with those properties. An example: when a person becomes angry, this often has results, not only for subsequent mental states, but for what happens in the physical world, results ranging all the way from loud and emphatic protests to acts of violence. And these results are different than what would have happened apart from that emotion. One might think that the emergence of such powers goes without saying, given the emergence of mental events and properties, but in fact it does not. Philosophers inclined toward physicalism often insist that the mental event (in this case, the feeling of anger) is itself the result of the person's physical brain-state, and that *any subsequent physical events (such as the physical reaction to the anger) are also the result of the physical brain-state*, operating throughout according to the standard laws of physics and chemistry. This strongly suggests that the view in question is one of *epiphenomenalism*, according to which mental properties and mental events are merely the accompaniments of the physical properties of body and brain; in themselves, the mental properties do no work and have no causal effects. This, however, is massively implausible, and most physicalists are determined to resist it, suggesting instead ingenious ways in which we may after all consider the mental properties as being

causally effective. The success of these attempts is dubious, but in any case physicalists uniformly insist on the *causal closure of the physical domain*, which means in effect that any event that has a cause at all has a sufficient physical cause. (The qualification is needed because quantum mechanics seems to tell us that some micro-events are purely random and uncaused.) This view, however, has consequences that are truly mind-boggling. For instance, it means that one never in fact performs an action *because* one consciously decided to perform it; rather, *both* the physical action and the decision to perform the action are entirely the result of physical brain-states, states over which one has no conscious control whatsoever. The view also implies that one never accepts the conclusion of a process of reasoning because one has followed the argument and has seen, rationally, that the conclusion follows from the premises; rather, one's acceptance of the conclusion is, once again, entirely the result of one's physical brain-state, a state which is in no way guided by the laws of logic or the principle of sound reasoning, but simply by the laws of physical cause and effect. The argument given earlier, to the effect that a reductionist view of evolution can have no explanation for the development of conscious thought, presents one more example of the unfortunate consequences of the doctrine of causal closure. Emergentists, I maintain, should reject that doctrine, and insist that *because of the occurrence of conscious mental events things go differently on the microlevel than they would go in the absence of such events*. This is an important conclusion, one that upsets much of the thinking that underlies contemporary naturalism concerning the human person.

When we affirm the emergence of conscious experience and of novel causal powers as sufficient to account for the human mind, we are in effect asserting that human beings consist, in the final analysis, of ordinary matter and nothing more. Such a view may be termed the theory of emergent material persons, or the EMP theory. There is, however, yet another answer to the question, "What emerges?" What emerges, on this third view, is not merely mental properties and experiences, and not merely new causal powers, but a *new individual*, a *subject* that *has* those experiences and *exercises* the causal powers in question. This new individual *is not composed* of the elementary particles of physics; rather, the new individual is an *emergent immaterial entity*, an "emergent self" which *as an undivided whole* undergoes conscious experiences of various kinds, acquires knowledge of itself and the world in which it lives, and carries out actions which serve (or are intended to serve) its ends and desires. By now this new individual is beginning to sound quite a bit like the soul posited by creationist versions of dualism; there are, however, some important differences. The emergent self is generated by the organic body through a natural process, rather than being inserted into the body from outside. It is also sustained in its continuing existence by the body and brain, and both its powers and its activities are intimately related to and dependent on the condition and functioning of the brain. And unlike most versions of creationist dualism, the emergent self is located in space: it exists in the region occupied by the brain and nervous system by which it is generated. Nevertheless, in spite of these differences the postulation of such a soul makes this view a version of dualism; indeed, of substance dualism and not of mere property dualism. Hence the name given to the view: *emergent dualism* (Hasker 1999, 171–203).

It has been argued in this chapter that an adequate theory of the mind-body relationship must be an emergentist theory. If this is the case, the EMP theory and emergent dualism will be the two most viable candidates of all the existing mind-body theories. If so, how should we go about deciding between them? The EMP theory has some distinct advantages. There is an obvious advantage in simplicity, in that the theory postulates that humans and other animals consist only of ordinary matter, rather than matter plus a rather mysterious



emergent immaterial soul. In taking this line, the EMP theory aligns itself with the materialism that is the prevalent view among both neuroscientists and philosophers who are interested in this topic. (To be sure, the postulation of emergent causal powers will be strongly resisted by many in the mainstream.) Furthermore, for some philosophers the sort of emergence postulated by emergent dualism will strain the limits of credibility. Timothy O'Connor, a leading proponent of the EMP theory, has remarked that "the idea of the natural emergence of a whole substance is perhaps a lot to accept" (O'Connor 1998).

There are reasons, however, to think the EMP theory may not be adequate in the end. Some of these reasons are more strictly theological; we will attend to them briefly later on. But there are also purely philosophical difficulties. Consider once again the emergent mental properties which play a role in this view. Timothy O'Connor has pointed out that these properties are *nonstructural*, meaning that the property's instantiation "does not even partly consist of the instantiation of a plurality of more basic properties by the entity or its parts" (O'Connor and Jacobs 2003, 541). Some examples may help to clarify this point. A person's standing upright consists in the fact that her limbs are aligned with one another, and with the surface of the earth, in a certain way. (These relationships would be quite different if she were sitting, or lying down, or running.) A person's weighing 150 pounds consists in the fact that each of his various parts (arms, legs, torso, head, and neck) has a certain weight, and these weights add up to 150 pounds. My shoveling the snow in my driveway consists in my hands gripping the handle of the shovel, my feet, legs, and torso being positioned so as to support my holding the shovel, and my arms pushing the blade of the shovel under a heap of snow, lifting it, and tossing it off to one side. All these are *structural* properties; the properties of the person as a whole (standing upright, weighing 150 pounds, shoveling snow) consist of properties of, and relations between, various parts of the person, in relation to the surrounding environment. But now consider my thinking to myself the thought that "Nonreductive physicalism is not adequate as a theory of the mind." How are we to analyze the structure of this event, in a way paralleling the other examples? Which parts of me are relevant, and what are those parts doing? Without doubt it is true that, when I am thinking this thought, certain neurons in my brain are firing in a particular pattern. But in which of these events does my thinking that thought *consist*? Certainly no one of my neurons is thinking that entire thought. Perhaps we could divide the thought into bits of some sort, and assign each of the bits to one neuron. But doing this would not, it seems, bring us any closer to our goal. We would then have many different neurons, each entertaining some tiny fragment of the thought, but we would still have no insight at all into whatever it is that *thinks the entire thought*, "Nonreductive materialism is an inadequate theory." What we need is a *single thing* that grasps and affirms that thought – and as yet, nothing of the sort has come into our view. Three centuries ago, the philosopher Leibniz wrote,

In imagining that there is a machine whose construction would enable it to think, to sense, and to have perception, one could conceive it enlarged while retaining the same proportions, so that one could enter into it, just like into a windmill. Supposing this, one should, when visiting within it, find only parts pushing one another, and never anything by which to explain a perception. Thus it is in the simple substance, and not in the composite or in the machine, that one must look for perception. (Leibniz 1991, par. 19)

It is a mistake to suppose that the problem arises because of the limitations of Leibniz's science. If we find his "parts pushing one another" implausible as the vehicle of thoughts, how would

replacing those parts with silicon chips, or with neurons, make things any better? The problem does not lie in the pushes and pulls but rather in the *complexity* of the machine, the fact that it is made up of many distinct parts, coupled with the fact that *a complex state of consciousness cannot exist distributed among the parts of a complex object*. The payoff of the argument comes in its final sentence: “it is in the simple substance, and not in the composite or in the machine, that one must look for perception.” This is the *unity-of-consciousness argument*, and it poses a serious problem for views which, like the EMP theory, affirm that human beings are composed of physical stuff and nothing else. What is sometimes said in response is that the subject of experience is the “person as a whole” – that is, the body as a whole. But if this is intended as an answer to the unity-of-consciousness argument it is simply an evasion. The body as a whole, on this view, consists ultimately of those microphysical particles and nothing more. To say that there is something going on here – something done *by that body* – which can’t be accounted for by anything done by those particles, either individually or in their interrelationships, verges on incoherence. To repeat: there is (by hypothesis) nothing there to do anything, *except* for those very same particles.<sup>5</sup>

The unity-of-consciousness argument takes a different turn if we introduce what might be termed “geographical” considerations, questions concerning the spatial location of a given experience. Suppose, for example, you are listening, with enjoyment, to a complex piece of music – say, the finale of Beethoven’s Ninth Symphony, the famous “Ode to Joy.” (Feel free to substitute another example according to your own musical taste.) We then pose the question, *where precisely does this experience occur?* Where is the property, *enjoying Beethoven’s Ninth Symphony*, instantiated? The point of this question becomes apparent when we consider the principle of *localized property instantiation*, which can be stated as follows:

- (LPI) The intrinsic (nonrelational) properties of a material thing are instantiated in the space occupied by that thing.

After a little consideration, this principle seems obvious. It makes no sense to suppose that a material object (say, a house) is in New York, but the instantiation of its F-ness occurs in Los Angeles (supposing, to be sure, that F-ness is an intrinsic, nonrelational property, for example the shape of the house). But when we ask where the experience, *enjoying Beethoven’s Ninth Symphony*, is instantiated, a further question raises its head. Since you, the experiencer, are by hypothesis a material object, the experience must occur somewhere within the boundaries of your physical existence – someplace inside your skin. Plausibly, it will occur someplace in the space occupied by your brain. We learn from physics, however, that the ultimate physical constituents of the brain, the elementary particles, are not packed in, so to speak, cheek to jowl, like riders in a crowded subway train. Rather, these particles are separated by comparatively vast amounts of empty space.<sup>6</sup> Now, if we suppose that the experience occurs in all of the space within the boundaries of the relevant brain-region, we encounter the fact that *in most of this space no material particles exist*. And yet, the brain consists precisely of those particles and nothing else. But if conscious experience literally exists in the interstices between the particles, it must nevertheless inhere in *something*, must it not? (To say that it need not inhere in anything is in effect to make consciousness a substance in itself, which is hardly a move a materialist will want to make.) So if consciousness exists in all of the space in the relevant brain-region, including the interstices between the particles, it must inhere in something other than the physical brain – and if so, what more is needed to make this hypothesis explicitly dualistic?

It would seem, then, that the EMP theorist needs to say that consciousness exists *in the brain's particles and nowhere else*; in particular, not in the comparatively vast empty spaces between the particles. And yet these particles are linked together in such a way that the very same conscious state occurs in all of them simultaneously *without there being any physical communication of information between the particles*.<sup>7</sup> Now, consider just one of the particles in your body, a quark named Jeremy. Jeremy is a *down* quark which is combined with two *up* quarks to make up one of the protons in the nucleus of a particular carbon atom in your brain. You are listening, with great enjoyment, to the triumphant final movement of Beethoven's Ninth Symphony. Pursuing our previous line of thought, we may ask, *What is going on in the tiny region of space occupied by Jeremy?* There is no question that part of the answer must be that this region of space contains the conscious state, *listening with enjoyment to the Ninth Symphony of Beethoven*. The unity-of-consciousness argument has shown that it can't contain just *part* of that conscious state – and if it contains none of the state (and the same is true of the other regions that are relevant), then the state just doesn't exist anywhere in that spatial region, which is to say it is not a state of any material object. We go on to ask, *Who or what is the subject of this conscious state?* Our initial answer, no doubt, will be that the subject of the state is you yourself, the person doing the listening. By hypothesis, however, this person consists of a certain physical body and nothing else. And of that body, nothing exists in the region in question except for the quark Jeremy. You, the person, are present in that region of space, and your presence there *consists entirely of the presence there of Jeremy*. There is a property, *enjoying Beethoven*, which literally is exemplified in that region of space. There is nothing physical in that region of space except for the quark, Jeremy. And there is nothing nonphysical in that region of space that might exemplify the property, *enjoying Beethoven*. The conclusion seems ineluctable: if the experience of enjoying Beethoven occurs in the space occupied by the quark Jeremy, the subject of that enjoyment must be Jeremy itself. It looks very much, then, as if the EMP theorist is obliged to say that yes, my quarks do, quite literally and not as a matter of any figure of speech, enjoy Beethoven. But this, I submit, is incredible.<sup>8</sup>

It is time to summarize. If the property, *enjoying Beethoven*, is a property of a physical object, it must be instantiated in physical space, presumably in (all or part of) the space occupied by a person's brain. In particular, it must be instantiated in the space occupied by the material particles (protons, electrons, and so on) within the brain; it can't be instantiated in the vast empty space between the particles, because in that space there is no physical thing that could instantiate it. But this entails that the physical particles in the person's brain (in this case, Jeremy) instantiate the experience, *enjoying Beethoven* – which is to say, that any such particle literally has the experience of enjoying Beethoven. Since this cannot possibly be true, it follows that the enjoyment of Beethoven cannot be a property of the brain, or of any other material thing. What is needed, then, is an experiencer that is not a physical thing – and precisely this is what is supplied by emergent dualism.

It was mentioned above that the EMP theory may run into theological difficulties. Now, it can't be emphasized too strongly that the Bible *does not have* a developed philosophical theory of the human person; this is an area in which proof-texting is very much out of place. But if scriptural passages consistently, and over a broad range of texts, seem to imply something about the human person that is contradicted by a philosophical theory of the person, this raises legitimate questions. There is reason to think that the New Testament consistently assumes that human beings are able to persist, without an ordinary physical body, between a person's death and the final bodily resurrection, something that would be

impossible if humans consist entirely of ordinary matter (Cooper 2000, 2009). An even more crucial question, however, is raised by the very concept of resurrection. Clearly, there is no difficulty in the thought that God is able to craft a “resurrection body” that is as similar as one might wish to the body of a person that has perished, and to do so either out of the very same matter (to the extent that it is still available) or out of new matter. But if this is done, what it makes it the case that the person thus constituted is *the very same individual* that previously perished? On the face of it, when a person, consisting solely of ordinary matter, dies, and her body decomposes (part of the matter perhaps being incorporated into other bodies or even annihilated in a nuclear explosion) – when that happens, that person ceases to exist, and any similar person subsequently created would be a new individual. Appeals to divine omnipotence at this point are of no use; before we can appeal to God’s power to accomplish a task, it must be established that the task in question can be described in a way that is logically coherent; otherwise, we are saying God can do something nonsensical. Now, this remains an area of major controversy, and we can’t go into the details here. However, there is reason to think that the solutions to this problem proposed by Christian materialists are unsuccessful; that at least is my own view of the matter. If so, then views in which human persons are wholly material beings are simply unacceptable for the purposes of Christian theology (Hasker 1999, 204–235; 2011).

Emergent dualism, on the other hand, suffers from no such objection. As is the case for creationist versions of dualism, the identity of the person is guaranteed by the soul, which persists after bodily death and is invested in a transformed, glorious body in the resurrection. To be sure, special divine assistance may be needed to sustain the soul during the interval when it lacks its ordinary support from the body. Something similar to this is needed also by creationists: creationist views generally imply that the soul is dependent on the body in some respects (e.g., sensory capacity and memory) that are essential to its normal function. But the problem of the identity of the resurrected person with the individual who previously lived, a problem which is extremely difficult for physicalist views, simply does not exist for emergent dualism.

Emergent dualism, it has to be said, is not the sort of theory that immediately strikes one as obviously correct; the notion of an immaterial soul emerging from the functioning of a physical brain can easily seem puzzling. The benefits of the theory have to be perceived in comparison with its competitors. (If the mind-body problem admitted of a simple, obviously correct, solution we philosophers would have a lot less work to do!) I have argued, nevertheless, that upon reflection emergentism is superior both to reductionism and to creationism. And among the emergentist options, emergent dualism is the one that best satisfies the requirements of both good philosophy and sound theology. I submit that it is worthy of our acceptance.

## Notes

1. It should be mentioned here that there is another variety of creationism, Thomistic dualism, which differs in some important ways from Cartesian dualism. See in this volume, Chapter 6, by Edward Feser, and Chapter 8, my “Critique of Thomistic Dualism.”
2. At present our most fundamental theory about the nature and behavior of matter is quantum mechanics. The theory is accepted because of the wide-ranging and astonishingly accurate predictions that are made from it, but it is a commonplace among the experts on this theory that no one really understands why matter behaves in the (sometimes very strange) ways the theory describes. Yet all of the “ordinary” causal interactions, which we imagine that we do understand, are the result precisely of these mysterious quantum interactions!

3. I find the evidence that an evolutionary process occurred completely convincing. This does not, however, entail that the prevailing “neo-Darwinian synthesis” provides a complete and adequate explanation of that process.
4. I mention here two especially eminent examples. Richard Swinburne’s book, *The Evolution of the Soul* (1998) is one of the best recent defenses of Cartesian dualism. Swinburne is deeply engaged with science and has no interest in rejecting evolution, but he does not discuss the tension between evolution and his creationist view of the soul. (See, however, his contribution to this volume for a different approach, one that might fairly be termed a version of emergent dualism.) In Alvin Plantinga’s book, *Where the Conflict Really Lies* (2011) he devotes two entire chapters (63 pages) to alleged conflicts between Christian faith and evolution, and evolution is referred to frequently throughout the rest of the book. Yet he also never mentions the tension between evolution and his own, presumably creationist, view of the soul.
5. For more on the unity-of-consciousness argument, see Hasker (1999, 122–146); see also Hasker (2016).
6. Electrons and quarks are each approximately one attometer (10–18 meters) in size; protons and neutrons one femtometer (10–15 meters); a hydrogen atom is 25 picometers ( $25 \times 10^{-12}$  meters). The diameter of the hydrogen atom, and therefore the orbit of its one electron, is around 25 million times the size of that electron! This is roughly equivalent to a single golf ball tracing an orbit 600 miles in diameter. It’s lonely in there! It is precisely this situation that enables the matter inside a star to achieve densities that are unimaginable to us planet-dwellers. Compressed by the enormous pressure within a neutron star, the empty space between particles is eliminated, and matter becomes so dense that a teaspoon-full of it would weigh 10 billion tons!
7. There will, of course, be the normal forces and interactions that occur between the microelements of any material thing whatever. It is not credible, however, that the information of the conscious field is transmitted by those forces and interactions.
8. It has been suggested to me that the experience occurs in the entirety of the space occupied by the brain, but not in any subregion of that space. This is incoherent. Consider once again the space occupied by Jeremy. The experience, enjoying Beethoven’s Ninth Symphony, either occurs in that region or it doesn’t. If it doesn’t (and the same is true of the other subregions of the brain-space), then the experience does not occur in space at all and it is not an experience of any physical thing. If it does, then experiencer has to be Jeremy, as was argued in the text.

## References

- Cooper, John W. 2000. *Body, Soul, and Life Everlasting: Biblical Anthropology and the Monism-Dualism Debate*, 2nd edn. Grand Rapids, MI: Eerdmans.
- Cooper, John W. 2009. “Exaggerated Rumors of Dualism’s Demise: A Review Essay on *Body, Soul, and Human Life*.” *Philosophia Christi*, 11(2): 452–464
- Hasker, William. 1999. *The Emergent Self*. Ithaca, NY: Cornell University Press.
- Hasker, William. 2011. “Materialism and the Resurrection: Are the Prospects Improving?” *European Journal for the Philosophy of Religion*, 3(1): 83–103.
- Hasker, William. 2013. “What Is Naturalism? And Should We Be Naturalists?” *Philosophia Christi*, 15(1): 21–34.
- Hasker, William. 2016. “Do My Quarks Enjoy Beethoven?” In *Neuroscience and the Soul: The Human Person in Philosophy, Science, and Theology*, edited by Thomas M. Crisp, Steve L. Porter, and Gregg A. Ten Elshof 13–40. Grand Rapids, MI: Eerdmans.
- Kim, Jaegwon. 1993. “The Myth of Nonreductive Materialism.” In *Supervenience and Mind: Selected Philosophical Essays*, 265–284. Cambridge: Cambridge University Press.
- Leibniz, Gottfried Wilhelm. 1991. *Monadology*. In *G. W. Leibniz’s Monadology: An Edition for Students*, edited by Nicholas Rescher Pittsburgh: University of Pittsburgh Press.
- O’Connor, Timothy. 1998. “Comments on Jaegwon Kim, ‘Causality and Dualism.’” Unpublished paper presented at the University of Notre Dame.
- O’Connor, Timothy, and Jonathan D. Jacobs. 2003. “Emergent Individuals.” *The Philosophical Quarterly*, 53(213): 540–555.
- Plantinga, Alvin. 2011. *Where the Conflict Really Lies: Science, Religion, and Naturalism*. Oxford: Oxford University Press.
- Swinburne, Richard. 1998. *The Evolution of the Soul*, rev. edn. Oxford: Clarendon Press.

# Against Emergent Dualism

BRANDON L. RICKABAUGH

Materialism permeates philosophy of mind. Yet, increasing difficulties are triggering dissent (Koons and Bealer 2010; Göcke 2012; Lavazza and Robinson 2013). William Hasker's insightful yet underappreciated work in the philosophy of mind is representative. Hasker does not favor the recent turn toward nonreductive physicalism either. Rather, his argument from the unity of consciousness entails *substance dualism*. Additionally, Hasker maintains that the soul is an emergent substance, a view he calls *emergent dualism*. Although Hasker's is a minority position, similar if not identical views are defended by LaRock (2013), Zimmerman (2011), and Eccles and Popper (1993). Hasker's criticisms of materialism are formidable, and his unity of consciousness argument deserves considerable attention. Still, emergent dualism faces difficulties. I argue that emergent dualism is not more attractive than nonemergent versions of substance dualism as Hasker suggests. I raise several new problems for emergent dualism that nonemergent versions of substance dualism evade.

## 5.1 What is Emergent Dualism?

According to substance dualism the following is true,

*Substance dualism:* Human persons are not identical to any physical body, but consist of a physical body and a nonphysical substantial soul.

The term “consist” leaves room for various views, according to which a human person is the soul, or is an embodied soul (see Evans and Rickabaugh 2015 for more on these distinctions). Hasker adopts a more specific version of substance dualism:

*Emergent dualism:* (a) Human persons are not identical to any physical body, but consist of a physical body and a nonphysical substantial soul, and (b) the human soul is naturally emergent from and dependent on the structure and function of a living human brain and nervous system (Hasker 2014, 215–216).

It is (b) that makes Hasker's view unique. Other types of substance dualism reject (b). What then is the nature of emergence and the emergent soul?

### 5.1.1 *The nature of emergence and emergent dualism*

According to the doctrine of emergence, when the parts of a system are arranged in the right way their collective activity brings something new into existence. The new emergent entity is not reducible to the parts of the system it came from. It is distinct in terms of its structure and causal powers. Hasker embraces four categories of emergence corresponding to various properties, events, and substances.

Emergence<sub>0</sub>: the arrival of new features (such as a fractal pattern or crystalline structure) that are logical consequences of a new arrangement of lower-level elements (Hasker 1999, 173).

Emergence<sub>1a</sub>: the arrival of higher-level features (such as solidity, liquidity, and transparency) resulting from causal interactions among lower-level elements (Hasker 1999, 173–174).

Emergence<sub>1b</sub>: the arrival of new modes of behavior, *emergent causal powers* (powers the brain has) describable only in terms of the operations of *new emergent laws* (systems of interaction between atoms and molecules of the brain) resulting from causal interactions among lower-level elements (Hasker 1999, 174; Hasker 2005, 77).

Emergence<sub>2</sub>: the arrival of a new entity (a nonphysical substantial soul) with libertarian freedom (powers of the soul), resulting from, but not consisting in, the proper configuration of a material substrate (Hasker 1999, 177; Hasker 2012a).

According to Hasker, emergent causal powers are, “already implicit in the physical ‘stuff,’ otherwise their emergence would sheerly be magical” (Hasker 2004, 112). They only become detectable once the base elements are appropriately arranged. It is at emergence<sub>2</sub> that human persons appear, as Hasker explains:

emergentism implies that consciousness, thought, rational volition, and so on make their appearance naturally as a result of the structure and functioning of the human brain and nervous system . . . Emergentists do not view the mind and its powers as being, as it were, injected from outside into the human biological system. Instead, the soul appears naturally, given the appropriate physical organization and function of the body and brain. (Hasker 2015, 152)

It is imprecise to say that, “the soul appears naturally.” Perhaps Hasker means to say the soul regularly appears under the right circumstances. He cannot mean “in a way explicable by naturalism.” Hasker is committed to a kind of panpsychism – protopanpsychism – that has historically rivaled naturalism (Skrbina 2005; more on this below). Secondly, it seems odd that a physical object generates the soul (Hasker 2012b, 196, quoted in Section 5.1.2.2). According to Hasker's view, the brain's mental potentialities (which I find perplexing) and the fact that causal dispositions are as much a part of a thing's essence as its categorical properties, mean that the soul is generated by a *psycho-physical object*. This is why many physicalists avoid emergentism.<sup>1</sup>

Lastly, according to emergent dualism, mental properties and mental events emerge from the brain (Hasker 2015, 157). They are distinct from and not determined by the operations of the brain or its physical properties (Hasker 2014, 217). Additionally, novel causal powers associated with these mental properties and mental events emerge. This makes emergent dualism a type of protopanpsychism. These novel causal powers also commit Hasker to the incompleteness of physics and the rejection of the causal closure of the physical. Most importantly, the nonphysical, substantive soul also emerges from the brain and brain stem. So, what is the nature of the emergent soul?

### 5.1.2 *The nature of the emergent soul*

#### 5.1.2.1 **The emergent soul as the person, the conscious self**

The soul of emergent dualism is the conscious self. It thinks, reasons, feels emotion, makes decisions, and is at the core of what we mean by “person” (Hasker 2005, 78). The emergent soul has mental experiences and exercises these emergent causal powers (Hasker 2015, 159). Hence, certain facts about the soul *wholly* ground facts about personhood, consciousness, intellect, affect, and will. This must be the case as Hasker maintains it is metaphysically possible for the soul to exist apart from the body. This is standard substance dualism.

#### 5.1.2.2 **The emergent soul as an immaterial, unified, substance**

The emergent soul is immaterial, as it has no material parts. It is a substance with its own causal powers. Additionally, the soul is unified, although isn’t always clear what Hasker means by this. “The self of emergent dualism,” says Hasker, “is not a Cartesian soul: it is generated by the functioning of a physical object and is itself spatially located, and it is not simple in the way that a Cartesian soul is simple” (Hasker 2012b, 196). Hasker speaks of the simplicity of the soul as “an undivided whole” (Hasker 2015, 159) Elsewhere, he states

the self that is the subject of experiences must function as an undivided unity and not as a system of parts. But this does not immediately carry with it all the freight traditionally attached to metaphysical doctrines of the “simplicity of the soul.” (Hasker 2012b, 187)

It appears that Hasker’s emergent soul is not mereologically simple, but only an undivided whole. In fact, the soul, according to Hasker, is a substance that could be, and in some cases is, divided. Hasker states,

I have repeatedly argued, for example, that the emergent self could under certain circumstances be divided – for instance, by the fission of the generating organism. (Arguably the famous cases of “brain bisection” through commissurotomy constitute partial examples of this possibility.) (Hasker 2012b, 187; see also Hasker 1999, ch. 7; 2010)

These remarks are in tension with a more recent statement of Hasker’s:

the self, the subject of experience, cannot be a complex physical object such as the human body or brain. Instead it must be a *simple substance*, one that has no parts that are themselves substances, and which cannot be divided into parts. (Hasker 2016, 28)



Taken together, these remarks are perplexing. Hasker holds that the soul is a unified whole, but is he talking of functional unity or the metaphysical unity had by a mereologically simple substance? His early remarks suggest not. However, his argument from consciousness, his recent endorsement of David Barnett's simple argument (Barnett 2010), and the quotation immediately above, strongly suggest that Hasker holds that the soul is mereologically simple (Hasker 2016). It has no separable parts. This is no trivial point, as this plays a significant role in an objection below.

### 5.1.2.3 The emergent soul as metaphysically and developmentally dependent

Lastly, Hasker thinks of his view as a type of integrative dualism. The soul and body form a completely integrated unity. The soul is nomologically dependent on the brain and brain stem for its existence and development. "The mind/soul," Hasker explains, "is both generated and sustained by the biological organism, and its activities are subserved and enabled by the function of the organism" (Hasker 2005, 79). The soul is developmentally dependent on the brain for the gradual development of its capacities in accordance with the developing brain and nervous system (Hasker 2014, 216).

### 5.1.3 Is the emergent soul unique?

According to Hasker, emergent dualism does not have the problems of nonemergent versions of substance dualism (Hasker 1999, ch. 6). However, Hasker's view of the soul does not substantially differ from nonemergent forms of substance dualism. Consider, for example, Descartes's view. Like Hasker, Descartes maintained that the soul is a unified, immaterial, substance, capable of being supernaturally sustained by God apart from its body. Hasker's account of the uniting of the soul with its resurrection body is more Cartesian than emergentist, as the renewed soul is infused into a new body (Hasker 1999, 135). Perhaps Hasker thinks his view departs from Descartes regarding the strong interactive dependency between soul and body. However, this is mistaken.

According to Descartes it is *metaphysically possible*, by the miraculous intervention of God, for the soul to exist without the body. This is far different than thinking the soul and body exist separately.<sup>2</sup> The mere metaphysical possibility of separation is all Descartes's arguments conclude. Descartes adamantly rejects the dualism of Plato, which sees the soul and body as entirely separate. The soul, says Descartes, does not simply reside in the body "as a pilot resides in a ship," but rather forms a kind of natural unity "most closely joined" and "as if intermixed" with it.<sup>3</sup> Descartes's view has more in common with scholastic-Aristotelian theories of soul-body union than is often portrayed.<sup>4</sup> For Descartes there is no "ghost in the machine."<sup>5</sup> He refers to the soul as a "substantial form,"<sup>6</sup> and "substantially united"<sup>7</sup> with the body. Descartes's view, like Hasker's, is that the mind and body are intimately connected.

Hasker's view differs little from nonemergent substance dualism, which he admits can account for the strong dependence of the mind on the brain (Hasker 1999, 155, 157). Where Hasker's view does diverge significantly has to do with the origin of the soul. Descartes, like Augustine and Aquinas, held that God creates the soul.<sup>8</sup> Hasker, however, thinks the soul naturally emerges from a complex arrangement of matter. That is, according to emergent dualism, there are psycho-physical laws that determine when a soul comes into existence. These laws kick in as it were, when an aggregate of matter reaches the necessary level of

complexity. It is not clear if these laws are necessary laws or contingent laws. However, I am persuaded by Rasmussen's argument in this volume that no psycho-physical laws are necessary.

I am convinced that Hasker's preference for emergent dualism boils down to aesthetic considerations. He seems to think it is aesthetically more pleasing not to have God intervening billions of times creating souls *ex nihilo*. However, Hasker has God intervening just as many times to directly preserve the disembodied soul in an intermediate state. If God has to do that, what is so aesthetically displeasing about God intervening to create souls *ex nihilo*? It is arguable that being able to create something *ex nihilo* is a great-making property. To think matter has this power is theologically bizarre.

## 5.2 Problems for Emergent Dualism

Hasker's view has been met with interesting objections.<sup>9</sup> In this section, I introduce problems unique to emergent dualism that nonemergent versions of substance dualism evade.

### 5.2.1 *A problem for emergence and explanatory power*

Dallas Willard argued that John Searle's use of emergence to explain consciousness is problematic. I suggest some of these objections apply to Hasker's view. Willard states,

we do have a pretty good story about what the atoms, molecules, etc., do to produce solidity, liquidity, boiling of the H<sub>2</sub>O. But in the case of the brain and its alleged emergent properties of consciousness, there is just no story at all. At best we have a rather crude set of brute correlations indirectly established. (Willard, n.d.)

Hasker admits that the emergence of new laws and presumably consciousness is, "a far cry from the 'emergence' of solidity and the like, which is based entirely on the operation of the micro-elements according to the standard laws of physics and chemistry" (Hasker 2016, 20). There remains an explanatory gap between how molecular states of H<sub>2</sub>O produce solidity and how brain states produce mental states, much less a soul. Willard rightly points out, "There is nothing about brain cell activity that would naturally associate it with intentionality, moods, qualia or subjectivity as these are present on the mental side of our life" (Willard, n.d.). Of course, Hasker is in agreement with Willard that the brain cannot be conscious. Why think the problem does not extend to the emergence of the soul from a brain?

Notice, this is a problem in principle, not one that merely reflects the current state of our knowledge. When emergent properties were characterized epistemically, they were described as those properties which an exhaustive, God's eye view of the subvenient base would not predict or be able to explain. This is true today even though emergent entities are characterized ontologically. They are so unique, that all one can say is that when matter reaches "the right stage" of complexity, emergence obtains (Hasker 2005, 76). However, there is no informative, noncircular way to specify what "the right stage" amounts to. All one can say is that it is the way things need to be for emergence to occur. But, this

borders on being tautological, rather than informative: when things are arranged so that emergence occurs, emergence occurs.

Moreover, Willard argues,

A striking lack of parallelism between the H<sub>2</sub>O cases and the brain is the following: We have absolutely no theory independent knowledge that the brain has properties Searle assigns to it, as we do that the water is boiling, solid, or liquid. (Willard, n.d.)

Likewise, we have absolutely no theory independent knowledge that the soul emerges from the brain. Much of Hasker's case for emergent dualism rests on emergence<sub>2</sub>, which is made plausible by the applicability of emergence<sub>1</sub> mental properties. Willard has given us reason to question that emergence<sub>1</sub> includes mental properties, and these arguments can be applied to emergence<sub>2</sub>. This raises concerns for Hasker's view. Talk of emergence does not illuminate the issue of the origin of the soul, as the comparison of H<sub>2</sub>O exhibiting solidity is "a far cry" from the brain producing a soul. In fact, I am inclined to think that talk of emergence is merely a label for the problem to be solved and not itself a solution.

Lastly, it seems that emergent dualism lacks the kind of explanatory power it is intended to provide. Dean Zimmerman, an advocate of emergent dualism, states,

The details of the mechanism by which brains generate souls remain, admittedly, as speculative as the search for a special conscious particle or a precisely demarcated conscious chunk of brain matter. (Zimmerman 2011, 195)

I think Zimmerman is correct. Consequently, we have reason to be suspicious of Hasker's claim that emergent dualism is more scientifically informed than nonemergent forms of substance dualism. If the brain produces consciousness, then there is no illumination regarding the origin of consciousness, as emergence is not illuminating but mysterious. Moreover, if emergentism includes the view that there are necessary psycho-physical laws, then I can't see how anything from science could in principle favor emergentism over the view that the psycho-physical laws are contingent.

### *5.2.2 An emergent problem of the many*

Here I apply Peter Unger's mental problem of the many for materialism to emergent dualism (Unger 2004). The argument begins by noting the composition of material objects. A chair, for example, is composed of billions of atomic simples arranged in the shape of a chair. However, the chair could have been made slightly smaller and thereby would have been composed of slightly less atomic simples. It seems that a chair composed of billions of atomic simples actually has overlapping parts or subregions that are themselves billions of atomic simples arranged in the shape of slightly smaller chairs. Therefore, for any chair there are actually billions of sub-chairs overlapping one another. We might not talk this way, but this is true of any macro-material object. Of course, on a mechanistic view such as Hasker's, this includes human brains.

This can be applied to develop an argument against emergent dualism:

P1. If emergent dualism is true, then wherever there is a properly complex arrangement of parts of kind-*k*, a soul emerges.

- P2. A brain is an aggregate of billions of parts of kind-*k* with many subregions of overlapping parts of kind-*k*.  
 P3. Therefore, if emergent dualism is true, there should be multiple souls emergent from any one brain.

Notice that P1 is a commitment of Hasker's view. As already explained, P2 is a commitment of Hasker's regarding all macro-material objects. A properly complex arrangement of parts of kind-*k* just is a living, properly functioning human brain. Like the table the brain could have been just a little smaller in billions of ways and still remained a brain. These billions of overlapping parts are all within any normal brain. Hence, P2 is justified, and the conclusion logically follows. Consequently, emergent dualism entails the incredible notion that from my brain emerge billions of souls with experiences just like mine, thinking my thoughts.<sup>10</sup> This is absurd. As those like Brentano (1973, 16), and Chisholm (1981, 86–88) have argued, this conclusion is incompatible with the basic awareness I have of myself as a single unified self.

One might object that what is true of chairs is not true of brains. Perhaps there is some specific number of parts necessary for kind-*k* such that a brain is not really a mass of overlapping smaller brains. There is empirical evidence that seriously undermines this notion. There are cases of people who function normally yet, either congenitally or due to surgery, are missing parts and even significant regions of their brain. For example, hemispherectomy involves the removal of parts of a hemisphere (functional hemispherectomy) or even an entire hemisphere (anatomical hemispherectomy). Many of these patients recover normal cognitive function.<sup>11</sup> Second, studies by John Lorber of patients cured of hydrocephalus revealed that 60 of the 600 patients had ventricular fluid occupying at least 95 percent of their cranial capacity! In spite of missing so much of their brain, and to the astonishment of neuroscientists, "half of Lorber's 60 cases were of above normal intelligence" (Forsdyke 2015; see also Lewin 1980). There are in fact subregions of the brain that have all the necessary parts and relations to function as a brain. In another context Hasker states that "presumably two generating bodies – two distinct brains and nervous systems – cannot occupy the same space" (Hasker 2012c, 224–225). However, Hasker simply assumes this, and I have given reason to doubt this presumption.

Perhaps Hasker can place a restriction on (P1), such that a complex arrangement of parts is necessary for the emergence of the soul, but not sufficient. It seems Zimmerman has something like this in mind where he writes,

I suppose that the following hypothesis is more likely: that many overlapping sets of events occur in the brain, none of which is the minimal cause of the soul's ongoing existence, nor the single cause of its overall phenomenal state. With many overlapping patterns of neural firings, each lawfully sufficient for the existence of a soul with the same phenomenal states, there could still be just one soul, its existence and phenomenal state simply *overdetermined*. There need be no vagueness about which activities in the brain generate the subject of consciousness – in fact, on this supposition, *many* precise (and largely overlapping) events are equally responsible – nor about how many there are. (Zimmerman 2011, 195; emphasis in original)

This reply might work regarding neural activities. However, the problem I have suggested is not with the vagueness of activities, but of the brain itself. Assume the emergent laws Zimmerman stipulates. It is still problematic that a brain has a vast number of subregions each of which is sufficient for the emergence of the soul. In this way, my argument sidesteps

views such as Trenton Merricks's (2001) that reject the notion of overlapping brains. Even if Merricks is right that there are not multiple overlapping brains, still you have multiple functions that are individually sufficient for the emergence of a soul (if any of them are). The problem isn't at the level of the laws, but at the level of the complex arrangement of atomic simples and their functions. Any restriction Hasker might place would be ad hoc. Perhaps Hasker might dig his heels in and assert that only one of the configurations yields a soul. This, I will argue, creates a sorites problem.

### 5.2.3 *A sorites problem*

Recall that according to emergent dualism the soul emerges once an aggregate of matter is arranged in the right way. Hasker writes, "prior to the emergence [of the soul] there was only an aggregate of simples, even if the arrangement of the simples was very similar to that of the body parts of a person" (Hasker 2012b, 185). As previously argued, there are problems for how to understand what "the right way" could mean. Here is an additional problem. Consider an aggregate of matter arranged in such a way as to be a properly functioning brain and central nervous system, and suppose that the number of parts arranged in the right way required for a soul to emerge is  $n$ . Consequently, there is some subset of those parts – atomic simples – and their arrangement such that they are sufficient, not for a soul to emerge, but for an organ just short of a brain to exist. Now, suppose God has this organism and is adding one by one and activating atomic simples until being just shy of  $n$  by one simple ( $n-1$ ). At the moment God adds the final atomic simple, the relevant emergence<sub>2</sub> laws activate, and the soul emerges.

Here is the problem. There is no metaphysically significant difference between having  $n-1$  atomic simples behaving in a certain way and having  $n$  atomic simples behaving in a certain way. The two cases do not differ in a metaphysically significant way at all. However, according to emergent dualism, the soul comes into existence once 1 is added to  $n-1$ . Hasker's view is committed to attributing a quite a significant degree of power to the addition of one atomic simple.

How could a defender of emergent dualism reply? Holding out for an empirical discovery to give us the precise number or range for  $n$  required for a soul to emerge wouldn't change the metaphysical problem. Perhaps one could reply that it isn't merely the number  $n$  of atomic simples that does all the work, but that we also need the correct arrangement of  $n$  atomic simples. This seems to be what Hasker has in mind. This does relieve the number of atomic simple of doing all the metaphysical work. But is this good enough? I think the answer is, no. Notice that the aggregate of  $n-1$  atomic simples is already structured. At the moment the final atomic simple is added, the structure is complete. It is still the addition of the final atomic simple to the aggregate that does all the work. It is simply added in the right place. Emergent dualism is left with the problem of attributing a significant degree of power to the addition of one atomic simple.

### 5.2.4 *A combination problem for emergent dualism*

A view gaining popularity among materialism dissenters is called panpsychism, according to which fundamental physical entities (perhaps quarks and photons) have conscious experience and bind together to form conscious human persons. Philosophers such as David Chalmers (2015) think panpsychism and protopanpsychism enjoy the advantages of

materialism and dualism without being encumbered by the problems either face. However, panpsychism faces the combination problem. Roughly, the problem is to explain just how the experiences of fundamental physical entities combine to yield the kind of conscious experience we have every day (Chalmers 2016).

The combination problem arises in virtue of attributing certain powers to matter. I think emergent dualism faces a similar problem.<sup>12</sup> “The mind/soul,” Hasker explains, “is both generated and sustained by the biological organism, and its activities are subserved and enabled by the function of the organism” (Hasker 2005, 79). Consequently, emergent dualism maintains that a unified nonphysical substance emerges from a physical aggregate of parts. To understand the problem, recall the dependence relation between body and soul Hasker is committed to. He states,

My view is that the conscious field is indeed divided as a consequence of the supporting organism’s being divided, and this seems an entirely plausible view, once we grant the possibility of the field’s being generated by the organism in the first place. (Hasker 2000, 203)

Note, this implies a grounding relation between body and soul. Accordingly, the body is ontologically more fundamental than the soul in certain respects. According to standard accounts, *As* are ontologically more fundamental than *Bs* in the relevant sense, if facts about the existence of *Bs* are grounded in facts about *As*.<sup>13</sup> In this case, facts about the soul (being divided) are grounded in facts about the body (the organism being divided). So, what facts about the soul are grounded in facts about the body? It seems that at least *structural facts about the body ground structural facts about the soul*. The body splits, so the soul splits. How is it then that the structure of the soul is such that it is a unified whole when the body, which grounds the structural facts of the soul, is not itself unified?

Let’s return to the combination problem for panpsychism. Regarding panpsychism, Hasker writes,

What we need is a *single* mind, and a single field of conscious awareness, for each sentient being . . . Panpsychism leaves this need unmet, and we would still need an explanation of the process by which those bits of mind-dust are fused into a single conscious mind. But given that this need would remain, not much is gained by postulating the mind-dust to begin with. (Hasker 2016, 40)

Quoting John Searle, Hasker states that what we need to know is “how the brain produces the peculiar organization of experiences that express the existence of the self” (Searle 2008, 136). Of course, Hasker’s view does not face the same combination problem as panpsychism as it does not attribute consciousness to bits of matter (Chalmers 2016). But Hasker’s view, which is a type of protopanpsychism, faces a similar problem.

Emergent dualism faces the problem explaining how bits of nonconscious matter combine to constitute mental properties and produce a conscious soul. More precisely, emergent dualism needs an explanation for the process by which a unified, immaterial, substance, with a complex organized system of mental properties, that is mereologically simple emerges from a material organism composed of billions of separable parts. It seems equally mysterious and equally problematic as to how a brain could be organized in such a way as to produce a soul. This seems more difficult than the original combination problem as Hasker posits emergent<sub>2</sub> causal powers and emergent<sub>2</sub> souls. Of course, souls seem to

solve the binding problem and the unity of consciousness in general. However, the cost of emergent dualism is just as great as the cost of panpsychism insofar as the emergence of a soul is just as mysterious and perhaps more problematic than getting conscious minds from conscious matter.

Perhaps one could reply that I've merely shifted the burden of proof onto the defender of emergent dualism, and that this isn't itself a reason to reject the view. Fair enough, although I've argued this is a heavy burden. Without a plausible solution, emergent dualism is at least unmotivated.

However, we can strengthen the objection beyond exposing a weakness in Hasker's view to provide a substantive defeater. The argument I have in mind is as follows:

- P1. If emergent dualism is true, then facts about the unity of soul are wholly grounded by facts about the unity of the brain.
- P2. The facts about the unity of the soul cannot be wholly grounded by facts about the brain.
- P3. Therefore, emergent dualism is false.

What can we say on behalf of this argument?

First, a distinction needs to be made regarding the nature of unity. That  $x$  is *functionally unified* simply means the various parts of  $x$  work together to accomplish some end. Artifacts and aggregates, for example, can be spoken of figuratively as functionally unified. But notice that since what is at work here is efficient causation,  $x$  literally has no objective function. Rather, the parts of  $x$  are assembled to imitate a function in the artificer's mind. This type of unity is external to  $x$ . However, metaphysical unity is distinct as it pertains to real (especially living) substances. Accordingly,  $x$  is *metaphysically unified* just in case  $x$  is a mereologically simple substance (a substance that has no separable parts). Here the principle of unity is internal.

In defense of P1, consider that, as already noted, Hasker holds that emergent causal powers are "already implicit in the physical 'stuff,' otherwise their emergence would sheerly be magical" (Hasker 2004, 112). The same thing should be said regarding features of the soul, such as unity. Recall that the only kind of unity a mere aggregate of parts can have is functional unity. However, the soul is metaphysically unified. Indeed, Hasker's own argument from the unity of consciousness established this (Hasker 1999, 122–146; 2010, 181–183). So, in what way is the metaphysical unity of the soul implicit in an aggregate of parts that is merely functionally unified?

This issue is close to the concern raised by the binding problems of consciousness. Consciousness is unified in a variety of ways, but brains (at least on atomistic views, such as Hasker's) are not. Therefore, it seems difficult if not impossible for a brain to have a unified consciousness (Bayne 2010, ch. 10). The problem for emergent dualism is that the brain is not metaphysically unified but atomistic and therefore cannot ground facts regarding the unity of the soul. There just is no fact regarding the metaphysical unity of the brain that could ground the fact that the soul is metaphysically unified. The brain just isn't metaphysically unified, but is, as Hasker says, an aggregate of atomic simples. Consequently, there is nothing true of the brain that can ground the truth of the metaphysical unity of the soul.

Here is another way to get at this objection. The unity of an aggregate, such as the brain, is one in which the relations among the parts are external relations. This is evident by the fact that the parts of the aggregate are separable parts. However, the unity of the soul is such

that any differentiation of faculties or powers within the soul must stand in internal relations to the soul itself. This is evidenced by the fact that these inseparable parts – faculties of intellect, emotion, will, and so on – cannot exist outside the whole of which they are inseparable parts. In turn, this raises a new way to look at both the sorites and combination problems together. It is implausible that by adding a single atomic simple we get a new whole constituted by internal relations from a subvenient structured object (the brain) constituted by external relations.

What could be said in reply? Perhaps Hasker could avoid this problem by widening the gap between body and soul, such that the facts about the unity of the soul are not grounded in facts about the brain. This comes at a cost. Hasker would lose an argument he has endorsed for favoring emergent dualism over nonemergent substance dualism. Namely, that his view makes better sense of the soul's dependence on the brain in cases of commissurotomy and associative identity disorder (Hasker 2010). Second, widening the gap between body and soul pushes against the strong body–soul interaction emergent dualism promises to provide. Again, Hasker would lose a motivating factor for his view. Widening the gap between body and soul is not an option. Consequently, emergent dualism faces a combination problem.

### 5.3 Conclusion

I have argued that emergent dualism faces a number of problems that nonemergent substance dualism evades. Perhaps these problems can be solved. Although, skeptical, I am curious how defenders of emergent dualism might respond. Until then I think we have compelling reasons to reject emergent dualism, reasons which do not extend to substance dualism.<sup>14</sup>

### Notes

1. It is worth noting that those who adopt an Aristotelian/Thomistic view of chemical change and substances have no reason to adopt the mechanistic and emergent view. See, for example, Richard J. Connell (1988).
2. Both versions of the argument appear in Descartes, *Meditations on First Philosophy*, Sixth Meditation. The first version is in CSM II 54, and the second version appears in CSM II 59. All references to Descartes work are from Descartes 1988 *Descartes Selected Philosophical Writings*, translated by John Cottingham, Robert Stoothoff, and Dugald Murdoch (Cambridge: Cambridge University Press, 1988), standardly referred to as CSM.
3. René Descartes, *Meditations on First Philosophy*, Sixth Meditation, 81; (CSM II 56). See also *Discourse on Method* 1637, part 5: 59 (CSM I 141), and *Meditations on First Philosophy*, Objections and Replies 228 (CSM II 160).
4. See for example Hoffman (1986). For a critical treatment of Hoffman, see Rozemond (1998). For a response to Rozemond, see Skirry (2005).
5. The “ghost in the machine” is a phrase coined by Gilbert Ryle (1949) to describe Descartes's mind-body dualism. The phrase was intended to portray the mysterious notion of mind/brain interaction, but to this day is often mistaken for an accurate statement of Descartes's view.
6. René Descartes, *Letter to Regius*, CSM III, 207–208.
7. René Descartes, *Letter to Mesland*, CSM III 243.
8. See for example, Descartes, *Discourse on Method*, 141; Aquinas, *Summa Contra Gentiles*, 2. 87, 3 and 4.
9. For example, Patrick Lee and Robert George (2008, ch. 2) argue that reflective self-awareness, universal concept formation, abstract reasoning, and free will must be independent of and transcend matter, which



means they could not have emerged from lower forces. Antonella Corradini (2008, 202–203) argues Hasker’s view is incompatible with holding that consciousness is dependent on purely material processes and that there are novel emergent powers at the level of consciousness.

10. This generates a problem most philosophers of mind take as a serious objection to other views, called the Too Many Thinkers problem. For a helpful introduction to the traditional Too Many Thinkers objection see Olson 2007, 29–39.
11. For a detailed account of one amazing case see Antonia Battro (2001). For a wide study see Moosa *et al.* (2013).
12. For a more detailed treatment of this argument and what I take to be its positive contribution to the ontology of human persons see Rickabaugh (forthcoming).
13. According to Jonathan Schaffer, “*x* is fundamental” just means that “nothing grounds *x*,” while “*x* is derivative” just means that “something grounds *x*.” See Schaffer (2009, 373), Rosen (2010), and Fine (2012).
14. I am indebted to several helpful comments on the first version of the paper from J. P. Moreland, Stephen C. Evans, Stuart Goetz, and Joshua Rasmussen.

## References

- Barnett, David. 2010. “You Are Simple.” In *The Waning of Materialism*, edited by Robert C. Koons and George Bealer, 161–174. New York: Oxford University Press.
- Battro, Antonia. 2001. *Half a Brain Is Enough: The Story of Nico*. Cambridge: Cambridge University Press.
- Bayne, Tim. 2010. *The Unity of Consciousness*. Oxford: Oxford University Press.
- Brentano, Franz. 1973. *Psychology from an Empirical Standpoint*. London: Routledge and Kegan Paul.
- Chalmers, David. 2015. “Panpsychism and Protopanpsychism.” In *Consciousness in the World: Perspectives on Russellian Monism*, edited by Torin Alter and Yujin Nagasawa, 246–276. Oxford: Oxford University Press.
- Chalmers, David. 2016. “The Combination Problem for Panpsychism.” In *Panpsychism: Contemporary Perspectives*, edited by G. Brüntrup and L. Jaskolla, 179–214. Oxford: Oxford University Press.
- Chisholm, Roderick. 1981. *The First Person: An Essay on Reference and Intentionality*. Minneapolis: University of Minnesota Press.
- Connell, Richard J. 1988. *Substance and Modern Science*. Houston, TX: Center for Thomistic Studies.
- Corradini, Antonella. 2008. “Emergent Dualism.” In *Psycho-Physical Dualism Today: An Interdisciplinary Approach*, edited by Alessandro Antonietti, Antonella Corradini, and Jonathan Lowe. Lanham, MD: Lexington Books.
- Descartes, René. 1988. *Descartes: Selected Philosophical Writings*, translated by John Cottingham, Robert Stoothoff, and Dugald Murdoch. Cambridge: Cambridge University Press.
- Eccles, John, and Karl Popper. 1993. *The Self and Its Brain*. London: Routledge.
- Evans, C. Stephen, and BRANDON L. Rickabaugh. 2015. “What Does It Mean to Be a Bodily Soul?” *Philosophia Christi*, 17(2): 315–330.
- Fine, Kit. 2012. “Guide to Grounding.” In *Metaphysical Grounding: Understanding the Structure of Reality*, edited by Fabrice Correia and Benjamin Schneider 37–80. Cambridge: Cambridge University Press.
- Forsdyke, Donald R. 2015. “Wittgenstein’s Certainty is Uncertain: Brain Scans of Cured Hydrocephalics Challenge Cherished Assumptions.” *Biological Theory*, 10(4): 336–332.
- Göcke, Benedikt Paul, ed. 2012. *After Physicalism*. Notre Dame, IN: University of Notre Dame Press.
- Hasker, William. 1999. *The Emergent Self*. Ithaca, NY: Cornell University Press.
- Hasker, William. 2000. “Reply to My Friendly Critiques.” *Philosophia Christi*, 2(2): 197–209.
- Hasker, William. 2004. “Emergent Dualism: A Challenge to a Materialist Consensus.” In *What About the Soul? Neuroscience and Christian Anthropology*, edited by Joel B. Green, 101–115. Nashville, TN: Abingdon Press.

- Hasker, William. 2005. "On Behalf of Emergent Dualism." In *In Search of the Soul*, edited by Joel B. Green and Stuart L. Palmer, 75–100. Downers Grove, IL: InterVarsity Press.
- Hasker, William. 2010. "Persons and the Unity of Consciousness." In *The Waning of Materialism*, edited by Robert C. Koons and George Bealer, 175–190. Oxford: Oxford University Press.
- Hasker, William. 2012a. "The Emergence of Persons." In *The Blackwell Companion to Christianity and Science*, edited by James Stump and Alan Padgett 480–490. London: Blackwell.
- Hasker, William. 2012b. "Is Materialism Equivalent to Dualism?" In *After Physicalism*, edited by Benedikt Paul Göcke, 180–199. Notre Dame, IN: University of Notre Dame Press.
- Hasker, William. 2012c. "Jaegwon Kim's Rejection of Substance Dualism." In *Philosophy and the Christian Worldview: Analysis, Assessment and Development*, edited by David Werther and Mark D. Linville, 215–228. London: Continuum.
- Hasker, William. 2014. "The Dialectic of Soul and Body." In *Contemporary Dualism: A Defense*, edited by Andrea Lavazza and Howard Robinson, 204–219. New York: Routledge.
- Hasker, William. 2015. "Why Emergence?" In *The Ashgate Research Companion to Theological Anthropology*, edited by Joshua Farris and Charles Taliaferro, 151–162. Burlington, VT: Ashgate.
- Hasker, William. 2016. "Do my Quarks Enjoy Beethoven?" In *Neuroscience and the Soul: The Human Person in Philosophy, Science, and Theology*, edited by Thomas M. Crisp, Steve L. Porter, and Gregg A. Ten Elshof, 13–40. Grand Rapids, MI: Eerdmans.
- Hoffman, Paul. 1986. The Unity of Descartes' Man. *The Philosophical Review*, 95: 339–369.
- Koons, Robert, and George Bealer, eds. 2010. *The Waning of Materialism*. Oxford: Oxford University Press.
- LaRock, Eric. 2013. "From Biological Naturalism to Emergent Subject Dualism." *Philosophia Christi*, 13(1): 97–118.
- Lavazza, Andrea, and Howard Robinson, eds. 2013. *Contemporary Dualism: A Defense*. New York: Routledge.
- Lee, Patrick and Robert P. George. 2008. *Body-Self Dualism in Contemporary Ethics and Politics*. New York: Cambridge University Press.
- Lewin, Roger. 1980. "Is Your Brain Really Necessary?" *Science*, n.s., 210(4475): 1232–1234.
- Merricks, Trenton. 2001. *Objects and Person*. Oxford: Oxford University Press.
- Moosa, A. N., A. Gupta, L. Jehi, A. Marashly, G. Cosmo, D. Lachhwani, et al. 2013. "Longitudinal Seizure Outcome and Prognostic Predictors after Hemispherectomy in 170 Children." *Neurology*, 80(3): 253–260.
- Olson, Eric. 2007. *What Are We? A Study in Personal Ontology*. Oxford: Oxford University Press.
- Rickabaugh, Brandon. forthcoming. "Emergent Dualism Cannot Save the Soul, But Neo-Aristotelianism Might."
- Rosen, Gideon. 2010. "Metaphysical Dependence, Grounding and Reduction." In *Modality: Metaphysics, Logic, and Epistemology*, edited by Bob Hale and Aviv Hoffmann, 109–135. Oxford: Oxford University Press.
- Rozemond, Marleen. 1998. *Descartes's Dualism*. Cambridge, MA: Harvard University Press.
- Ryle, Gilbert. 1949. *The Concept of Mind*. London: Hutchinson.
- Schaffer, Jonathan. 2009. "On What Grounds What." In *Metametaphysics: New Essays on the Foundations of Ontology*, edited by David Manley, David J. Chalmers, and Ryan Wasserman, 347–383. Oxford: Oxford University Press.
- Searle, John. 2008. "The Self as a Problem in Philosophy and Neurobiology." In *Philosophy in a New Century: Selected Essays*, 137–151. Cambridge: Cambridge University Press.
- Skirry, Justin. 2005. *Descartes and the Metaphysics of Human Nature*. New York: Thommes-Continuum Press.
- Skrbina, David. 2005. *Panpsychism in the West*. Cambridge, MA: MIT Press.

- Unger, Peter. 2004. "The Mental Problems of the Many." In *Oxford Studies in Metaphysics*, edited by Dean Zimmerman, vol. 1, 196–222. Oxford: Clarendon Press.
- Willard, Dallas. n.d. "Non-Reductive and Non-Eliminative Physicalism?" Accessed June 7, 2017. <http://www.dwillard.org/articles/artview.asp?artID=48>.
- Zimmerman, Dean. 2011. "From Experience to Experiencer." In *The Soul Hypothesis*, edited by Mark Baker and Stewart Goetz, 169–196. London: Continuum Press.

# Debating Thomistic Dualism

# Aquinas on the Human Soul

EDWARD FESER

The biggest obstacle to understanding Aquinas's account of the soul may be the word "soul." On hearing it, many people are prone to think of ghosts, ectoplasm, or Descartes's notion of *res cogitans*. Of course, none of these has anything to do with the soul as Aquinas understands it. But even the standard one-line Aristotelian-Thomistic characterization of the soul as the form of the living body can too easily mislead. When those unfamiliar with Aristotelian metaphysics hear the word "form," they are probably tempted to think in terms of shape or a configuration of parts, which is totally wrong. Or perhaps they think of it in Platonic terms, as an abstract universal that the individual human being participates in – also totally wrong. Or they suppose that if the soul is the form of the living body then it is terribly mysterious, and perhaps even incoherent, to say (as Aquinas does) that the soul can persist beyond the death of the body. Totally wrong again. Whether or not one agrees with Aquinas's account of the soul, when rightly understood it is not really all *that* mysterious how the soul can both be the form of the body and nevertheless survive the death of the body. Or so I will argue.

So, what I propose to do in what follows is to avoid using the word "soul" as long as possible, and at first talk instead about what it is, on Aquinas's view, to be a human being. The first thing to say about that is that a human being is a kind of substance. Part of what this entails is that a human being is not *more* than a substance. In particular, a human being is not a composite of *two* substances, as Descartes thought. Part of what it entails is that a human being is not *less* than a substance. In particular, a human being is not a mere collection of psychological attributes, as contemporary theorists of personal identity inspired by John Locke's "continuity of consciousness" account might suggest. And part of what it entails is that a human being is a composite of substantial form and prime matter, as Aquinas, being a good Aristotelian, held all material substances are.

The second thing to address is what *kind* of substance a human being is, and Aquinas's answer, like Aristotle's, is that a human being is a rational animal. What that entails is, among other things, that a human being is neither an angel nor an ape but something in between. That is to say, it entails the falsity both of Platonic and Cartesian accounts of

human nature on the one hand, and materialist accounts on the other. Human beings are at the same time both far more closely tied to matter than Platonists and Cartesians suppose, and far less closely tied to matter than materialists suppose.

Once all of this is set out we can return to the question of what the soul is and understand how it can be that the soul is the form of the body and yet survives the death of the body. It is sometimes suggested that Aquinas's account of the soul is very strange compared to its allegedly more straightforward Cartesian and materialist rivals. I would suggest that it is not Aquinas's account that is strange but rather human nature that is strange. Aquinas merely captures that strangeness faithfully, whereas Cartesian and materialist views seek to explain it away. Angels and apes are, after all, easier to understand than is something which is a little of both. What Aquinas is telling us is that we are neither angels nor apes, but rather precisely that hybrid of both. The message is odd but true. In any event, we shouldn't blame the messenger.

## 6.1 What a Substance is

Consider a stone of the sort you might pluck from a river bed. Suppose it is solid, gray in color, round in shape, with a smooth surface and weighing two ounces. The solidity, grayness, roundness, smoothness, and weight are *attributes* of the stone, and the stone itself is the *substance* which bears these attributes. The attributes exist *in* the stone whereas the stone does not exist *in* any other thing in the same sense. Substances, in general, just are the sorts of things which exist in themselves rather than inhering in anything else, and which are the subjects of those attributes which do of their nature inhere in something else. This is true of material substances like stones, and it is true of angels, which as creatures of pure intellect are immaterial substances.

What concerns us here, though, are material substances in particular. One way material substances differ from immaterial substances is that they are *perishable*. An angel would be annihilated if God stopped preserving it in being, but it has no *inherent tendency* toward corruption. Hence nothing in the natural order can destroy it. Material substances, by contrast, can be destroyed by other natural objects because they do have an inherent tendency toward corruption. Another way material substances differ from immaterial substances is that there can be multiple instances of the same species of material substance. By contrast, each angel is in Aquinas's view the unique member of its own species.

What makes it possible for material substances to exhibit these features is that they are composites of *form* and *matter*. Consider once again the stone of our example. Its solidity, grayness, roundness, smoothness, and weight are among the forms it exhibits. (Note that while the round shape is among the forms the stone exhibits, it is not the only one. All shapes are forms, but not all forms are shapes.) Now, a form of itself is universal rather than particular. A form like solidity or roundness can exist not only in the particular stone you've plucked from the riverbed here and now, but in other stones that exist at other times and places, and in things other than stones. A form is also of itself imperishable. If you were to crush the stone into dust, or even if you were to destroy every solid and round thing that exists, you wouldn't thereby destroy solidity and roundness themselves, since they could in principle come to exist in some new object.

Now, matter is the principle by virtue of which a form, which is otherwise universal, is "tied down," as it were, to a specific individual thing, time, and place. And it is the principle

by virtue of which a thing having a certain form, which is of itself imperishable, is susceptible of perishing. For the matter of a thing is essentially its *potential* to receive form, whereas the form of a material thing is what *actualizes* its matter so as to make of it a concrete thing of specific kind. That is to say, the distinction between matter and form is a special case of the more general Aristotelian distinction between potentiality and actuality. It is because matter just is the potentiality to receive form that material things can perish, since a thing's matter is never "locked on" to the form of that thing. It is always ready in principle to take on some other form instead.

If we abstract from our notion of matter *all* form, leaving nothing but the pure potential to receive form, we arrive at the idea of *prime matter*. Matter having some form or other is *secondary matter*. There is a corresponding distinction between kinds of form. A form which makes of prime matter a concrete *substance* of a certain kind is a *substantial form*. A form which merely modifies some secondary matter – matter which already has a substantial form – is an *accidental form*.

Needless to say, much more could be said about all of these concepts, but the one we need to examine in a little more detail for present purposes is the notion of substantial form.<sup>1</sup> The distinction between substantial form and accidental form is illuminated by comparison with the different but related Aristotelian distinction between *nature* and *art* – that is to say, between natural objects on the one hand, and everyday artifacts on the other. Hence, consider a *liana vine* – the kind of vine Tarzan likes to swing on – as an example of a natural object. A *hammock* that Tarzan might construct from living liana vines is a kind of artifact, and not a natural object. The parts of the liana vine have an inherent tendency to function together to allow the vine to exhibit the growth patterns it does, to take in water and nutrients, and so forth. By contrast, the parts of the hammock – the liana vines themselves – have no inherent tendency to function together as a hammock. Rather, they must be arranged by Tarzan to do so, and left to their own devices – that is to say, without pruning, occasional rearrangement, and the like – they will tend to grow the way they otherwise would have had Tarzan not interfered with them, including in ways that will impede their performance as a hammock. Their natural tendency is to be liana-like and not hammock-like; the hammock-like function they perform after Tarzan ties them together is extrinsic or imposed from outside, while the liana-like functions are intrinsic to them.

Now the difference between that which has such an intrinsic principle of operation and that which does not is essentially the difference between something having a substantial form and something having a merely accidental form. Being a liana vine involves having a substantial form, while being a hammock of the sort we're discussing involves instead the imposition of an accidental form on components each of which already has a substantial form, namely the substantial form of a liana vine. A liana vine is, accordingly, a true *substance*, as Aristotelian philosophers understand substance. A hammock is not a true substance, precisely because it does not qua hammock have a substantial form – an *intrinsic* principle by which it operates as it characteristically does – but only an accidental form. In general, true substances are typically natural objects, whereas artifacts are typically not true substances. A dog, a tree, and water would be true substances, because each has a substantial form or intrinsic principle by which it behaves in the characteristic ways it does. A watch, a bed, or a computer would not be true substances, because each behaves in the characteristic ways it does only insofar as certain accidental forms have been imposed on them from outside. The true substances in these cases would be the raw materials (metal, wood, glass, etc.) out of which these artifacts are made.

It is important to emphasize, however, that the correlation between what occurs “in the wild” and what has a substantial form, and the correlation between what is human-made and has only an accidental form, are only rough correlations. For there are objects that occur in nature and apart from any human intervention and yet have only accidental forms rather than substantial forms, such as piles of stones that gradually form at the bottom of a hill, tangles of seaweed that wash up on the beach, and beaver dams. And there are human-made objects that have substantial forms rather than accidental forms, such as babies (which are in an obvious sense made by human beings), water synthesized in a lab, and breeds of dog. Of course, no one would be tempted in the first place to think of these as “artifacts” in the same sense in which watches and computers are artifacts. But even objects that are “artificial” in the sense that they not only never occur “in the wild” but require significant scientific knowledge and technological expertise to produce can count as having substantial forms rather than accidental forms. Eleonore Stump suggests Styrofoam as a possible example (Stump 2003, p. 44).

Stump’s rationale is that it seems to be essential to a thing’s having a substantial form that it has properties and causal powers that are irreducible to those of its parts (Stump 2006). Hence water has properties and causal powers that hydrogen and oxygen do not have, whereas the properties and causal powers of an ax (to borrow Stump’s example) seem to amount to nothing over and above the sum of the properties and powers of the ax’s wood and metal parts. When water is synthesized out of hydrogen and oxygen, what happens is that the prime matter underlying the hydrogen and oxygen loses the substantial forms of hydrogen and oxygen and takes on a new substantial form, namely that of water. By contrast, when an ax is made out of wood and metal, the matter underlying the wood and the matter underlying the metal do not lose their substantial forms. Rather, while maintaining their substantial forms, they take on a new accidental form, that of being an ax. The making of Styrofoam, Stump suggests, seems to be more like the synthesis of water out of hydrogen and oxygen than it is like the making of an ax. For Styrofoam has properties and causal powers which are irreducible to those of the materials out of which it is made, and which therefore indicate the presence of a substantial form and thus a true substance.

But there is one further complication to add to the story. On Aquinas’s account, a substance’s properties “flow” or follow from its having the substantial form it does. Being four-legged, for example, flows or follows from having the substantial form of a dog. But this “flow” can, as it were, be blocked. For instance, a particular dog might, as a result of injury or genetic defect, be missing a leg. It wouldn’t follow from its missing that leg that being four-legged is not after all a true property of dogs, nor would it follow that this particular creature was not really a dog after all. Rather, it would be a *damaged or defective instance* of a dog. When determining the characteristic properties and causal powers of some kind of thing, then, we need to consider the *paradigm* case, what that kind of thing is like when it is in its mature and normal state.

So, a thing counts as a true substance when, in its mature and normal state, it exhibits certain properties and causal powers that are irreducible to those of its parts. It is in this sense that a human being is a substance. But what *kind* of substance is that?

## 6.2 What a Rational Animal is

The answer is that a human being is a rational animal. That is to say, a human being is the kind of substance which, in its mature and normal state, exhibits both the properties and



causal powers characteristic of animality and those characteristic of rationality. Let me elaborate.

An animal is, of course, a kind of living thing. Living things, for the Aristotelian, are to be understood as substances which exhibit *immanent* causal processes as well as *transeunt* (or “transient”) causal processes. Immanent causal processes begin and remain within the agent or cause (though they may also have some external effects), and typically they in some way involve the fulfillment or perfection of the cause. Transeunt causal processes, by contrast, are directed entirely outwardly, from the cause to an external effect. An animal’s digestion of a meal would be an example of an immanent causal process, since the process begins and remains within the animal and serves to fulfill or perfect it by allowing it to stay alive and grow. One boulder knocking another off the side of a cliff would be an example of a transeunt causal process. Living things can serve as transeunt causes, but what is characteristic of them is that they are also capable of immanent causal processes in a way that nonliving things are not. A living thing can undertake activity that is *perfective* of it, that *fulfills* it or *furtheres its own good*, while nonliving things cannot do this.

Aquinas, like other Aristotelians, distinguishes between three basic kinds of living thing. The first is the *vegetative* kind. There are a great many forms vegetative life can take, but what makes them all vegetative is that they carry out three basic sorts of operation. First, they take in nutrients so as to preserve themselves in existence; second, they go through a growth cycle, and third, they reproduce themselves. If, in their mature and normal state, they do all that but do nothing *more* than that, nothing that is not a mere variation on these activities, then they are *merely* vegetative.

The second kind of living thing is the *animal* or *sensory* kind. Animal forms of life incorporate the basic functions of vegetative forms of life – nutrition, growth, and reproduction – but on top of that carry out three further and distinctive sorts of operation. First, they take in information via specialized sense organs, and in many cases can later bring to mind sense images of the things revealed by sensation; second, they exhibit appetites or inner drives, such as an impulse to pursue or avoid something sensation has revealed to them; and third, they have the power of locomotion or self-movement, by which their appetites might be efficacious in actually getting them toward or away from what sensation has revealed. Animals too can come in a very wide variety of forms, but if in its mature and normal state a thing carries out the activities described but does nothing more than that, nothing that is not a mere variation on those activities, then it is *merely* animal.

The third kind of living thing is the *human* kind. Human life incorporates the basic functions of animal or sensory life, and thus of vegetative life – nutrition, growth, reproduction, sensation, appetite, and locomotion – but on top of that carries out two further and distinctive sorts of operation. The first is intellectual activity, which essentially involves the capacity to form abstract or universal concepts, the capacity to put these concepts together into complete thoughts or propositions, and the capacity to reason from one proposition to another in accordance with standards of logical inference. The second is volition or will, which is the capacity to pursue what the intellect judges to be good or avoid what the intellect judges to be evil. Now, to be a rational animal just is to have the capacities of intellect and will as well as the capacities characteristic of animality. Anything which, in its mature and normal state, carries out these activities – nutrition, growth, reproduction, sensation, appetite, locomotion, intellection, and volition – is a human being or rational animal.

Now, like other Aristotelians, Aquinas takes each of these forms of life to be irreducibly different. Living things of any sort are irreducible to inorganic phenomena; animal life is

irreducible to merely vegetative life; and rational or human life is irreducible to vegetative or animal life. This idea is, of course, routinely dismissed today as having been refuted by modern science. However, Aristotelians have offered arguments which purport to show that science has done no such thing.<sup>2</sup> Furthermore, there are at least three considerations that should give pause even to those not already convinced by the distinctively Aristotelian arguments. First, that it is by no means obvious that the powers of the rational form of life can be reduced to those of the merely sensory or animal form is evident from the well-known difficulties facing attempts by contemporary philosophers of mind to provide a naturalistic account of the propositional attitudes (i.e., belief, desire, and the like). Second, that it is by no means obvious that the powers of the sensory form of life can be reduced to those of the vegetative form is evident from the intractability of the “qualia problem” (also known as the “hard problem of consciousness”), also much discussed in contemporary philosophy of mind. And that the organic in general cannot be reduced to the inorganic is evident from the difficulties facing attempts to provide a naturalistic analysis of the notion of biological function, as well as the absence of any plausible naturalistic account of the origin of life. These are *exactly* the sorts of problems you’d expect contemporary science and philosophy would face if the Aristotelians had been right all along.

Be that as it may, there is, for Aquinas, more to the uniqueness of human life than its irreducibility. Vegetative life is irreducible to what is inorganic, and merely animal life is irreducible to vegetative life, but merely animal and vegetative forms of life are nevertheless as corporeal as inorganic phenomena are. Human beings are different. Their distinctive activities are irreducible in a special way in that they are *incorporeal*, not essentially tied to any bodily organ. In particular, our intellectual powers are immaterial in a strict sense that entails that they are not the powers of any bodily organ, not even the brain.

There are a variety of arguments that Thomists have given for this conclusion, but for present purposes I’ll just mention one of them. The core idea of the argument goes back at least to Plato’s “affinity argument,” but it has been given an especially powerful expression in recent philosophy by the late James Ross (1992, 2008, ch. 6), and I have defended Ross’s version at some length myself (Feser 2013). Simplifying greatly, the argument is this: (1) All formal thinking is determinate, but (2) no corporeal state or process is determinate, so (3) no formal thinking is a corporeal state or process. By “formal thinking” what is meant is thinking that conforms to patterns of the sort familiar from mathematics and logic, such as adding, subtracting, squaring, reasoning via *modus ponens* or *modus tollens*, and so forth. By saying that such thinking is “determinate,” what is meant is that there is a fact of the matter about whether one is really employing one of those patterns of reasoning rather than another. For example, when you reason from the premises that *if it is raining then the streets are wet* and *it is in fact raining*, to the conclusion that *the streets are wet*, it is just a fact of the matter that you are using *modus ponens* rather than some other pattern of reasoning. There is no ambiguity or indeterminacy about it.

Ross offers several considerations in support of premise (1), including arguments to the effect that the premise cannot coherently be denied. For one thing, to defend a rejection of premise (1) will require making use of the very patterns of reasoning the rejecter denies we ever really apply. (For instance, you will have to apply forms of reasoning like *modus ponens* in an argument to the effect that we never determinately reason according to *modus ponens*.) For another thing, even to deny premise (1) requires that one determinately grasp precisely the patterns one denies we have a determinate grasp of. (For instance, you have determinately or unambiguously to grasp what *modus ponens* is in the first place and how it

differs from other patterns of reasoning in order to go on to deny that we ever determinately or unambiguously grasp what *modus ponens* is.)

In defense of premise (2), Ross draws on a number of thought experiments from contemporary analytic philosophy, including Kripke's "quus" paradox and Quine's "gavagai" example. The upshot of these thought experiments is that no collection of facts about physiology, behavior, bodily sensations, or mental imagery, and so forth, could in principle by themselves determine that you are having a thought with *this* particular content rather than *that* one. All such facts and collections of facts are inherently indeterminate in the sense of ambiguous between different possible attributions of content. It follows that no such collection of facts could by itself ever determine that it is *modus ponens* (say) rather than some other pattern of reasoning that you are employing in any particular case.

Suppose this or some other argument for the incorporeality of the intellect is correct. Then, given what has already been said, it follows that what a human being is, is the kind of substance which in its mature and normal state carries out both corporeal *and* incorporeal activities. Some of what we do is the sort of thing that nonhuman animals do, and some of what we do is the sort of thing that angels do. That is very unusual in nature. Everything else there is falls on one side or other of the divide between the corporeal and the incorporeal. To be a human being just is to be the kind of substance which straddles that divide, which has a foot in both worlds, as it were. That might seem not only unusual but impossible. How could one thing be both corporeal and incorporeal? Isn't that contradictory? But it is not contradictory, any more than it is contradictory to say that the Great Pyramid is both triangular and square. The Great Pyramid is triangular on its sides and square on its bottom. It can be both triangular and square if it is triangular and square *in different respects*. It is in that way that a human being can be both corporeal and incorporeal.

Now, recall that a substance can fail to manifest some characteristic property or power if it is damaged. Again, though every dog by virtue of its substantial form will tend to have four legs, this tendency is frustrated in a dog which has lost a leg in an injury, or which never developed one of its legs in the first place due to some genetic defect. And this failure of a substance to manifest all its properties or powers can be far more extreme than that. Imagine, for example, a dog which has lost all four of its legs. Or imagine a dog which as a result of some horrific accident has lost not only its legs, but also its eyes, ears, nose, teeth and tongue, most of its skin and many of its internal organs, and in effect has been reduced to its bare vegetative functions, kept alive on a slab in a laboratory somewhere via a feeding tube. It would still be a dog rather than some other kind of substance. It hasn't *literally* become a vegetable, and if you were somehow able to treat it in such a way that it could regenerate its lost organs, the organs it would grow back would all be dog organs. But it would be a severely damaged substance, an incomplete substance. It would be a dog reduced to the stub of a dog, to the bare minimum consistent with there being a dog at all. And if you destroyed *all* of its corporeal functions, including the most rudimentary vegetative functions, there would be no dog left at all.

A human being can, of course, suffer damage of a similar sort. People lose arms, legs, eyes, ears, and other organs. It is also possible for a human being to suffer the sort of horrific damage I imagined the dog of my example suffering, as does the protagonist of Dalton Trumbo's novel *Johnny Got His Gun*. We'd still have a human being in this case. We wouldn't have a different substance, but the very same substance in a radically damaged or incomplete state. Now I noted that when the last of a dog's corporeal functions are destroyed, the dog itself is destroyed. The reason is that there's nothing *more* to a dog than

its corporeal functions. But the same is not true of a human being. If you destroy all of a human being's corporeal functions, it doesn't follow that the human being is gone, because a human being is the kind of substance which has *incorporeal* activities as well as corporeal ones.

Hence when the human *body* is destroyed, it doesn't follow that the human *being* is destroyed, that the *substance* is destroyed. It is *not* destroyed, any more than a dog is destroyed when you reduce it to its vegetative functions. Rather, it continues on as a radically incomplete substance, as the stub of a human being, reduced to the bare minimum consistent with there being a human being at all. The difference with the case of the dog is that whereas the bare minimum consistent with there being a dog is something corporeal, the bare minimum consistent with there being a human being is something *incorporeal*. It is the human substance reduced to its intellectual and volitional functions, with all the corporeal functions being prevented from manifesting. This is why, for us unlike dogs, death is not the end. Death is more like an amputation than it is like annihilation. It is a "full body amputation," as it were.

That sounds like a pretty severe amputation, and it is even more severe than you might expect given that at least the intellect survives the death of the body. On Aquinas's view, though concepts are different from sense images and intellect therefore a different faculty than sensation or imagination, our intellects nevertheless require sensation as a source of information and imagery as an aid to abstract thought. This is why, even though the concept *triangularity* differs from a mental image of a triangle or a mental image of the word "triangle," you nevertheless cannot entertain the concept without calling to mind images like those. But sensation and imagination are corporeal activities, requiring a brain. So, since the brain is among the organs "amputated" at death, the intellect cannot do much on its own despite surviving the "amputation."

The following analogy might help. In the movie *Guardians of the Galaxy*, there is a character named Groot who is a sentient, thinking, walking and talking tree able to regenerate any organs he might lose in battle. At one point in the movie he is blown to bits and reduced to a single twig. He is not thereby destroyed, however. The twig is planted in a flower pot and he very slowly begins to grow back his arms, legs, eyes, mouth, and other body parts. Presumably, while still a mere twig and before these organs have grown back, he not only lacks locomotion but also sentience and thought. These powers are all dormant, as it were – ready to "flow" from Groot's substantial form but prevented from doing so by the severe damage he has suffered. What is left of a human being after death is like the twig. The differences are, first, that unlike the twig, what is left of a human being after death is *incorporeal*; and second, that unlike the twig, what is left of a human being after death lacks the power to regenerate the lost, corporeal organs on its own. Their restoration would require divine intervention.

So, death is far from the liberation Plato implies that it is. Since we are rational *animals*, the absence of our animal, corporeal powers cannot fail to be an extremely grave diminution. All the same, since we are *rational* animals and rationality is essentially *incorporeal*, even death is not the end of the human being, contrary to what the materialist supposes.

As I have said, this is very strange. Indeed, human beings, as hybrids of the corporeal and the *incorporeal*, are arguably the strangest things in all creation. We are, metaphysically speaking, real weirdos. And this is the source of many of our persistent moral difficulties. For example, where sex is concerned, we tend constantly either to overstate or to understate

its importance, and toward either licentiousness or prudery. The reason we are so prone to these extremes, I would say, has precisely to do with our highly unusual place in the order of things. Angels are incorporeal and asexual, creatures of pure intellect. Nonhuman animals are entirely bodily, never rising above sensation and appetite, and our closest animal relatives reproduce sexually. Human beings, as rational animals, straddle this divide, having, as I have said, one foot in the angelic realm and the other in the animal realm. Metaphysically, this is just barely a stable position to be in, and sex makes it especially difficult to maintain. The unique intensity of sexual pleasure and desire, and our bodily incompleteness qua men and women, continually remind us of our corporeal and animal nature, pulling us “downward” as it were. Meanwhile our rationality continually seeks to assert its control and pull us back “upward,” and naturally resents the unruliness of such intense desire. This conflict is so exhausting that we tend to try to get out of it by jumping either to one side of the divide or the other. But this is an impossible task and the result is that we are continually frustrated. And the supernatural divine assistance that would have remedied this weakness in our nature and allowed us to maintain an easy harmony between rationality and animality was lost in original sin.

So, behaviorally, we have a tendency to fall either into prudery or into sexual excess. And intellectually, we have a tendency to fall either into the error of Platonism – treating man as essentially incorporeal, trapped in the prison of the body – or into the opposite error of materialism, treating human nature as entirely reducible to the corporeal. The dominance of Platonism in early Christian thought is perhaps the main reason for its sometimes excessively negative attitude toward sex, and the dominance of materialism in modern times is one reason for its excessive laxity in matters of sex. The right balance is the Aristotelian-Thomistic position – specifically, Aristotelian-Thomistic philosophical anthropology, which affirms that man is a single substance with both corporeal and incorporeal activities; and Aristotelian-Thomistic natural law theory, which upholds traditional sexual morality while affirming the essential goodness of the body.

Other aspects of our moral condition can also be illuminated by this uneasy metaphysical straddle entailed by being a rational animal, but I’ll leave that as homework.

### **6.3 What the Soul is**

Now, I have written over eight pages discussing Aquinas’s philosophical anthropology, immortality, and even the war of the flesh and the spirit, without once using the word “soul.” But really I have been talking about the soul the whole time, and now it is time to make it explicit where it fits in. To some extent this is easy. As is well known, the word “soul” is in Aristotelian-Thomistic philosophy essentially a technical term for the substantial form of a living thing. This is why plants and animals have souls just as human beings do. There’s nothing terribly remarkable about that because plants and animals, like everything else, have substantial forms, and the soul is just a kind of substantial form. So, you could just plug in the word “soul” wherever I’ve talked about the substantial form of a human being or other living thing, and there you’ll have the Thomistic account of the soul. At one level it is no more complicated than that.

That makes it sound like the question of where the soul fits in is a matter of semantics, and I think that’s essentially correct. However, the semantics of the term are complicated, even as Aquinas uses it, so more needs to be said. In particular, the Aristotelian-Thomistic position is often expressed as the view that the human soul is the substantial form of the

living human *body*. And that is perfectly true. But it is also potentially misleading. Indeed, I think it has misled many readers, who find it terribly mysterious how the soul could be the form of the body and yet survive the death of the body. They suspect that Aquinas is trying to “pull a fast one” here, trying to have his cake and eat it too, trying to marry two irreconcilable positions, trying to be an Aristotelian philosopher and a Christian theologian at the same time and not succeeding. In fact, none of that is correct, and I think what I’ve already said so far shows why.

But let’s try to understand why people find the formulation in question puzzling, before making it explicit why they shouldn’t. What happens when they hear the sentence “The human soul is the substantial form of the living human body” is, I think, this. They know that, on the Aristotelian analysis, when a material substance like a stone loses its substantial form, nothing of that individual stone survives. The form of the stone carries on only in the sense that some other thing could always come to have the form of a stone. But that particular stone is gone for good, and there is no sense in which its form, considered as a particular thing, carries on or subsists. So far so good. Then the listener recalls that for the Aristotelian, a soul is just the substantial form of a living thing, and that the souls – that is to say, the substantial forms – of plants and nonhuman animals don’t subsist or carry on after their deaths any more than the substantial form of a stone carries on after its destruction. Again, so far so good.

But now the listener makes a mistake. He supposes that when Aquinas, following Aristotle, says that:

- 1 The human soul is the substantial form of the body.

then he is saying something that entails:

- 2 The human soul is the substantial form of a substance which is entirely bodily or corporeal.

As a result the listener is puzzled when Aquinas goes on to say that the soul persists beyond the death of the body. After all, stones, trees, and nonhuman animals are all entirely bodily or corporeal, and their substantial forms don’t carry on when the substances in question perish. And the human body is by definition bodily or corporeal. So why should *its* substantial form carry on after it perishes any more than these other substantial forms do? Hence the listener concludes that if Aquinas were consistent, then he ought to think *either* that the soul survives the death of the body but therefore is not really the form of the body but rather a substance in its own right; or that the soul is the form of the body but therefore does not persist beyond the death of the body. But to think both that it is the form of the body and that it survives the death of the body is (the listener judges) not consistent.

But in fact there is no inconsistency, because *proposition (1) simply does not entail proposition (2)*, and Aquinas would reject (2). For in Aquinas’s view, the human soul is the form of a *substance*, that substance is a *human being*, and a human being has both corporeal and incorporeal operations. Hence the soul is not the form of a substance which is *entirely* bodily or corporeal. Rather, it is the form of a substance which is corporeal *in some respects* and incorporeal *in others*. Now, those corporeal respects are the ones summed up in the phrase “the body.” Hence the soul is, naturally, the form of the body. But it simply doesn’t follow that the soul is the form of a substance which is *exhausted* by its body, that is, by its bodily operations.

This is why there is nothing terribly mysterious about why the soul, as Aquinas understands it, can persist beyond the death of the body. For the *substance* of which the soul is the form does not go out of existence with the death of the body. Rather, the *corporeal or bodily* properties and powers of that substance are no longer manifest, while *the incorporeal properties and powers continue*. To be sure, the substance in question has been severely reduced or damaged. That is why Aquinas thinks of the disembodied soul as an “incomplete substance.” But an incomplete substance is not a nonsubstance. Thus, to say that the soul persists beyond the death of the body is *not* to say that the form of a substance persists after the substance has gone out of existence (which certainly *would* be a very mysterious thing for an Aristotelian like Aquinas to say!)

I would suggest that these considerations shed light on a dispute that has arisen among Thomists in recent years over how to understand the relationship between the postmortem soul and the human being whose soul it is. There are two main views. One of them has come to be called *corruptionism*, and it holds that at death the human being ceases to exist until the resurrection, even though the soul carries on. The other view is called *survivalism*, and it holds that the human being persists in existence after death and even before the resurrection, though only as constituted by his soul. Among the many philosophers who have contributed to this debate are Patrick Toner (2009) on the corruptionist side, and David Oderberg (2012) on the survivalist side.<sup>3</sup> There are two main questions where these views are concerned. First, which of them is the most plausible view to take in light of Thomistic metaphysical principles? Second, which of them did Aquinas himself actually hold? I will focus on the first question rather than say much by way of exegesis of Aquinas’s texts. It is worth noting, however, that the two questions are not unrelated. If one of the two views is in fact more plausible in light of Aquinas’s own principles, then that is at least some evidence favoring the attribution of that view to him, and reconsidering our interpretation of those passages that seem at odds with it.

It will no doubt be obvious from what I’ve said so far that the position I favor is the survivalist one. For I have said that after death the human being persists, albeit in a severely reduced or incomplete state. Now, the survivalist view is characterized as the view that after death but before resurrection, the *soul* constitutes the human being. The way I would suggest interpreting this claim is as follows. Consider first that the soul, being a kind of form, cannot *by itself and without qualification* either subsist or constitute anything, because a form *qua form* exists only together with the substance of which it is the form. However, since the substance of which the human soul is the form is not entirely corporeal, that substance subsists and continues beyond the death of the body because the death of the body involves the cessation of only the corporeal aspect of said substance, not the incorporeal aspect. Hence, because the substance that the human soul is the form of persists beyond the death of the body, the *soul* persists beyond the death of the body.

Consider next that the substance in question (i.e., the human being) persists only in a greatly diminished and incomplete state, indeed diminished to the bare minimum possible consistent with the substance’s existing at all. So, you might say that the substance has gotten as close as anything possibly could get to being a substantial form existing all by itself. Moreover, the material side of this substance is completely gone, so that we do have a form without *matter*, even if it is not exactly a form existing apart from any *substance*. So, there is a loose sense in which we can say that we have a substantial form by itself; and there is a strict sense in which we can speak of a substantial form without a *body*, specifically. So, since the soul just is that substantial form, it is quite natural to speak of the soul existing all by

itself – even if, were we speaking more strictly, we’d speak instead of the soul continuing to exist together with the substance of which it is the form, but without the *corporeal* features of that substance being present any longer.

So, since there is a sense in which we might call this thing that persists beyond the death of the body the soul, and this thing that exists beyond the death of the body is also a substance – albeit an incomplete one – and, since this incomplete substance just is the human being in a radically impaired state, it is quite natural to say that the soul that survives death just is the human being surviving death (again, in a radically incomplete state). Now, that is just what the survivalist position says. So, the position I have been developing in this paper is essentially the survivalist position.

Now, I think that one of several very good reasons for a Thomist to take the survivalist position rather than the corruptionist position is that the corruptionist position would make Aquinas’s view as mysterious as people sometimes suppose it to be (mistakenly, as I have argued). For the corruptionist position holds, again, that at death the human being ceases to exist, even though the soul carries on. Yet the human being is the substance of which the soul is a form. So, if the human being ceases to exist at death, then that means that the substance of which the soul is the form ceases to exist at death. And in that case, how could the soul carry on? How could a form exist apart from the substance of which it is the form? Corruptionism seems to make Aquinas’s position as incoherent as its critics accuse it of being. But survivalism does not have this problem, precisely because it does *not* say that the human being ceases to exist at death.

There are two further important arguments for survivalism, one philosophical and one theological. The first is this. On Aquinas’s philosophical anthropology, just as it is the *human being* who sees, and not the eye, it is the *human being* who thinks, and not the intellect. The eye sees only in a loose sense, and the intellect thinks only in a loose sense. Now, on both the survivalist view and the corruptionist view, the intellect survives the death of the body and thought occurs as well (albeit only with divine assistance since the intellect’s normal corporeal sources of information are gone). But if thought occurs and if on Aquinas’s own principles it is strictly speaking only the human being, and not the intellect, which thinks, then it follows that there must be a sense in which the human being survives as well (Oderberg 2012, p. 8).

The second, theological argument is this. After death the soul is rewarded, punished, or purged in light of the deeds of this present life. But it makes sense to reward, punish, or purge only persons, not mere parts of persons. It makes no sense, for example, to speak of rewarding or punishing Bob’s foot or his pancreas for Bob’s good or bad deeds. But then the soul as it exists after death must in some sense be the human person existing after death, rather than a mere part of the person (see Hershenov and Koch-Hershenov 2006).

Naturally, corruptionists present arguments of their own. One of them appeals to the *weak supplementation principle*, a widely accepted thesis of mereology (the study of parts and wholes). The principle holds that a thing cannot have only a single proper part, where a proper part is a part that is less than the whole. Now, the disembodied soul is merely part of a human being. If the human being persisted after death as the disembodied soul, then the human being would exist as a single proper part of a human being, thus violating the Weak Supplementation Principle. So, concludes the corruptionist, the human being does not persist.

There are several things that could be said in response to this. Oderberg suggests that the weak supplementation principle seems like a universal truth of mereology because the



examples we tend to focus on when thinking about mereology are material objects, whose parts are spatially smaller than the whole. But what is true of material things will not necessarily be true of immaterial things, and the soul is immaterial.

It seems to me, though, that a more important point is that here as elsewhere when thinking about the metaphysics of substance, we need always to keep in mind the distinction between the properties which flow from a thing's substantial form and the actual manifestation of those properties, and the corresponding distinction between the normal or paradigm case and the aberrant case. Again, all dogs by virtue of their substantial form are four-legged, and this is not falsified by the existence of three-legged dogs, because such dogs are defective instances. The manifestation of one of the properties they would naturally tend to exhibit (four-leggedness) is in this case being blocked. And if we consider the more radically damaged dog of my example – reduced to its merely vegetative functions – we can see how a thing might be reduced to something close to a single one of its proper parts.<sup>4</sup> Now, I would suggest that this is exactly what happens in the case of the disembodied soul. The human being has been reduced to a single one of its proper parts. This doesn't violate the Weak Supplementation Principle if we interpret that principle as applying to a thing *in its mature and normal state*. For the human being qua disembodied soul is *not* in his normal state.

Another corruptionist argument appeals to Aquinas's famous remark in his Commentary on St. Paul's First Letter to the Corinthians that "I am not my soul," and similar remarks. Doesn't this entail that Aquinas himself denied that the human being survives death as a disembodied soul, and was therefore a corruptionist rather than a survivalist?

It does not entail that. It would entail that only if, when making such remarks, Aquinas was addressing the *specific* issues in dispute between corruptionism and survivalism. And he was not. I would suggest that the target of such remarks was not what we would today call survivalism, but rather Platonism. The Platonist, recall, takes the view that a person is an *entirely* incorporeal thing, which is only *contingently* related to the body. The survivalist certainly rejects that view, but it was one which would have been very familiar in Aquinas's day, and one to which many thought (and many still think) belief in the immortality of the soul tends to lead to. What Aquinas intends in making remarks like "I am not my soul" is to indicate that he rejects this Platonist view. He means "I am not *merely* my soul," or "I am not *reducible to* my soul," because the body is essential to me and thus something I would have when in my mature and normal state. It doesn't follow that there is no sense at all in which I am my soul. Nor does this follow from Aquinas's view that it is St. Peter's soul, rather than St. Peter, to whom we pray. For just as it is a person who thinks, rather than part of a person, it is only a person to whom we can intelligibly pray, not part of a person. It would make no sense to say "Left foot of St. Peter, pray for us," or "Pancreas of St. Peter, pray for us." But it does make sense to say "Soul of St. Peter, pray for us." The only way that can be true is if there is a *sense* in which St. Peter's soul just is St. Peter.

Certainly this is what Aquinas *should* say given the general metaphysics of substance in the context of which he develops his philosophical anthropology, or so I would argue.<sup>5</sup>

## Notes

1. See Feser (2014, ch. 3) for a thorough exposition and defense of Aquinas's understanding of substance.
2. For an overview of the traditional Aristotelian position on the irreducibility of the different forms of living thing, see Koren (1955); and for a recent defense, see Oderberg (2007, chs. 8–10).
3. See Oderberg (2012, n2) for a list of philosophers on each side of the debate.

4. Oderberg makes a similar point and illustrates it with the vivid example of a human being reduced to just his head, kept alive through futuristic technology as in a science fiction story. His point is well taken, but I prefer my example of the radically damaged dog, since (a) the human head example brings to mind the sorts of thought experiments characteristic of the complicated contemporary debate over personal identity, and (b) from a Thomistic point of view, even the human being preserved as a severed head is not *just* a head, but the head plus the incorporeal aspect of the human being. These factors threaten to lead the discussion down irrelevant side paths which can be avoided if we stick with the dog example.
5. For comments on an earlier version of this paper, I thank audience members at Mount Saint Mary College in Newburgh, New York, on June 5, 2015, and at Harvard University on October 2, 2015.

## References

- Feser, Edward. 2013. "Kripke, Ross, and the Immaterial Aspects of Thought." *American Catholic Philosophical Quarterly*, 87(1): 1–32. DOI: 10.5840/acpq20138711
- Feser, Edward. 2014. *Scholastic Metaphysics: A Contemporary Introduction*. Heusenstamm, Germany: Editiones Scholasticae.
- Hershenov, David, and Rose Koch-Hershenov. 2006. "Personal Identity and Purgatory." *Religious Studies*, 42: 439–451.
- Koren, Henry. 1955. *An Introduction to the Philosophy of Animate Nature*. St. Louis, MO: B. Herder.
- Oderberg, David S. 2007. *Real Essentialism*. London: Routledge.
- Oderberg, David S. 2012. "Survivalism, Corruptionism, and Mereology." *European Journal for Philosophy of Religion*, 4: 1–26.
- Ross, James. 1992. "Immaterial Aspects of Thought." *Journal of Philosophy*, 89(3): 136–150. DOI: 10.2307/2026790
- Ross, James. 2008. *Thought and World: The Hidden Necessities*. Notre Dame, IN: University of Notre Dame Press.
- Stump, Eleonore. 2003. *Aquinas*. London: Routledge.
- Stump, Eleonore. 2006. "Substance and Artifact in Aquinas's Metaphysics." In *Knowledge and Reality: Essays in Honor of Alvin Plantinga*, edited by Thomas M. Crisp, Matthew Davidson, and David Vanderlaan, 63–80. Dordrecht, Netherlands: Springer.
- Toner, Patrick. 2009. "On Hylemorphism and Personal Identity." *European Journal of Philosophy*, 19(3): 454–473.

# In Defense of a Thomistic-like Dualism

J. P. MORELAND

I confess, I am, indeed, a Peeping Thomist. That is, as Thomistic purists always remind me, my view is not Aquinas's own view; indeed, mine departs from his at crucial points. Even so, I believe my view does stand, in important ways, within a Thomistic framework or is its near cousin. Technically speaking, my position is actually an updated view in analytic philosophy drawing heavily from the late medieval Aristotelians (Des Chene 2000; Pasnau 2011). But Aquinas was, of course, a late medieval Aristotelian, and I often side with him in the disputes of his time. So I call my view a Thomistic-like dualism.

In what follows, I lay out the details of my position and I argue that it has certain advantages over physicalist treatments of the human person, and, to a lesser degree, over alternate versions of substance dualism (see the Introduction for the wide sense we give to "substance dualism"), then I respond to some objections against my position.

To begin, I want to disclose certain ontological assumptions in what follows. First, part/whole relations are important for treatments of substances, and there are two kinds of parts relevant to our discussion: separable and inseparable.

*p* is a *separable part* of some whole *W* =<sub>def.</sub> *p* is a particular, *p* is a part of *W*, and *p* can exist if it is not a part of *W*.

*p* is an *inseparable part* of some whole *W* =<sub>def.</sub> *p* is a particular, *p* is a part of *W*, and *p* cannot exist if it is not a part of *W*.

In contemporary philosophy, Brentano, Husserl, and their followers most fruitfully analyzed inseparable parts (Moreland 2002; Smith 1982) whose existence and identity come from the whole of which they are parts. The paradigm case of an inseparable part in this tradition is a (monadic) property-instance or relation-instance. Thus, if substance *s* has property *P*, the-having-of-*P*-by-*s* is (1) a property-instance of *P*; (2) an inseparable part of *s* which we may also call a mode of *s*.

Second, a *substance* =<sub>def.</sub> an essentially characterized particular that (1) has (and is the principle of unity for its) properties but is not had by or predicable of something more basic than it; (2) is an enduring continuant; (3) has inseparable parts but is not composed of separable parts; (4) is complete in species. It is an axiom of Aristotelian-substance theory that no genuine substance is composed of or contains within itself other substances. Substances are to be clearly distinguished from mereological aggregates, which are particular wholes constituted by (at least) separable parts and external relation-instances between and among those separable parts. The structure of a mereological aggregate is a specific ordering of external relation-instances between and among its various separable parts.

Finally, I accept constituent realism regarding properties (and relations), according to which properties (and relations) are universals that, when exemplified (and they need not be to exist), become constituents of the ordinary particulars that have them (Moreland 2013).

With this framework in mind, I offer my view of a human person.

## 7.1 What is Thomistic-like Dualism?

### 7.1.1 *The soul*

The human soul (hereafter, simply soul) is a simple (containing no separable parts), spatially unextended substance containing the capacities for consciousness and for animating, enlivening, and, for some, developing teleologically its body. The essence of the soul is constituted by determinate/determinable properties, namely, human personhood. Thus, being a human is a sufficient condition for being a person. The faculties of the soul (e.g., the mind, will) are inseparable parts of the soul containing a group of naturally resembling powers/capacities (see Perler 2015). The essence of the soul is in the category of property and not relation – especially external relation (see Koslicki 2008, 200–233), it grounds membership in a thing's natural kind and it should be understood in terms of Aristotelian essentialism.

The late medieval Aristotelians drew a distinction between a thick particular (the entire concrete organism; the thin particular plus accidents) and the thin particular (the essence/form, the nexus of exemplification, and an individuator, in their case, prime matter) (Pasnau 2011, 99–114). In my view, the human person is identical to his soul (the thin particular), which contains three metaphysical constituents – a human essence, exemplification, and a bare particular (see Moreland 1998).

As we turn to an analysis of the body, some words by Francisco Suárez (1548–1617) provide a fitting transition:

If a sensory soul were to intercede [between matter and the intellectual soul], then the intellective soul would be a pure principle of thought. A pure principle of thought, however, is not suited to inform the body . . . Therefore, for a rational soul to be a true form of the body, it must be the principle not only of thinking but also of the operations that are exercised by the body. (Suárez 1597)

### 7.1.2 *The body and the body–soul relationship*

In this section I offer two versions of Aristotelian-style dualism that provide two different understandings of the body and the body–soul relationship: first, a strictly metaphysical

thesis I shall call metaphysical Aristotelianism (MA), and second, what I call organicism, a more metaphysical/scientific thesis than MA that, among other things, implies certain scientific theses that while currently in disfavor, are making a comeback in recent years.

My delineation of these two distinct Aristotelian-style views has been noted by what is most likely the most authoritative treatment of the Aristotelian metaphysics of substance in the late Middle Ages – Robert Pasnau’s *Metaphysical Themes: 1274–1671* (Pasnau 2011). Says Pasnau:

Scholastic authors do offer metaphysical entities as principles of explanation on a concretely physical level, as efficient causes in competition with a corpuscular-mechanistic account of the natural world. The hylomorphic theory admits of an alternative formulation, however, as an explanatory schema at a different level of analysis, not competing with a corpuscular-mechanistic theory, but accounting for abstract, structural features of the world – in particular, the unity and endurance of substances . . . One diagnosis of the decline of scholastic thought . . . is that the scholastics lost their grip on hylomorphism as a metaphysical theory, conceiving of it instead as a concrete, physical hypothesis. (Pasnau 2011, 100–101, cf. 558–565)

### 7.1.2.1 Metaphysical late medieval Aristotelianism

According to MA, a living organism is not a mereological-aggregate system composed of separable parts, bundles of properties, nor is it a concrete organism construed as some sort of whole. Rather, it is a thin particular, namely, an essence exemplified by an individuator (usually prime matter) standing under (sub-stands) the accidental features of the organism, including its body (Pasnau 2011, 99–134). The thin particular is identical to the organism’s soul, it is mereologically simple (not composed of separable parts), metaphysically complex (containing a complex essence, exemplification, and an individuator), and hollenmerically present throughout the organism’s body (i.e., fully present to the body as a whole and fully present at each part of the body).

The thin particular played three central metaphysical roles: (1) It metaphysically grounded the special sort of synchronic unity of living things, especially compared to mereological aggregates/systems. (2) It metaphysically grounded a living thing’s ability to be a continuant, sustaining strict, absolute identity through certain changes (including part replacement in the organism’s body). (3) It provided the metaphysical ground both for placing the organism in its natural kind and for unifying that kind.

MA’s advocates clearly distinguished attempts to provide an ontological classification of the nature of various capacities and their possessors from proffering an explanation of the bodily conditions required for the exercise of those capacities; and MA advocates were clearly interested in the former, not the latter. As Dennis Des Chene points out:

The Aristotelians, while acknowledging, even insisting, on the necessity of a material basis for the instantiation and exercise of vital powers, did not seek to reduce them to complexes of powers found also in inanimate things . . . For them, the project was not to find a chemical basis for life, but to describe and classify vital powers, and then, in keeping with the scheme of Aristotelian natural philosophy, to define the genera and species of living things in terms of those powers. (Des Chene 2000, 7)

For MA, then, the body is key for both the functioning of the thin particular’s (soul’s) powers and the actualization of its various capacities. Speaking of the human soul, Des

Chene observes, “The human soul is not merely joined with the body in fact. It is the *kind* of soul which, though capable of separate existence . . . nevertheless by its nature presupposes union with a body, and moreover with a particular kind of body, a body with organs, in order to exercise all its powers – even reason” (Des Chene 2000, 71). Elsewhere, Des Chene notes: “Even the intellect requires, so long as the soul is joined with a body, a certain disposition of the brain” (Des Chene 2000, 96).

Thus, MA remains consistent with, and even entails the search for specific neurological causal/functional/dependency conditions associated with the actualization of the soul’s capacities for consciousness. Such a search would not provide information about the intrinsic nature of the capacity or the property it actualizes (e.g., pain) nor about the possessor of that capacity (the soul, not the brain). But it would provide information about the bodily conditions required for its actualization. This form of dualism (as with organicism; see below) is quite at home with the existence of contemporary neurological findings.

Moreover, while physicalism may be fundamental to a neuroscientific research program, in the specific sense in which this is true (there will be neurophysiological conditions in deep causal/functional dependency with the various capacities for life and consciousness), physicalism is also fundamental to an MA research program.

#### 7.1.2.2 Scientific late medieval Aristotelianism (organicism)

A second view among the late medieval Aristotelians, distinct from MA, which I call “organicism,” has certain things in common with vitalism, though classifying it as such is a matter of controversy; more on this later. Pasnau notes that on this view, the soul “plays a straightforwardly causal role, explaining both the behavior and the physical structure of an animal’s body” (Pasnau 2011, 558, cf. 549, 560–565). In this sense, the soul is not only the formal/essential cause of the body, but also becomes (1) an internal efficient first-moving cause of the development and structure of the body, and (2) the teleological guide for that development and structure (thus, form determines function). As with MA, organicism identifies the organism with the thin particular, the soul.

Here, the soul is a substance with an essence or inner nature containing, as a primitive unity, a complicated, structural arrangement of capacities/dispositions for developing a body. Taken collectively this entire ordered structure is unextended, *holenmerically* present throughout the body, and constitutes the soul’s principle of activity that governs the precise, ordered sequence of changes the substance will (normally) undergo to grow and develop.

The various physical/chemical parts and processes (including DNA) are tools – instrumental causes – employed by higher-order biological activities to sustain the various functions grounded in the soul. Thus, the soul is the first efficient cause of the body’s development as well as the final cause of its functions and structure, which are internally related to the soul’s essence (see Kaiser 2014). The functional demands of the soul’s essence determine the character of the tools, but they, in turn, constrain and direct the various chemical processes taking place in the body as a whole. Organicism, then, implies the organism as a whole (the soul) is ontologically prior to its bodily parts. Along with the soul’s *holenmeric* presence in and to the body, this understanding of the soul’s essence makes it very similar to the notion of information, as used in biology today.

While scientists can usually tell you what (nonpropositional) information does or how it is measured, they find it difficult to define. At a minimum, information is the reduction of

possibilities (Dembski 2014, 17–19). If I tell you I live in California, for example, that small bit of information leaves open many possibilities. But if I give you my town and street address, this new information eliminates a significant number of possibilities.

In addition, biologist Jonathan Wells claims that information is a *sui generis*, irreducible entity that is an immaterial, unextended, multifaceted blueprint for organismic development. As such, information is present/available to the organism as a whole, and fully present/available to each cell integrated into the organism (organisms have numerous bacterial cells that, while they play an important symbiotic role with organisms, they are not, strictly speaking, “parts” of the organism) and, as such, teleologically guides the organism’s development toward maturity (personal email communication, Jonathan Wells, April 3, 2015). If information is not identical to an Aristotelian essence, it seems at least to be very close.

A minority of biologists are returning to a type of Aristotelian essentialism (Webster and Goodwin 1996, 3–100, especially 17–25), and a growing minority of biologists are reintroducing irreducible teleology into the field (Toepfer 2012; Aranda-Anzaldo 2011).

Moreover, an organism’s parts are inseparable parts that stand in internal relations to other parts and to the soul’s individuated essence; they are literally functional entities constituted by their role in the organism as a whole. The body is developed and grows teleologically, by means of law-like developmental events, rooted in the internal essence of the soul. The first efficient cause of the characteristics of an organism’s body is its soul (containing a blueprint or information in its individuated essence); the various body parts, including DNA and genes, are important instrumental causes the soul uses to produce the traits that arise. This sort of view, along with the holism with which it is associated is also gaining ascendancy in biology (Denton, Kumaramanickavel, and Legge 2013).

Since my preferred view is organicism, I assume it in what follows. In closing this section, I summarize the Thomistic-like organicist view and present two final reflections entailed by it: (1) the organism as a whole (the soul) is ontologically prior to its inseparable parts; (2) the parts of the organism’s body stand in internal relations to other parts and to the soul’s essence; they are literally functional entities (the heart functions literally to pump blood); (3) the body’s operational functions are rooted in the soul’s internal structure; the internal structure or essence is the blueprint, the information responsible for the body’s structure and functions; (4) the body is developed and grows teleologically as a series of developmental events occurring in a law-like way, rooted in the internal essence of the human soul; (5) the first efficient cause of the characteristics of the human body is the soul; various body parts, including DNA and genes, are important instrumental causes the soul uses to produce the traits that arise; (6) the body is a mode of the soul (the soul could exist without the body but not conversely; a body without a soul is a corpse); as such it is an ensouled physical structure; thus, there are two aspects to the body – a soulish, immaterial and a physical aspect.

I now turn to two final reflections. First, I want to explain how conscious states – e.g., thoughts, memories, sensations – are and are not in the body. As usual, the methods and findings of neuroscience are unable to address the question and, in general, are largely irrelevant to the central questions constituting philosophy of mind (Moreland 2003). Consider, for example, the discovery that if one’s mirror neurons are damaged, then one cannot feel empathy for another. How are we to explain this? Three empirically equivalent solutions come to mind: (1) strict physicalism (a feeling of empathy is identical to the firings of mirror neurons); (2) mere property dualism (a feeling of empathy is an irreducible state

of consciousness in the brain whose obtaining depends on the firing of mirror neurons); (3) substance dualism (a feeling of empathy is an irreducible state of consciousness in the soul whose obtaining depends (while embodied) on the firing of mirror neurons). Of these three, no empirical datum can pick which is correct, nor does an appeal to epistemic simplicity help. Epistemic simplicity is a tie-breaker, and the substance dualist will insist that the arguments and evidence for substance dualism are better than those for the other two options mentioned above.

Now consider a music CD. Strictly speaking, there is no music in the CD; there are only pits (and the old vinyl records contained grooves.) But CDs, if not damaged, when placed in the right retrieval system, cause musical sounds. According to my Thomistic-like view, the body is an ensouled physical structure. The soul is fully present at each point of the body, and its essence informs the body and gives it its nature as a *human* body. Thus, for a current human body to be a body, it must have a soulish and a physical dimension to it.

Now certain pits or grooves associated with memories, thoughts, sensations, and so forth, are formed and stored in the physical dimension of the body (since according to physical theory the body's physical aspect is brute matter and a complex aggregate, it cannot literally store conscious states since if it did, the unity of consciousness would be lost. That unity seems to require that all conscious states belong to a simple subject of consciousness; more on this later. Moreover, whatever the physical aspect of the body stores is spatially extended, but most, if not all, mental states are unextended.) But when these are triggered, whether spontaneously by getting hit in the knee or by the mind searching to bring back a memory, the conscious state will obtain in the body's soulish aspect. Since the body's soulish aspect is just the soul being *holenmerically* present to and in the body, the soul exemplifies conscious properties, not the physical body.

Second, on this view, the role of DNA and genes can be likened to the materials needed to construct a house. Three things are needed for such a construction: specified building materials, a complete floor plan or blueprint, and a specified order in which things will be assembled. In terms of this analogy, DNA's role is to specify the patterns for making the materials (proteins) to be used in assembling the organism. Genes play a role in stabilizing certain aspects of the spatial and temporal order of growth and development, but they do not generate that order. Genes produce cell materials but not the overall plan or internally related organization among the organism's parts. According to organicism, living organisms are wholes irreducible to and ontologically prior to their parts.

Considerable evidence supports an organocentric view. First, the two main functions of DNA (being copied in the process of cell division, and serving as a template for protein synthesis) require the coordinated activity of numerous complex molecules, and can occur only within the context of an entire cell. In fact, as Richard Lewontin says, "DNA is a dead molecule, among the most nonreactive, chemically inert molecules in the living world. That is why it can be recovered from ancient plants and long-dead animal. It has no power to reproduce itself and, while it is promoted as producing proteins, in fact proteins (enzymes) produce DNA" (Lewontin 1992; cf. Noble 2012). The feedback process between DNA and the rest of the cell is species-specific; that is, it is unique to each species and depends on the nature of the specific organism for its distinct activity.

Second, more than DNA is passed on in reproduction. A single-celled zygote contains intricate machinery without which the DNA is biologically inert. This extra material is *always* co-present with DNA, and DNA requires the former for its specific functioning. Experiments have shown that "if a nucleus [which contains DNA] of one species is



transplanted to the enucleated egg [an egg from which the nucleus has been removed] of an unrelated species, the egg may continue to develop for a while, following the pattern characteristic of its own species, rather than the injected nucleus—but the end result is premature death” (Wells 1993, 13).

Can organicism explain how changing a gene can alter characteristics of an organism? Consider an artist using a fine paintbrush to produce a painting. If an artist altered the brush, say one used to paint houses, and replaced it with a fine brush or even an ice pick, this change will alter his product. But neither an altered brush (or ice pick) nor its correspondingly altered product means there is no artist. Instead, the brush is a tool used by the artist; if something happens to either, the result will change.

Similarly, genes comprising DNA sequences are tools – instruments – and that is all. As H. F. Nijhout notes, certain genes produce (via interaction with other parts of the cell) certain materials, which, in turn, help determine which of various possible developmental pathways is actualized. According to Nijhout, “Such genes can thus be said to control alternative developmental pathways, just as the steering wheel of a car controls the direction of travel. However, this is far from equating the steering wheel with the driver” (Nijhout 1990, 442).

What plays the role of driver? Brian Goodwin says the organism is an autonomous, irreducible center of activity, a whole with its own internal nature, its own species-specific principle of development in which the various parts are genuine functional entities that exist for and by means of each other and the whole of which they are parts (Goodwin 1994). Clearly, this language expresses that organisms are substances, not ordered aggregates. If by “soul” we mean an individuated nature, the thin particular, then every living organism is identical to its soul and it is plausible to take the soul to be what Goodwin is getting at when he talks about the organism as a whole.

## **7.2 Three Reasons Embracing an Immaterial Substantial Self is Superior to Physicalism**

Many arguments claim that views entailing an immaterial substantial self are superior to physicalist views, and these are peppered throughout this volume. I wish to mention three such arguments, to add my own reflections to their treatment. If successful, while these arguments count against physicalism, providing support for some version of substance dualism (as defined in this book’s introduction), they do not support my Thomistic-like dualism over alternatives.

### *7.2.1 The unity of consciousness*

More than anyone else, William Hasker has championed this argument for substance dualism (Hasker 1999, 122–144). By the unity of consciousness, say, of one’s visual field, I mean two things. First, there is what Bayne and Chalmers call subsumptive phenomenal unity: Necessarily, all of one’s experiences are subsumed within a single, totalizing state of consciousness. This totalizing state is a conscious state in its own right, and there is a what-it-is-like to be in that state (Bayne and Chalmers 2003, 26–27). For example, consider states A (sensing a chair) and B (sensing a lamp). There is a what-it-is-like to be in A, a what-it-is-like to be in B, and a what-it-is-like to be in A and B together. The total phenomenal unity

thesis says that there is always a single phenomenal state that subsumes all of one's other phenomenal states at a time.

The notion of "subsuming" is a bit unclear, so let me state the second thing I mean by the unity of consciousness, also from Bayne, and add my own metaphysical clarification to it. According to Bayne, an atomistic theory of consciousness states that the phenomenal field is composed of "atoms of consciousness" – independently conscious states (Bayne 2010, 225–229). By contrast, Bayne – and I – accept holism: The components of the phenomenal field are conscious only as components of that field (it is interesting to note that diachronically, consciousness changes as a continuous flow, but the brain changes states in a discrete, atomistic way). I add the qualification that the phenomenal field is a whole in which subsumptive components are modes or inseparable parts of the whole field.

Now consider the following principle: (F) For any complex object (one with a plurality of separable parts)  $O$ , if  $O$  performs function  $F$ , then  $O$ 's performing function  $F$  consists in parts  $p_1$ – $p_n$  and subfunctions/activities  $f_1$ – $f_n$ , such that  $p_1$  performs  $f_1$ , . . .  $p_n$  performs  $f_n$ . For example, a computer performing function  $F$  just is a certain set of its parts performing their own subfunctions. Principle  $F$  can also be stated in terms of properties such that an object  $O$  having some property  $P$  consists in each part having some property or other. This is clearly the case with additive properties, for example, mass. It does not, however, rule out emergent properties. Given the reasonable assumption that supervenience for simple, emergent properties is local (the supervenient simple property obtains and is dependent on what is going on right there at the subvenient base), the principle disallows emergent properties exemplified by complex objects like  $O$  taken as an irreducible whole. But it does not disallow each of the relevant parts of  $O$  to have an emergent property.

The following argument, then, is an attempt to show that the unity of consciousness cannot be explained if one is a brain, because a brain is just an aggregate of different physical (separable) parts. It is only if the self is a single, simple subject that one adequately accounts for the unity of consciousness.

To grasp the argument, consider one's awareness of a complex fact, say one's own visual field consisting of awareness of several objects at once, including a number of different surface areas of each object. One's entire visual field contains several different experiences, for example, being aware of a desk toward one's left side and being aware of a podium in the center of one's visual experience of an entire classroom. Corresponding to such an experience, numerous different light waves bounce off of different objects (and off of different locations on the surface of the same object, say different areas of the desk's top side), they all interact with the subject's retinas, and they all spark signals that terminate in myriad locations of the brain, breaking objects down into constituents (LaRock 2013). If we add local emergence, then we could hold that each relevant part of the brain instantiates an atomistic sensory experience.

Accordingly, a physicalist may claim that such a unified awareness of the entire room by means of one's visual field consists in a number of different physical parts of the brain each terminating a different wavelength, aware only of part – not the whole – of the complex fact (the entire room). But this cannot account for the single, unitary awareness of the entire visual field. There is a what-it-is-like to have the whole visual field. If we terminate our search for an explanation for this with a holistic phenomenal field, then two problems arise. First, it is hard to see how myriad atomistic parts could give rise to a single, nonatomistic, holistic field; we are owed an account of this within the constraints of subject physicalism.

Second, a basic datum of our experience is not simply this or that item of awareness in the room, but that *I have and am not identical* to the totalizing state. In the history of philosophy, classic substances have served to unify things in this way, and I believe this ontology provides the best answer for how we could have a totalizing, unified field of consciousness. The very same substantial soul is aware of the desk to the left, the podium at the center, and, indeed, each and every distinguishable aspect of the room. But no single part of the brain is correspondingly activated as a terminus for the entire visual fields. Only a single, uncomposed mental substance can adequately account for the unity of one's visual field or, indeed, the unity of consciousness in general.

The most prominent physicalist rejoinder attempts to explain objectual phenomenal unity in terms of synchronicity: All the different locations of the brain processing electrical signals associated with different aspects of the object of perception (e.g., color, size, shape, etc.) fire together at the very same time, and this explains objectual unity. Unfortunately, a growing amount of empirical evidence refutes this thesis (see LaRock 2015, 12–15). And, philosophically, the connection between synchronicity and objectual unity is unclear. Consider LaRock's analogy: "If five chefs are located in separate kitchens and each chef is consciously aware of only part of the same recipe, it does not follow that any one chef is consciously aware of the recipe as a whole – *even if all of the chefs are consciously aware of their respective recipe parts at the same time*" (LaRock 2015, 15). The synchronicity solution, then, fails to satisfy.

### 7.2.2 *Human persons are continuants that sustain strict identity through part replacement*

Consider the following argument:

- 1 If something is a physical object composed of parts, it does not survive over time as the same object if it comes to have different parts.
- 2 My body and brain are physical objects composed of parts.
- 3 Therefore, my body and brain do not survive over time as the same object if they come to have different parts.
- 4 My body and brain are constantly coming to have different parts.
- 5 Therefore, my body and brain do not survive over time as the same object.
- 6 I do survive over time as the same object.
- 7 Therefore, I am not my body or my brain.
- 8 I am either a soul or a body or a brain.
- 9 Therefore, I am a soul.

Premise (2) is commonsensically true. Premise (4) is obviously true as well. Our bodies and brains are constantly gaining new cells and losing old ones, or at least, gaining new atoms and molecules and losing old ones. So understood, bodies and brains are in constant flux. I assume (8) represents the only live options for most ordinary people. This leaves premises (1) and (6).

Let's start with (1). Why should we believe ordinary material objects composed of parts do not remain the same through part replacement? Why is mereological essentialism problematic for virtually all versions of physicalism besides those who identify us with an atomic simple? Because these versions of physicalism identify us as (or claim we

are constituted by) mereological aggregates, and, though it is unpopular to say so, mereological essentialism cannot be avoided for such wholes. Given part alteration, physicalism does not have the ontological resources capable of providing/grounding the type of unity needed for a mereological aggregate to be a continuant. It may even be that physicalism does not have the ontological resources to avoid eliminativism regarding mereological aggregates.

Here is a definition of a mereological aggregate: It is a particular whole composed of (at least) separable parts and external relation-instances between and among those separable parts.

Why think mereological essentialism characterizes mereological aggregates? Because a proper metaphysical analysis of such wholes provides no entity adequate to ground metaphysically their literal identity through part alteration. To illustrate, suppose we have some mereological aggregate *W*, say a car, in the actual world *w* at some time *t*, and let “the *ps*” refer distributively to all and only the atomic simples (assuming such) that make up *W*. Given that the *ps* just are a specific list of simples taken distributively regardless of structure, a different list of simples, the *qs*, would not be identical to the *ps*, even if both lists shared all but one part in common. This same insight would be true if we took “the *ps*” and “the *qs*” collectively as referring to some sort of mereological sum. In either case, no entity “over and above” the parts exists to ground sameness through part alteration.

*W* has different persistence conditions from – and, thus, is not identical to – the *ps*. *W* could be destroyed and the *ps* (taken in either sense) could exist. Let *S* stand for all and only the various relations standing between and among the *ps*. *S* is *W*’s type of structure. Is *W* identical to *S* and the *ps*? No. *W* has its own structure, say in comparison to some other whole *W\** exactly similar in structure to *W*. *W* and *W\** have their own structures. Given that *S* is a universal, it is not sufficient for individuating *W*’s specific structure. For that we need *SI*, *W*’s structure-instance, *W*’s token of *S*, and *SI* will consist of all and only the specific relation-instances that are instantiated between and among the *ps*. Let “the *rs*” stand for all and only the relevant relation-instances that compose *SI*. I think it is now obvious that *SI* is a mereological aggregate composed of the *rs*. If the *rs* undergo a change of relation-instances, it is no longer the same list of relation-instances. Given that *SI* just is a mereological aggregate or, perhaps, a specific ordering of the *rs*, if the *rs* undergo a change of relation-instances, *SI* ceases to exist; a different structure (perhaps exactly similar to *SI*) obtains, since no entity serves as a ground for *SI*’s sameness through part replacement. If *W* is the *ps* plus *SI*, it follows that *W* is subject to mereological-essentialist constraints.

What about premise (6): I do survive over time as the same object. Why should we think we survive as the same object over time? Suppose you are approaching a brown table and you undergo a series of table experiences during the process.

In this series of experiences, you are aware of different aspects of the table at different moments. However, at each moment you are also aware of a self having those experiences, uniting them into one field of consciousness across time. Moreover, you are also aware that the very same self had all the experiences during the process. Finally, you are aware that the self that had all the experiences is you yourself.

Through introspection, you are aware that you are the self that owns and unifies your experiences at each moment of time and that you are the same self enduring through time. This is pretty obvious to most people, a basic datum of experience. On the basis of this datum, the belief that we are enduring subjects or selves is properly basic.

### 7.2.3 *The modal argument and near-death experiences*

The core of the modal argument for the soul is fairly simple: (1) I am possibly disembodied (I could survive without my brain or body). (2) Neither my brain nor body is possibly disembodied (they could not survive without being physical). (3) So I am not my brain or body. (4) I am either a soul or a brain or a body. (5) So I am a soul.

The most controversial premise is (1) and defenders of the argument have typically taken two strategies to support it. First, those like Stewart Goetz defend what is called the Simple Argument (I am aware of myself as a simple substance and of my body as a complex thing, so I am not identical to my body) to serve as an epistemic support for (1): Since I know I am not my body, then it is metaphysically possible that I exist without it in the absence of overriding defeaters of which there are none (Goetz 2001, 89–104).

Second, those like Charles Taliaferro defend (1) on the grounds that most people can conceive of and have the basic modal intuition that one could exist disembodied (Taliaferro 1986). I favor Goetz's approach, but I believe that for those who deny we are directly aware of ourselves as simple substances, Taliaferro's approach would still be available. However, there is a third approach for defending (1) that, surprisingly, has rarely been used in philosophical circles. In this approach, I appeal to the numerous, credible near-death experiences (NDEs; These should be called death experiences because there is nothing *near* about them. In most cases, the person's heart stopped beating and brain stem activity went silent).

NDEs aren't rare. A Gallop poll and other studies report 4 percent of Germans and 4 percent of Americans have had an NDE (one out of twenty-five). Nine to twelve million Americans have had an NDE and an estimated 200–300 million worldwide have had such an experience. Interestingly, the core elements of an NDE are remarkably the same across various cultural and religious traditions (Kellehear 2009, 134–158; Long 2010, 149–171), going back to the ancient Near East. Moreover, 900 journal articles about NDEs have been published in the scholarly literature. Finally, most NDE researchers were skeptics before they did their research and many of them have PhDs or MDs. (Three key books containing much of what is laid out above are Holden, Greyson, and James 2009; Long 2010; Miller 2012; a solid, popular treatment of NDEs is Atwater 2000; see also, Morse 1990, 1992; Sabom 1998.)

The evidence derived from actual NDE cases is powerful and persuasive. For one thing, NDEs happen to little children who know nothing about heaven, dying or the afterlife, yet they learn things while dead they could not have known were the NDE inauthentic (Morse 1990). For example, in the popular book *Heaven is for Real* Todd Burpo, a pastor and respected man in the community, tells of his 4-year-old son's trip to heaven during a medical crisis. The boy saw and learned things only a real NDE could explain adequately: he met his miscarried sister about whom he had never been told; he told details of a great-great grandfather he couldn't have known (and, while he did not recognize an older picture of the great-great grandfather, after leafing through family albums, he did recognize him as he looked in his prime); he saw and accurately described Todd's actions while praying in a separate hospital room. All these details were confirmed.

Moreover, carefully researched cases exist in which people born blind gain sight during an NDE only, and they describe, with limited vocabulary, their rooms and what is happening outside; they return to being blind when they reenter their bodies (Ring and Cooper 2008).

Too many people have had NDEs and too many carefully researched cases exist for people just to dismiss their veridicality out of hand. Additionally, no naturalistic explanation exists. In my view, the collaborative evidence confirmed by other people shows beyond a reasonable doubt that NDE experiencers leave their bodies and see/hear things. Moreover, there is worldwide consistency to the core components people experience during an NDE (Long 2010, 149–171). These events are widespread and frequent; they happen to atheists and very young children; the deaf hear and the blind see and the colorblind see colors. They have enhanced consciousness when they should have no consciousness at all, and incredible life changes result from them.

Physicalists argue that the modal argument establishes only the epistemic or conceptual possibility of disembodiment, not metaphysical possibility. But this is just a question-begging assertion, since its only “evidence” requires a prior commitment to physicalism. The reality of NDEs, or even the metaphysical possibility of their being true due to the complete coherence of the accounts, adequately demonstrates the metaphysical possibility of disembodiment, and according to the argument presented earlier, this supports the truth of some form of substance dualism.

### **7.3 Advantages of Thomistic-like Dualism (Organicism) Over Other Substance Dualisms**

#### *7.3.1 The causal pairing problem*

Various arguments solve the causal pairing problem, but in my view, the most effective finds a connection between soul/body or mind/brain more primitive than causation which grounds the latter. Such a connection is at the very heart of my Thomistic-like dualism. The soul animates, unifies, forms, informs and is hollenmerically present to its body. My mind and brain, then, causally interact with and only with each other, because my mind is a faculty of my soul, my brain is an inseparable part of my body, and my soul relates to my body in the ways just specified.

#### *7.3.2 The correlation of an increase in brain complexity and mental functioning*

In general, as we observe animals with less brain complexity, we also notice less mental functioning, which seems more or less correlated up to human persons. Why? As evolution produces increasingly complex brains, the most fashionable physicalist argument goes, that increase creates new thresholds of complexity appropriate for more sophisticated emergent properties of consciousness to be instantiated. But an appeal to emergent properties has always seemed suspect to me: “emergence” is not a solution, but a name for the problem to be solved. Further, it seems to be a case of getting something from nothing, a case of magic without a Magician. And if one loads subvenient entities with dispositional power to actualize the emergent property, one had difficulty explaining the simple unity of the emergent property from innumerable dispositional parts. Finally, sorites problems lurk in the neighborhood. Could this emergent property be instantiated with one less atomic part in the subvenient base? How about two less atomic parts? At some point, the emergentist must say the addition of one small atomic part had a huge, disproportionate metaphysical effect – the instantiation of the new emergent property.

My Thomistic-like dualism solves this naturally: Since the soul's intrinsic essence/ordered set of capacities/information forms the body to express its powers that depend on certain bodily structures, the soul unfolds a brain suited for its faculty of mind to be fully operative. Brain complexity and mental sophistication correlate because the brain was formed to express mental functioning.

### 7.3.3 *Intrinsic normative proper functioning*

My organicism clearly explains how biological parts function properly. A part functions properly when it functions the way it ought to function, according to how its internal essence was conceived, intended, and designed to function by God.

This may become clearer if we compare the human essence, humanness, to a blueprint. In creating human beings, God first conceived of humanness as a blueprint. Second, when he created Adam and Eve, that blueprint acted in them as an internal organizing principle. When we say *the heart functions to pump blood*, we are not talking about any particular heart or any statistical collection of hearts, but about *the ideal heart*. How is it ideal? The ideal heart is a part of ideal human nature, as it was conceived and designed by God.

On this picture, statements about proper function (*the heart functions to pump blood*) have two features. First, they are normative. A heart that pumps blood functions the way it *ought* to function according to the role in human nature it was intended to play. Second, the functional language about the heart is literal. The blueprint of what a heart is supposed to do is within the being of each human person and the heart is a literal functional entity internally related to other parts of the body, which works with them in a real, teleological way.

Contrast this view with that embraced by most scientific naturalists, which does away with this notion of designed proper function. In its place, proper function is understood as a usual function that is common and statistically most prevalent, which confers advantage in the struggle for survival. The heart, then, ought to pump blood because most (reproductively advantageous) hearts, in fact, pump blood. The Thomistic-like dualist takes this naturalistic understanding of proper function to be neither necessary nor sufficient for real proper function. While I won't do it here, it is easy to come up with counterexamples that support this claim.

So much for the naturalist account of "proper" function. The Thomistic-like dualist view seems superior to naturalistic approaches. Unfortunately, things are not this straightforward when it comes to comparing the theist Thomistic-like substance view with the theist mereological-aggregate treatment of proper function (views according to which the soul is a substance but the body is a mere physical mereological aggregate). The difference between them lies in two key points.

First, the Thomistic-like advocate views the macro-parts of the body as inseparable; mereological-aggregate advocates depict them as separable. Second, the Thomistic-like substance theist views the human essence as a blueprint literally designed and *placed within* the individual human being as the internal, metaphysical ground of genuine internal relations, real functional entities, and actual teleology within human bodily development. The mereological-aggregate theist depicts the human person's body as an artifact only. The blueprint of a car in a designer's mind causes him to assemble certain parts in a certain way, but those parts, for example, a carburetor, are really only mechanical devices externally related to other parts, engaging in efficient causality only. Strictly speaking, no internal relations, no teleology, no functional movements exist *within* the car itself. The car's parts behave *as if* they

were real functional entities. Still, there is a real sense in which the proper versus improper function is genuine and normative. A carburetor functions properly if its movements are isomorphic with and “realize” the concept of a carburetor in the mind of the car’s designer.

For Thomistic-like substance versus mereological-aggregate theists, then, the issue comes down to this. Both can allow for true, normative statements about the proper functioning of the parts of the human body. The Thomistic-like substance theist will claim the human essence is a blueprint designed by God, placed within the individual human being such that the parts of the human being are themselves literally constituted by internal relations to other parts, which play a functional, teleological role metaphysically contained in the organism itself. The mereological-aggregate theist admits the human essence is a blueprint designed by God that serves as the standard for proper versus improper functioning. But he insists that the individual human body is still a physical mereological aggregate composed of separable parts with no internal relations, irreducible functional roles, or teleology literally within the body of the individual human.

I cannot argue, in detail, the relative merits of the two positions here and I admit that this is a limitation of my exposition. But I hope I have made the issue clear enough to foster further dialogue. At the very least, I should point out that it has been almost impossible for biologists to avoid functional, teleological language and the Thomistic-like view explains why.

### *7.3.4 Comportment with neuroscientific data*

When it comes to neuroscientific data regarding correlations and dependencies both ways, my model is empirically equivalent with strict physicalism or mere property dualism. And it comports with those data without advancing the absurd notion that conscious states could exist in the brain and nervous system qua brute matter. Whatever is in the brain (and the “in” is a spatial in) is spatially located, extended, and shaped, but mental states like beliefs, thoughts, desires have no spatial extension or shape and, arguably, no location. They are in the soul by way of an “in” of exemplification. Further, as Leibniz’s mill thought experiment showed long ago, if we leave the macroworld and its inaccurate manifest image, and travel to the more accurate level of microphysics – if we were shrunk in order to enter the brain, we would see a galaxy with particles far from each other in largely empty space. It is hard to see where one would put an unextended memory. Inside an atom? Spread out in the empty space between atoms and molecules? My view avoids these difficult questions and this fact counts in its favor.

## **7.4 Arguments Against Thomistic-like Substance Dualism**

Some have raised arguments against substance dualism per se, for example, the interaction problem(s), private language difficulties, and so on. But in this section I limit my remarks to three criticisms. The first is not often brought up in mainstream philosophy of mind, yet it deserves a response. The next two are directed at Thomistic and Thomistic-like views.

### *7.4.1 Specific mind/brain dependencies undermines substance dualism*

According to Nancey Murphy, “science has provided a massive amount of evidence suggesting that we need not postulate the existence of an entity such as a soul or mind



in order to explain life and consciousness” (Murphy 1998, 18). This evidence claims that “biology, neuroscience, and cognitive science have provided accounts of the dependence on physical processes of *specific* faculties once attributed to the soul” (Murphy 1998, 17, see also 13, 27). Elsewhere she claims: “My argument in brief is this: all of the human capacities once attributed to the mind or soul are now being fruitfully studied as brain processes – or, more accurately, I should say, processes involving the brain, the rest of the nervous system and other bodily systems, all interacting with the socio-cultural world” (Murphy 2006, 56).

When I first came across this argument, I could hardly believe anyone could take it seriously, and C. Stephen Evans puts his finger on why:

What, exactly, is it about these findings that are supposed to create problems for dualism? . . . Is it a problem that the causal effects should be the product of specific regions of the brain? Why should the fact that the source of the effects are localized regions of the brain, rather than the brain as a whole, be a problem for the dualist? It is hard for me to see why dualism should be thought to entail that the causal dependence of the mind on the brain should only stem from holistic states of the brain rather than more localized happenings. (Evans 2005, 333–334)

In addition to Evans’s bewilderment about why these discoveries should be thought of as counting against dualism, I offer three further responses. First, many substance dualists – including me – believe in a substantial ego not because it is a theoretical postulate with superior explanatory power, but because it is something of which people are directly aware. Thus, belief in a substantial, simple soul is properly basic and grounded in self-awareness. The point is not that dualists are right about this; the point is, given this dualist approach, that advances in our knowledge of mental/physical dependencies are simply beside the point. Further debate about which approach is the fundamental one for defending substance dualism is not something for which advances in scientific knowledge are relevant.

Second, in those cases where substance dualism *is* postulated as having superior explanatory power, typically the explanations are not of scientific facts like those Murphy mentions, but are distinctively philosophical. Arguments from the unity of consciousness, the possibility of disembodied survival or body switches, the best view of an agent to support libertarian agent causation, the metaphysical implications from the use of the indexical “I” are typical of arguments offered by substance dualists, and the facts Murphy mentions are not particularly relevant for assessing these arguments. As I have already tried to show, when views like Thomistic-like dualism entail certain biological facts about organisms, it seems a growing number of scientists acknowledge those facts.

Finally, the discovery of “the dependence on physical processes of *specific* faculties once attributed to the soul” does not sufficiently ground attributing those faculties to the brain. There is an important distinction between describing the nature, proper categorization and possessor of a capacity versus explaining what conditions are necessary for its actualization. To see this it is important to get clear on the use of “faculty” as the term has been historically used in discussions of substances in general and the soul in particular. Roughly, a faculty of some particular substance is a natural grouping of resembling capacities or potentialities possessed by that thing. For example, the various capacities to hear sounds would constitute a person’s auditory faculty. Moreover, a capacity gets its identity and proper metaphysical categorization from the type of property it actualizes – its manifestational property. The nature of a capacity-to-exemplify-F is properly characterized by F itself. The capacity, then, to reflect light is properly considered a physical, optical capacity. For property dualists, the

capacities for various mental states are mental, not physical, capacities. Thus, the faculties constituted by those capacities are mental, not physical, faculties.

Arguably, a particular is the kind of thing it is in virtue of the actual and potential properties/faculties essential and intrinsic to it. To describe the faculties of a thing, then, provides accurate information about the kind of particular that has those faculties. Moreover, a description of a particular's capacities/faculties is a more accurate source of information about its nature than is an analysis of the causal/functional conditions relevant for the particular to act in various ways. The latter can be either clues to the intrinsic nature of that particular or information about some other entity that the particular relates to in exhibiting a particular causal action. Remember, there is a difference between attempts to describe, categorize and identify a capacity's nature and possessor, and proffering an explanation of the functional/causal conditions that must be present for that capacity to be actualized.

I have already pointed out that the brain is not the right sort of thing to possess mental capacities or actualized properties (e.g., it raises unity-of-consciousness problems). If we set aside the manifest image and look at the brain at the level of microphysics, it is unclear where these mental capacities would dwell. And it would seem that everything in the brain has spatial extension, but mental capacities do not. The dual nature of the body that is part of Thomistic-like dualism is a much better way to handle these specific correlations/dependencies than is Murphy's attempt to stick them in the brain.

#### *7.4.2 Thomistic-like dualism is a version of a rightly abandoned vitalism*

William Hasker has forcefully raised this objection (Hasker 2013, 503–504), but vitalism has been misunderstood frequently because the concepts of that debate have been used in many different ways. For example, during its zenith as a scientific research program, there were several distinct forms of vitalism. The more crude forms of vitalism have rightly been rejected because of their tendency to depict the individuated essence as either a spatially located vital entity, a force, or a fluid (like caloric or phlogiston) viewed as a mechanistic entity alongside other mechanical parts. This strategy reduced the living organism to a special sort of mereological aggregate with just another mechanical (though immaterial) part.

A more adequate vitalism, if we wish to use this term of the Thomistic-like substance view (or, as I prefer, "organicism"), grounds its explanatory power in factors like the identification of an adequate driver for DNA, the irreducible organic, holistic relation among parts to parts and parts to whole, the species-specific immanent law of organization and teleological development, and the individuated, internal structural essence holomergically present and very much like the current notion of an organism's information, the reality of inseparable parts and normative functioning found in living things.

Hasker is well aware of my response. As a rejoinder, he claims a majority of biologists reject my view; further, he doubts my view will provide a promising line for future research (Hasker 2013, 504). I believe Hasker is correct about his former claim, and I am no sociologist of biology. But the central components of my view with scientific implications are, indeed, being embraced by more and more biologists. A number of biologists have announced the end of the old mechanistic approach to organisms (for a list see Wells 2014; see also Webster and Goodwin 1996, ix, 193). At the same time, Michael Denton, Govindasamy Kumaramanickavel and Michael Legge proclaim, "that despite the dominance of the mechanistic-reductionist paradigm through most of the past century the

possibility of a twenty-first-century organicist revival cannot be easily discounted” (Denton, Kumaramanickavel, Legge 2013, 31).

Moreover, old paradigms die hard, and biologists have been so deeply indoctrinated with the machine metaphor for organisms that the sociology of the biological community both blinds biologists to the inadequacy of the metaphor and prevents biologists from being open to organicism. Stephen Talbott bluntly calls biology’s refusal to disown the machine metaphor “an inexcusable mistake” that “has gripped the scientific community for decades, severely perverting biological understanding” (cited in Wells 2014, 2). As an example, Talbot cites a widespread complaint voiced at a conference at Harvard Medical School that, even though fifty years have passed since biologists have known that some proteins exhibit a disorder-dependent function, not a single biology text cites an example (Talbott 2010, 35–36). Thus, the fact that a majority of biologists reject views like mine may be due more to the sociology of biology than to the explanatory impotence of organicism.

There are three key elements in the growing acceptance of organicism relevant to my thesis:

- 1 A return to Aristotelian essentialism: According to Webster and Goodwin, organisms have real species-essences grounding their membership in real natural kinds and, as such, these essences are Aristotelian in nature (Webster and Goodwin 1996, 17–25, 31, 36, 52–53, 66–70, 98–100). Moreover, essences are *holenmerically* present to the organism, constituted by the blueprint/information for forming and maintaining the organism (Dembski 2014, 47–54). This leads Kaiser to say “the soul is the form or actuality that makes the body organized, and therefore able to act” (Kaiser 2014, 19). Elsewhere he says:

the parts of the organism, even cells which are capable of living separately from the body, act as parts of the one and the same substance; they participate in the one life of the whole organism. So the soul, by being the actuality of the whole and all the parts is the principle of unity of the organism. DNA, on the other hand, although present in nearly all cells of the body (e.g., red blood cells lack a nucleus or DNA), is not a principle of unity; rather, its activity is regulated by the whole organism. (Kaiser 2014, 20)

- 2 The return of irreducible teleology: According to Georg Toepfer, “Nothing in biology makes sense, except in the light of teleology” (Toepfer 2012, 113). Accordingly, there is a return to the idea that form determines function (Kaiser 2014).
- 3 Holism and inseparable parts: The rise of organicism implies a return to a view of organisms as irreducible wholes, ontologically prior to their parts, and whose parts are inseparable parts literally constituted by the role they play in the whole (Kaiser 2014; Talbott 2010; Toepfer 2012; Denton, Kumaramanickavel, Legge 2013).

Organicism, then, is gaining steam. But advocates of organicism tend to identify the key entity as the organism as a whole – the thick particular – rather than the organism’s soul – the thin particular. Thus, I want to close this section by explaining why I think we should prefer the thin particular.

Organicists place the emergence of the thick particular against the backdrop of the standard mereological hierarchy: at the biological level emergent wholes consist in systems or networks of mutually interacting parts and interpenetrating processes acting for the whole (Phillips 1970, 427). Given this framework, the thin particular view avoids five problems.

First, the organicist must load brute matter, as described by chemistry and physics, with emergent-property potentialities. Regarding the emergence of sentience, panpsychism must be embraced as Denton, Kumaramanickavel, and Legge (2013, 39–40) explicitly acknowledge. This is hardly good news for those hoping to avoid “spooky” entities like a vital spirit. Is it less “spooky” to attribute an attenuated form of sentience to electrons? Further, these emergent potentialities are actualized by the mere appearance of a new, and in this sense, more complex spatial configuration of parts. Second, at this “proper” level of complexity, one suddenly has parts shifting from standing in external to internal relations. Third, all the laws in the hierarchy below living organisms are efficient causal laws. But at the biological level, suddenly we get emergent teleological laws. In these three cases, at the right level of complexity, presto, we get all these new entities by the mere spatial rearrangement of parts. This sounds to me like magic without a magician.

Fourth, the view additionally suffers from the combinatorial problem: How can mereologically simple, *sui generis* properties or wholes emerge from a combination of separable parts? The issue here is not the sheer coming-to-be of these emergent entities. Rather, the focus is on the simple unity of genuinely emergent properties. It is difficult to imagine how merely rearranging myriad parts could give rise to simple, uncomposed, nonstructural entities.

Fifth, as Thomas Kaiser points out, the thick particular view cannot account for the special unity and order of the organism’s parts. It can simply notice them (Kaiser 2014, 2). Kaiser argues that a system, no matter how complex, is such that if you change the parts or order, you no longer, in the strict philosophical sense, have the same system (Kaiser 2014, 2–12, 15–17); instead, a genuine substance that stands under, organizes and develops the “system” is needed – none other than the thin particular of the organism.

Qua substance, the thin particular addresses how the organism remains the same through part or order replacement. It also avoids the first four difficulties by depicting the thin particular’s origin as an example of substantial change. When certain conditions are met (e.g., the union of sperm and egg each of which have soulish potentialities) a new essence appears and, along with it, a new substance. This has been an intelligible view for 2,500 years and, even if wrong, it does not face the four conceptual difficulties that plague the thick particular view. Nor is there a difficulty with the combinatorial problem.

Moreover, as Talbott points out, the parts of an organism grow within an integral unity existing from the very start. He also notes that the form, existence, and activities of the parts depend on and arise from and are caused by the whole (Talbott 2010, 38, 40). This is best explained by a thin particular that exists from the very beginning of the organism’s existence and, subsequently, teleologically develops inseparable parts in a particular structure for the sake of the whole.

Of course, the adoption of the thick versus the thin particular may be simply an aesthetic preference. As Denton, Kumaramanickavel, and Legge acknowledge, the embracing of organicism “[is] aimed at replacing the actions of an immaterial spirit with what was seen as an equivalent but perfectly natural agency – the *emergent* autonomous activity of the *whole* organism” (Denton, Kumaramanickavel, and Legge 2013, 31).

The sort of essentialism entailed by my view is not at home in evolutionary theory. As E. Mayr has said:

The concepts of unchanging essences and of complete discontinuities between every *eidos* (type) and all others make genuine evolutionary thinking impossible. I agree with those who

claim that the essentialist philosophies of Aristotle and Plato are incompatible with evolutionary thinking. (Mayr 1970, 4)

I have already addressed this above. But as a further response, I must agree with Thomas Nagel who expresses considerable skepticism in regard to the Darwinian story of the origin of life and the development of all current life forms by Darwinian mechanisms. Says Nagel about his skepticism:

It is just a belief that the available scientific evidence, in spite of the consensus of scientific opinion, does not in this matter rationally require us to subordinate the incredulity of common sense. (Nagel 2012, 7)

Like Nagel, I do not accept the blind watchmaker thesis, nor do I believe in the thesis of common descent. So Mayr's problem doesn't bother me. In fact, my Thomistic-like dualism should be considered an external conceptual problem for standard evolutionary theory. Indeed, Michael Chaberek has persuasively shown that Aquinas's philosophy is inconsistent with theistic evolution (Chaberek 1997).

As philosopher of science Larry Laudan taught us long ago, an external conceptual problem (e.g., philosophical arguments against the possibility of running through an actual infinite by successive addition) is an intellectual difficulty first raised in a discipline outside of science but which, upon further inspection, may count directly against a certain scientific theory that entails things that are contrary to the external conceptual problem (e.g., an infinitely old, steady state universe) (Laudan 1977). When scientists weigh the rationality of accepting their theory, one key factor is its ability to handle external conceptual problems. If the theory cannot do so, the external conceptual problem counts against the theory's rationality.

I have tried to clarify my specific version of substance dualism and interact with arguments for and against it. Whether or not you, the reader, accept my view, it still seems to be the case that it is worthy of being a greater part of the conversation. Perhaps my chapter will facilitate that.<sup>1</sup>

## Note

1. I want to thank Paul Nelson for several helpful insights he gave me regarding an earlier draft of this chapter.

## References

- Aranda-Anzaldo, Armando. 2011. "Assuming in Biology the Reality of Real Virtuality (A Come Back for Entelechy?)." *Ludus Vitalis*, 19: 333–342.
- Atwater, P. M. H. with David H. Morgan. 2000. *The Complete Idiot's Guide to Near-Death Experiences*. Indianapolis, IN: Alpha Books.
- Bayne, Tim. 2010. *The Unity of Consciousness*. Oxford: Oxford University Press.
- Bayne, Tim, and David J. Chalmers. 2003. "What is the Unity of Consciousness." In *The Unity of Consciousness*, edited by Axel Cleeremans, 23–57. Oxford: Oxford University Press.
- Chaberek, Michael. 1997. "Thomas Aquinas and Theistic Evolution." *AAS*, 89: 186–190.
- Dembski, William. 2014. *Being as Communion*. Burlington, VT: Ashgate.

- Denton, Michael, Govindasamy Kumaramanickavel, and Michael Legge. 2013. "Cells as *Irreducible Wholes*: The Failure of Mechanism and the Possibility of an Organicist Revival." *Biology & Philosophy*, 28: 31–52.
- Des Chene, Dennis. 2000. *Life's Form: Late Aristotelian Conceptions of the Soul*. Ithaca, NY: Cornell University Press.
- Evans, C. Stephen. 2005. "Separable Souls: Dualism, Selfhood, and the Possibility of Life After Death." *Christian Scholar's Review*, 34: 327–340.
- Goetz, Stewart. 2001. "Modal Dualism: A Critique." In *Soul, Body and Survival*, edited by Kevin Corcoran, 89–104. Ithaca, NY: Cornell University Press.
- Goodwin, Brian. 1994. *How the Leopard Changed Its Spots*. New York: Simon & Schuster.
- Hasker, William. 1999. *The Emergent Self*. Ithaca, NY: Cornell University Press.
- Hasker, William. 2013. "The Dialectic of Soul and Body." *American Catholic Philosophical Quarterly*, 87: 495–509.
- Holden, Janice Minor, Bruce Greyson, and Debbie James, eds. 2009. *The Handbook of Near-Death Experiences*. Santa Barbara, CA: Praeger Publications.
- Kaiser, Tom. 2014. "Is DNA the Soul?" Paper presented at the West Coast Meeting of the Society for Aristotelian and Thomistic Studies, June 14.
- Kellehear, Allan. 2009. "Census of Non-Western Near-Death Experiences to 2005: Observations and Critical Reflections." In *The Handbook of Near-Death Experiences*, edited by Janice Holden, Bruce Greyson, and Debbie James, 135–158. Santa Barbara, CA: Praeger Publications.
- Koslicki, Kathrin. 2008. *The Structure of Objects*. Oxford: Oxford University Press.
- LaRock, Eric. 2013. "From Biological Naturalism to Emergent Substance Dualism." *Philosophia Christi*, 15(1): 97–118.
- LaRock, Eric. 2015. "Emergent Dualism is Theoretically Preferable to Reductive Functionalism." Unpublished manuscript, March 31.
- Laudan, Larry. 1977. *Progress and Its Problems*. Berkeley: University of California Press.
- Lewontin, Richard. 1992. "The Dream of the Human Genome." *New York Review of Books*, May 28.
- Long, Jeffrey. 2010. *Evidence of the Afterlife*. New York: HarperOne.
- Mayr, E. 1970. *Populations, Species, and Evolution*. Cambridge, MA: Harvard University Press.
- Miller, J. Steve. 2012. *Near-Death Experiences as Evidence for the Existence of God and Heaven*. Acworth, GA: Wisdom Creek Press.
- Moreland, J. P. 1998. "Theories of Individuation: A Reconsideration of Bare Particulars." *Pacific Philosophical Quarterly*, 79: 251–263.
- Moreland, J. P. 2002. "Naturalism, Nominalism and Husserlian Moments." *The Modern Schoolman*, 79(January/March): 199–216.
- Moreland, J. P. 2003. "A Christian Perspective on the Impact of Modern Science on Philosophy of Mind." *Perspectives on Science and Christian Faith*, 55(March): 2–12.
- Moreland, J. P. 2013. "Exemplification and Constituent Realism: A Clarification and Modest Defense." *Axiomathes*, 23: 247–259.
- Morse, Melvin. 1990. *Closer to the Light: Learning the Near-Death Experiences of Children*. New York: Ballantine Books.
- Morse, Melvin. 1992. *Transformed by the Light*. New York: Villard Books.
- Murphy, Nancey. 1998. "Human Nature: Historical, Scientific and Religious Issues." In *Whatever Happened to the Soul?*, edited by Warren S. Brown, Nancey Murphy, and H. Newton Malony, 1–29. Minneapolis, MN: Fortress Press.
- Murphy, Nancey. 2006. *Bodies and Souls, or Spirited Bodies?* Cambridge: Cambridge University Press.
- Nagel, Thomas. 2012. *Mind & Cosmos*. Oxford: Oxford University Press.
- Nijhout, H. F. 1990. "Metaphors and the Role of Genes in Development." *BioEssays*, 12(September): 441–446.

- Noble, Denis. 2012. "A Theory of Biological Relativity: No Privileged Level of Causation." *Interface Focus*, 2: 55–64.
- Pasnau, Robert. 2011. *Metaphysical Themes: 1274–1671*. Oxford: Clarendon.
- Perler, Dominik, ed. 2015. *The Faculties*. Oxford: Oxford University Press.
- Phillips, D. C. 1970. "Organicism in the Late Nineteenth and Early Twentieth Centuries." *Journal of the History of Ideas*, 31(July–September): 413–432.
- Ring, Kenneth, and Sharon Cooper. 2008. *Mindsight: Near-Death and Out-of-Body Experiences of the Blind*. New York: iUniverse.
- Sabom, Michael. 1998. *Light & Death*. Grand Rapids, MI: Zondervan.
- Smith, Barry, and Wolfgang Kunne. 1982. *Parts and Moments: Studies in Logic and Formal Ontology*. Munich: Philosophia Verlag.
- Suárez, Francisco. 1597. *Disputationes metaphysicae*. n.p.
- Talbott, Steve. 2010. "The Unbearable Wholeness of Beings." *The New Atlantic Monthly*, Fall: 27–51.
- Taliaferro, Charles. 1986. "A Modal Argument for Dualism." *Southern Journal of Philosophy*, 24: 95–108.
- Toepfer, Georg. 2012. "Teleology and Its Constitutive Role for Biology as a Science of Organized Systems in Nature." *Studies in History and Philosophy of Biological and Biomedical Sciences*, 43: 113–119.
- Webster, Gerry, and Brian Goodwin. 1996. *Form and Transformation: Generative and Relational Principles in Biology*. Cambridge: Cambridge University Press.
- Wells, Jonathan. 1993. "The Dogma of DNA." *Bible-Science News*, 31(8): 15.
- Wells, Jonathan. 2014. "Revolutionary Biology." *Evolution News and Views*, October 20.

# A Critique of Thomistic Dualism

WILLIAM HASKER

The Thomistic doctrine of the soul as the form of the body has many of the right intentions. It aims to promote a close integration of soul and body, and more broadly of the human person with the overall world of nature. Yet it does this without denying or minimizing the distinctive attributes of human beings as rational, moral, and religious creatures. And while emphasizing that the normal and the best state for human beings is as embodied persons, it makes room for their persistence disembodied after biological death and their eventual re-embodiment in the resurrection.

It is all the more regrettable, then, that the view as usually understood cannot accomplish these goals in a way that makes it a good candidate for our acceptance. So, at any rate, I shall argue. I will then describe a modification of the view that remedies some of the flaws noted in the standard version, but brings with it additional difficulties. Finally, I will present a different view, one that shares enough of the assumptions and motivations of the Thomistic view to qualify as a worthy successor.

## 8.1 A Summary of Aquinas's View

I begin with a brief survey of Aquinas's view of the soul, following the exposition of that view by Eleonore Stump (2003). The place to begin is with Aquinas's notion of form; specifically, with his notion of *substantial form*. (There are also accidental forms: when Socrates, who was formerly standing, sits down, he acquires the accidental form, *being seated*. For the most part accidental forms will not concern us here.) According to Stump, "A substantial form is the form in virtue of which a material composite is a member of the species to which it belongs, and it configures prime matter" (Stump 2003, 194). (Prime matter is "stuff" stripped of all distinguishing characteristics; it is the basic constituent in all material existence. Prime matter, however, never exists on its own, but always comes configured by a form of some kind.) All material things are composed of matter and form, and:



The complete form (the substantial and accidental forms taken together) of a non-human material object is the arrangement or organization of the matter of that object in such a way that it constitutes that object rather than some other one and gives that object its causal powers. (Stump 2003, 194)

The final clause here is important: the form accounts for the fact that the object has properties, including causal powers, which the matter by itself cannot have. Also worthy of note is the specification to *nonhuman* material objects. Humans are indeed material objects, but the substantial form of a human being is of a different sort than the forms of other such objects. This contrast holds in particular between humans on the one hand and plants and animals on the other. Plants and animals have souls (the “vegetative soul” and the “sensitive soul”), but:

Unlike human souls, the souls of plants and nonhuman animals are nonetheless material forms, and even a material form that is a soul goes out of existence when the material composite it configures goes out of existence. (Stump 2003, 201)

Such a form is characterized by Stump as a “configurational state” of the matter which it informs.

Human souls, however, are another sort of thing. In addition to material forms, there are “subsistent forms.” These are immaterial rather than material; an angel, for example, is a subsistent immaterial form. Such subsistent forms can exist without there being any matter which they configure: according to Aquinas, “if there is a subsistent form, it is immediately an entity and one” (Stump 2003, 198). He also says, “nothing keeps a form from subsisting without matter, even though matter cannot exist without a form” (Stump 2003, 199). Now, it may occur to us at this point that, in employing the same term to designate both these immaterial subsisting entities and the configurational states of material composites, Aquinas is simply equivocating on the word “form.” I am inclined to think there is something in this complaint,<sup>1</sup> but Stump explains the commonality of meaning by saying that “for Aquinas, to be is to be configured or to have a form; and everything, material or immaterial, is what it is in virtue of a form” (Stump 2003, 200). Since this is so, “although Aquinas is perfectly content to deny matter of God, he refuses to deny form of God: being, even divine being, is configured” (Stump 2003, 200).

The human soul, however, is different from other subsistent immaterial forms, in that the human soul, unlike an angel, does configure matter. Aquinas puts it like this:

the human soul has subsistent being, insofar as its being does not depend on the body but is rather elevated above corporeal matter. Nevertheless, the body receives a share in its being, in such a way that there is one being of soul and body, and this is the being of a human. (Stump 2003, 201)

Stump sums up the situation by saying:

So, for Aquinas, the human soul is the noblest and highest of the forms that configure matter, but it is the lowest in the rank of intellectual subsistent forms, because it is mixed with matter, as the intellectual subsistent forms that are angels are not. (Stump 2003, 201)

It remains to say something about the way in which the material composite which is a human being begins and ends its earthly existence:

Aquinas thinks that a human being is generated when the human soul replaces the merely animal soul of the fetus in the womb and that a human being is corrupted or decomposed when the human soul leaves the body and is replaced by whatever other substantial form is in the dead corpse. (Stump 2003, 203)

Note that it is not a *preexisting* human soul that replaces the animal soul of the fetus; rather, the soul is created and infused in the same instant; the soul is created as the soul of this particular body. According to Aquinas, this occurs some 40 days after conception for a male, and 80 days for a female. An interesting consequence of this is that an early stage fetus not only is not a human being, but *will never become a human being*. When one substantial form replaces another, the commonality between the two consists merely in the “prime matter” which is the same in both (see Brower 2012). What we can say about the fetus, then, is that it consists in part of the prime matter that, if all goes well, will in time come to be the prime matter of a human being.

There is much more in Aquinas’s view that invites discussion, but the bare bones as set forth here should be sufficient to enable us to see how the view meets the desiderata outlined at the beginning of this chapter. By making the soul the form of the body, which structures the body and enables its distinctive powers and activities, the view points us to a closer and more profound integration of the two than is apparent either in the Platonic dualism which Aquinas rejected, or in modern Cartesian dualism. At the same time, the distinctive nature of the human soul, as contrasted with the souls of animals, makes it plain that humans are not *merely* animals; rather, they are “rational animals,” with all that implies. The fact that souls are subsistent, immaterial forms means that they are fully capable of continuing to exist after physical death, albeit in an imperfect state which looks forward to fulfillment in the resurrection. One might well ask, what is there in all this not to like?

## 8.2 Objections to Thomistic Dualism

Unfortunately, the answer to that question is, “Quite a bit.” The problems I shall indicate become evident if the view is scrutinized in the light of certain more recent developments in both science and philosophy – developments, obviously, which Aquinas could not have taken into account. I will now argue that (1) this Thomistic view fails to integrate human beings with the rest of nature convincingly, (2) the work actually done by the human soul, following Aquinas’s theory, is surprisingly limited, (3) the case for including such souls in our system, as opposed to thinking of human beings as composed of matter and nothing else (or, nothing but material forms), is quite weak, and (4) the view in its traditional form is very difficult to combine in any plausible way with the well-established facts of biological evolution.

I begin by pointing out that, by making the human soul so fundamentally different from the souls of animals, the view already postulates a pretty wide gap between human beings and the rest of the animate creation – a gap which is papered over but not narrowed by the claim that each has a soul which is the “form of its body.” Beyond this, however, I have claimed that the work actually done by the human soul is surprisingly limited. In fact, I shall

argue that this work amounts practically just to the difference made in human life by the fact that humans are rational creatures. That is no small matter, to be sure. But it falls considerably short of the wide-ranging influence on human biology one would expect, given that the soul is said to be the substantial form of the body. To see why I say this, consider that, by the time the soul is infused into the fetus (which may be a relatively late-stage fetus, if we follow Aquinas), most of the essential biological structures are already in place, albeit in early stages of development. Consider, also, that very similar structures exist in the fetal development of an animal – say, a gorilla or a chimpanzee – at a comparable stage. Now, the chimpanzee fetus, quite unaided by any “subsisting immaterial form,” will naturally develop into the magnificent animal we have recently learned so much about. This animal will function biologically in ways that are essentially similar to the functioning of a human being, and will exhibit a rich and complex mental, emotional, and social existence which, while lacking in some of the distinctive features of human life, nevertheless compels our admiration and wonder. Must we not suppose that the human fetus, if it were not infused with an immaterial soul, would be capable of the same sort of development?

To this it will be replied that the “sensitive soul,” which is said to guide the development of the animal organism, does not continue in the human fetus but is rather replaced by the subsisting human soul. And on account of this, it is indeed the immaterial human soul that guides, and accounts for, the development of the biological structures and functions of the mature organism. To which I reply: that is what the theory says, but this metaphysical fact, if it is a fact, seems to make no biological difference at all. According to the theory, nothing carries over from the early fetus to the human being but prime matter. But a physician observing a fetus at the moment of infusion will not see its characteristic structures and processes suddenly disappear, in order to reappear an instant later under the supervision of the subsistent soul. Biologically, everything carries on just as before. The only difference, even granting the theory, is that certain distinctively rational capacities will gradually become apparent, capacities which by hypothesis would not be present in the absence of an infused soul. It is very difficult to avoid the impression that the configurational state of the organism remains (whether or not it is still called the “sensitive soul”), and that it retains the causal efficacy which it had before the soul’s infusion. And this, of course, leaves quite a lot less work to be done by the immaterial soul. That is what I meant by saying that, practically speaking, the difference made by the soul is limited to human rationality.

The other criticism is that the case for the existence of subsistent souls is weak. The main philosophical reason given for their existence is that the activity of reason has no material organ. Unfortunately, this is one point at which it is very difficult for us to agree, in the light of contemporary brain science. Stump, for one, simply admits that the traditional view was mistaken about this (Stump 2003, 213).<sup>2</sup> We should not, however, overlook the point that this is very nearly the *only* philosophical argument which is available for the immateriality of the soul. (There will of course be the theological argument from an afterlife. But we would greatly prefer not to have a vitally important metaphysical position rest solely on such a theological argument.)

But why must this be so? Are there not other arguments which can be used, and which are in fact used, by dualists and others opposed to materialism? Indeed there are. There is the notorious “hard problem” of consciousness – the problem, as some have put it, of how soggy gray brain-stuff gives rise to technicolor phenomenology. There is the problem of intentionality: how is it that a state of a physical system can represent, can “mean,” something entirely different from itself, perhaps something that does not exist at all? There

is the problem of teleology: how can mere matter behave in a way that is genuinely purposive, as opposed to merely giving the impression of purpose due to a clever design (say, like a thermostat)? In particular, how can the course of our inquiries be oriented to the desire for truth, rather than being guided merely by the mechanistic processes of our bodies and nervous systems? There is – one of my own favorites – the unity-of-consciousness argument: how can a mere collection of physical parts exhibit a unified consciousness, given that a complex conscious state cannot exist “parceled out” among the many pieces of a complex physical system?

There are, then, all of these arguments, though of course there are also answers to them by materialists, and so the controversy rages on. But for the Thomistic dualist, there is a more fundamental problem, namely this: *the Thomistic dualist cannot use any of these arguments, because she has already conceded all of the points in dispute*. She holds that animals, who by hypothesis have no subsistent souls but only the appropriate configuration of the material organism, exhibit all of the phenomena in question: sensory experience, pleasure and pain, intentionality (who that has loved, and been loved by, a dog can doubt that the dog has ideas of particular individuals?), purposefulness, and unified states of consciousness. For her, none of these phenomena gives any purchase for an argument to the existence of an immaterial soul. And other arguments that will serve that purpose for her may be hard to come by.

Finally, there is the fact – and I do take it to be a fact – of biological evolution. Evolution fits uncomfortably with the Thomistic view for at least a couple of reasons. For one thing, there is the “species essentialism” clearly articulated by J. P. Moreland:

For the Thomist, a genus and a species in the category of substance are not degreed properties. That is, either they are fully predicable of an entity or they are absent . . . An entity either is or is not a human person or some other type of person. (Moreland and Rea 2000, 224–225)<sup>3</sup>

This rules out immediately the widely held notion that, in a broad evolutionary perspective, biological classifications such as genus and species are blurred and mutable. To be sure, it need not be the case that “species” in Moreland’s metaphysical sense maps directly onto the biological concept of species, a concept which is itself in dispute. But in view of the detailed way in which the soul directs and energizes the development of the organism, the kinds of souls must themselves be very numerous; kinds of organisms that are different in any major or substantive way will need to have different kinds of souls overseeing their growth and functioning. These “species” of souls, furthermore, are immutable, and immutably distinct. This fits poorly, to say the least, with the evolutionary idea of gradual development of new species. Furthermore, the sharp break posited between human beings and other animals belies the fact that many of the distinctive human attributes, psychological as well as physical, have their precursors in the apes from which we have come. To say that is not to deny the important differences that clearly exist between humans and all other animals. But underlying those massive differences there is enough continuity that it becomes less than credible to explain the differences by saying that humans have immaterial souls and animals do not.

### 8.3 A Modified Thomistic View

We now turn to a modified version of the Thomistic view, as presented by J. P. Moreland (Moreland and Rea 2000).<sup>4</sup> This view differs from the standard version in several ways, of

which two are most important. First, the possession of immaterial subsistent souls is not limited to human beings, but rather is common to all animals, indeed to all living things.<sup>5</sup> A little reflection shows that this modification overcomes one of the main objections offered above against the original Thomistic view. By extending the possession of subsistent souls to all living things, the view bridges the gap between humans and other living creatures. This does not mean, of course, that animal souls, let alone those of plants, are similar in all respects to those of humans. We can gain some grasp of the differences by observing the differences in the capacities typically exercised by well-developed adult members of the various species. Nevertheless, the souls of animals and plants belong to the same metaphysical category as the souls of human beings.

A second modification is that the soul must be supposed to be present, and operative, *from the very beginning of the organism's existence*. This removes what can only be called the absurdity of human beings who were never conceived, and never existed as early term fetuses! (Remember that, on the Thomistic view, the infusion of the rational soul brings into existence a *new organism*, sharing nothing with the previously existing fetus but the prime matter of which each is composed.) This change also avoids the awkwardness of supposing that the physical development of the body is at first accomplished by the sensitive soul, a material form, but is then taken over by the rational soul, a quite different sort of entity. Further, these two modifications, taken together, mean that the revised view no longer concedes that a purely material being (such as an animal, or an early term fetus is conceived to be on the traditional view) can exhibit many of the distinctive mental functions and properties.

By positing the infusion of souls at the very beginning of an organism's life, the view is able to assert that infused soul is unambiguously involved in, and necessary for, both the development of the essential biological organs and their successful functioning. Moreland develops this relationship with considerable sophistication, as is shown in the following quotations:

The soul is a substance with an essence or inner nature . . . [which] contains, as a primitive unity, a complicated, structural arrangement of capacities and dispositions for developing a body. Taken collectively, this entire ordered structure can be called the substance's *principle of activity* and will be that which governs the precise, ordered sequence of changes that the substance will go through in the process of growth and development. (Moreland and Rea 2000, 204; emphasis in original)

The various physical and chemical parts and processes (including DNA) are tools – instrumental causes employed by higher-order biological activities in order to sustain the various functions grounded in the soul. Thus the soul is the first efficient cause of the body's development as well as the final cause of its functions and structure internally related to the soul's essence. The functional demands of the soul's essence determine the character of the tools, but they, in turn, constrain and direct the various chemical and physical processes that take place in the body. (Moreland and Rea 2000, 205)

Finally, the modified Thomistic view overcomes the objection to the effect that there is little or no evidence for the existence of a subsistent soul. That objection, it will be recalled, was based on the fact that all of the phenomena appealed to in standard arguments against materialism – sensory experience, pleasure and pain, intentionality, purposefulness, and unified states of consciousness – are attributed on the Thomistic view to beings that are

purely material, namely to nonhuman animals. Since on the modified view these animals also are in possession of subsistent souls, those arguments come back into play.

On the other hand, these modifications do nothing to overcome the problem with evolution noted for the original Thomistic view. For those who find the evidence for evolution compelling, this must remain a serious objection to the modified Thomistic view.<sup>6</sup> There is, furthermore, an additional objection that comes into view when we confront the modified view with recent developments in biological science. It is abundantly clear from the material quoted above that the view is committed to *vitalism*, which holds

first, that in every living organism there is an entity that is not exhaustively composed of inanimate parts and, second, that the activities characteristic of living organisms are due, in some sense, to the activities of this entity. (Beckner 1967, 254)

Now it is an undisputed fact that vitalism has had an extended history in the biological sciences, and has been advocated by many prominent biologists. It is also a fact, however, that the view has lost credibility for almost all biologists for about a century; it is now universally regarded as a failed research program that has been abandoned for good reason.<sup>7</sup> The association of Moreland's version of dualism with this failed research program is not, to put it gently, a point in its favor.

Moreland is aware of this complaint, and devotes some effort to defending his view against it. In brief, his answer is that the cruder forms of vitalism were rightly rejected because they made unjustified assumptions concerning the "individuated essence," assumptions which his view has no need to accept. He goes on to say that

A more adequate vitalism – if we wish to use this term of the Thomistic substance view (and we prefer the term *organicism*) – grounds the doctrine of substance in factors like the irreducible organic, holistic relation among parts to parts and parts to whole, the species-specific immanent law of organization and development, and the internal structural form and normative functioning found in living things. (Moreland and Rae 2000, 217)

Through these considerations Moreland seeks to insulate his view from the failure that such endeavors have met with in the past. Whether this is a promising and attractive line to take must be left to the reader to decide.

## 8.4 A Way Forward?

Without doubt, it is somewhat discouraging that a proposal which starts out with such good intentions finds itself surrounded by intractable difficulties. I will now suggest that a further modification of Thomas's view can yield a result that overcomes these difficulties and is a viable contender against other existing mind-body theories. The result may not qualify strictly as a Thomistic view, but it will arguably be in the spirit of Aquinas (who, let us not forget, took a good deal of flak over his incorporation into his theology of the then-novel insights of Aristotelian science and philosophy).

As with Moreland's modified Thomistic view, the human soul must be accounted as a thing of the *same ontological kind* as the souls of other organisms – at least, of other animals that are judged to have some degree of sentience or awareness. This emphatically does not

mean that “all souls are created equal.” The differences between humans and other animals are both real and vitally important, and must be respected. But there are powerful indications of similarity and continuity underlying those differences, and they also need to be respected. If a human soul needs to be an immaterial substance (and I believe that it does), the same must be said of the souls of lesser creatures. The wide gulf between them postulated by the traditional Thomistic view simply is not credible. This in turn means that the view does not concede that a purely material being (such as an animal, or an early term fetus is conceived to be on the traditional view) can exhibit many of the distinctive mental functions and properties. In this way it leaves the arguments against materialism in play, as they are not for traditional Thomistic dualism.

In view of the evolutionary context of contemporary thought, a further modification that is needed is to abandon the essentialism that takes genus and species among living creatures to be immutable “kinds.” Species are generally stable in the short term, but over the vast stretches of geological time the transmutation of species simply cannot be denied.

And finally, I propose that, rather than consider souls as being created individually by God and infused into their organisms, we should embrace an emergentist view in which the souls are generated naturally as a result of the structure and functioning of the biological organism – in the case of humans and the higher animals, of the brain and nervous system. This of course is a large subject; I refer the reader to the discussion of it under the heading of “emergent dualism” (see Chapter 4, this volume). This is the kind of view I am proposing as, if not actually a version of Thomistic dualism, at least a worthy successor to it. An emergentist view clearly fits far more naturally with biological evolution than does creationism. If it pleased God to create a situation in which our bodies would naturally evolve, it makes a more harmonious picture if the same is true of our minds and souls. And there at least a couple of additional advantages to such a view. It avoids the question, which can be awkward for a creationist, as to which of the nonhuman creatures possess souls. Descartes, troubled by this difficulty, cut the knot by answering, None do! Animals, one and all, are mere automata, giving the illusion of conscious experience without really having any. Surely we cannot follow Descartes in this. But the question *is* awkward, in that any possible answer tends to be embarrassing. Set the cutoff high – say, mammal and birds – and you seem to have arbitrarily excluded from conscious experience many types of creatures that give every appearance of having such experience. But setting it low has problems as well; there is something off-putting in the notion of God creating souls individually for worms, mosquitoes, and intestinal parasites! Emergent dualism responds that all creatures possess souls if the biological organism has developed in a way that enables it to be the “emergence base” for a soul. We do not know, in detail, exactly what is required for this – but whatever the answer turns out to be, no problem is created thereby for the emergentist. Beyond this, an emergentist view can more readily make sense of the pervasive and intimate dependence of our thought-processes on the fine-grained working of our brains and nervous systems, a dependence for which creationists are driven to give what appear to be ad hoc explanations.

If Thomas had lived today, would he be an emergentist? It’s hard to say! On this topic, I am happy to yield to one far better placed than I am to estimate the result of placing him in a twenty-first-century context. Eleonore Stump poses the question, “Would Aquinas think that the mind is identical to the brain if he had known enough neuroscience? Given what he says about the separated soul, the answer, of course, has to be ‘no.’ But even if we ask about the mind before death, in its natural, embodied state, it seems less misleading to say that he could have thought that the mind emerges from the functioning of the brain, since the

human form on his account is dynamic rather than static” (Stump 1995, 520). If he would have thought that, then I say good for him!<sup>8</sup>

## Notes

1. David Braine, who is in general a defender of Aquinas, nevertheless suggests that Aquinas may have been deceived by confusing two different uses of the word “form” of quite different origins: on the one hand, we have the use of the word form to refer to “natures” or “predicates” and, on the other hand, we have this notion of “form” as a correlate of matter originating with the idea of shapes (Braine 1992, 499n).
2. “Aquinas mistakenly supposes that the intellect is tied to no particular body organ” (Stump 2003, 213).
3. The book as a whole is co-authored, but Moreland is responsible for the material on the metaphysics of the soul–body relation.
4. It should be pointed out that Moreland’s view comes in two versions, creationist and traducian. Only the creationist version will be considered here.
5. For the most part the discussion in *Body and Soul* focuses on animals and does not mention plants as such. But Moreland (2000, 213) states, “If by *soul* we mean individual nature, then every living organism is identical to its soul.”
6. Moreland’s view harmonizes nicely with a “progressive creationist” view in which God from time to time creates major new types of organisms at the point at which the environment has developed so as to be able to support them. I believe Moreland is disposed favorably toward progressive creationism for other reasons, so he would not find the evolutionary objection to his view to be a serious problem.
7. According to Beckner, “vitalism showed a curious tendency to come out on the losing side of biological controversy: after Darwin, it was anti-Darwinian, and it supported the view that organic syntheses could be effected only in a living organism. It also supported the useless and misleading conception of a primordial living substance, the protoplasm, a term and idea that unfortunately still survive” (Moreland and Rae 2000, 255).
8. My thanks to the *American Catholic Philosophical Quarterly* for permission to reuse material contained in “The Dialectic of Soul and Body,” *ACPQ* 87: 3.

## References

- Beckner, Morton O. 1967. “Vitalism.” In *The Encyclopedia of Philosophy*, edited by Paul Edwards, vol. 8, 254. New York: MacMillan.
- Braine, David. 1992. *The Human Person: Animal and Spirit*. Notre Dame, IN: University of Notre Dame Press.
- Brower, Jeffrey E. 2012. “Matter, Form, and Individuation.” In *The Oxford Handbook to Aquinas*, edited by Brian Davies and Eleonore Stump, 85–103. Oxford: Oxford University Press.
- Moreland, J. P., and Scott B. Rea. 2000. *Body and Soul: Human Nature and the Crisis in Ethics*. Downers Grove, IL: InterVarsity Press.
- Stump, Eleonore. 1995. “Non-Cartesian Substance Dualism and Materialism without Reductionism.” *Faith and Philosophy*, 12(4): 505–531.
- Stump, Eleonore. 2003. *Aquinas*. London: Routledge.



# Debating Cartesian Dualism

# Cartesian Substance Dualism

RICHARD SWINBURNE

## I

In his *Discourse on the Method* published in 1637 René Descartes wrote:

Examining attentively that which I was, I saw that I could conceive that I had no body, and that there was no world nor place where I might be; but yet that I could not for all that conceive that I was not. On the contrary, I saw from the very fact that I thought of doubting the truth of other things, it very evidently and certainly followed that I was; on the other hand if I had only ceased from thinking, even if all the rest of what I had ever imagined had really existed, I should have no reason for thinking that I had existed. From that I knew that I was a substance the whole essence or nature of which is to think, and that for its existence there is no need of any place, nor does it depend on any material thing; so that this “me,” that is to say, the soul by which I am what I am, is entirely distinct from body, and is even more easy to know than is the latter; and even if body were not, the soul would not cease to be what it is. (1972, 1.101)<sup>1</sup>

Descartes’s argument begins from one obviously true premise that (at the time when he was considering this argument) Descartes is thinking. It then proceeds by means of two principles about what is “conceivable” to the conclusion that Descartes is essentially “a thinking substance distinct from his body, which he calls his ‘soul.’” And of course if the argument shows this for Descartes it will also show the same for each human, that we are essentially souls. In this chapter I shall argue that although Descartes’s argument is not sound as it stands, it can easily be improved so as to make it sound. I shall then go on to reject what I consider to be the major objection to the argument, spell out the nature of the soul which the argument shows us to have and reject objections to this account of its nature, and finally support Descartes’s conclusion by a further argument using modern neurophysiology.

Many people have the reaction to arguments of this kind, that it could not be possible to reach substantial conclusions about human nature from an argument which seems to be mainly about what “we can conceive.” Thus Paul Snowdon (2014, 48) writes that an argument of this kind, “does fall foul of the general precept that serious ontological

conclusions should never be derived from theoretical requirements supposedly discerned in thought about personal identity.” All the great discoveries about the nature of the world made by science in the last five hundred years have been the result of scientists doing experiments and making observations to test theories formulated using technical terms and difficult mathematics. So how could we reach such a big conclusion about human nature as Descartes purports to have reached by a short argument which relies on no results of experiments and uses no sophisticated mathematics? That is an understandable but mistaken reaction. Although there is much in this argument about what Descartes “can conceive,” it starts from a crucial first premise which for Descartes and so for each other conscious human is, “I am thinking.” All that the subsequent argument does is to draw out what is involved in this premise which is apparently trivial but really astounding. Almost all of what science of the past five hundred years has taught us about the world concerns publicly observable objects including our own bodies and the predictable ways in which they behave. What Descartes does is to draw attention to something totally different from the publicly observable, our own conscious awareness, something about which we can be more certain than about anything else, and merely asks us to face up to what that involves.

So let us now look in more detail at Descartes’s argument. I need to begin by explaining some of the terminology which Descartes uses. Descartes make claims about what in the above standard English translation is translated as, he can “conceive,”<sup>2</sup> in the sense of “suppose to be true.” The argument has force only if we assume that what Descartes can conceive is conceivable, and what Descartes cannot conceive is not conceivable, so that “conceivable” means (as indeed it is often used in modern analytic philosophy to mean), “logically possible,” understood as “does not entail a contradiction.”<sup>3</sup> I now spell out what this involves in my own words. An indicative sentence normally expresses a proposition; it makes a claim about how things are in the world. Different sentences may express the same proposition, that is make the same claim about the world; for example, “The king is dead” expresses the same claim as any sentence of another language into which it is translated, such as “*Le roi est mort*” or “*Rex mortuus est.*” I understand by a proposition  $p$  “mini-entails” a proposition  $q$ , that in virtue of the meanings of the sentences which express  $p$  and  $q$ , anyone who asserts  $p$  is explicitly committed to  $q$ . Thus the proposition expressed by “it is 3 meters long” mini-entails the proposition expressed by “it is more than 2 meters long” because anyone who asserts “it is 3 meters long” is explicitly committed to “it is more than 2 meters long.” Someone wouldn’t have understood “it is 3 meters long” if they did not recognize that the statement commits them to “it is more than 2 meters long.” The explicit commitments of a sentence help to determine what the sentence means, and so help to make the proposition which it expresses the proposition it is. What determines whether one proposition  $p$  entails another proposition  $w$  which it does not mini-entail is whether there is a chain of mini-entailments from  $p$  to  $w$ , that is  $p$  mini-entails  $q$ ,  $q$  mini-entails  $r$ ,  $r$  mini-entails  $t$ , and so on until we reach  $w$ . Thus there is a chain of mini-entailments from the proposition expressed by “this is a closed rectilinear figure (in a Euclidean space) with three equal sides” to the proposition expressed by “this is a closed rectilinear figure (in a Euclidean space) with three equal interior angles.” Someone could believe the proposition expressed by the former sentence without realizing that the proposition implicitly committed them to the proposition expressed by the latter sentence, and so the former does not mini-entail the latter. But textbooks of geometry contain proofs beginning from the former proposition and proceeding by short steps, that is, each proposition mini-entailing the next proposition,

until we reach the latter proposition. And philosophical arguments also are often expressed in this form. Such proofs show that the former proposition entails the latter proposition. Although asserting the former proposition does not explicitly commit you to the latter, the latter is – as it were – buried in the former; and the proof makes this explicit.

A contradiction is a proposition of the form “p and not-p,” such as “he is more than 6 feet tall and he is not more than 6 feet tall.” It makes no sense to suppose such a proposition to be true; it asserts something and then denies it. So, also, any proposition which entails a contradiction could not be true under any circumstances, and so the proposition which it expresses is “logically impossible,” that is, “inconceivable.” Thus, a logically possible or “conceivable” proposition is one which does not entail a contradiction (see Note 2), and is therefore one in which it makes sense to suppose that it could be true – in this world or another world it is or might be true. A logically necessary proposition is one whose negation (the proposition which says that the former proposition is not true) entails a contradiction, and so must be true under all circumstances. “All squares have four sides” is logically necessary, because “it is not the case that all squares have four sides” entails a contradiction. To “conceive” a proposition is then to suppose of a conceivable proposition that it is true. And although anyone who understands a proposition knows all its mini-entailments, he may not realize what are all of the mini-entailments of the mini-entailments and so on; and hence he may believe that it does not entail a contradiction, when in fact it does. So we may be mistaken about whether a proposition is conceivable. But for many conceivable propositions it is very obvious that they are conceivable. All true propositions are conceivable; and innumerable false propositions are also obviously conceivable, since we can grasp fully and – if necessary – spell out at length how the world would be different if they were true, and “see” that nothing we say entails a contradiction. My desk is brown; but “my desk is red” is obviously conceivable; it entails no contradiction. Newton’s law of gravitational attraction – that (within limits imposed by relativity theory and quantum theory) every body attracts every other body with a force proportional to its mass and inversely proportional to the square of their distance apart – is true, but it is conceivable that it might be false. It makes sense to suppose that the universe might be such that the force of gravitational attraction is proportional to the cube (instead of the square) of the distance between bodies. Such a universe would be very unstable, and humans could not survive in it; but there is fairly obviously no contradiction entailed by supposing the universe to be like this. Likewise, it is often obvious that some proposition is inconceivable without our needing to show this by deriving a contradiction from it. It is obviously inconceivable that I now have the power to cause the tree outside my window never to have existed, but it might take a complicated argument to derive the contradiction from it. And in logic as in ordinary life, it is always rational to believe that things are as they seem to be, in the absence of contrary evidence; and so to believe that an apparently conceivable proposition is conceivable and that an apparently inconceivable proposition is inconceivable (see Note 3).

Descartes’s argument is concerned with the consequences of his premise, the proposition “I am thinking” which, at the time he is considering it, he surely knows infallibly to be true. Descartes normally uses “thinking” in a wide sense. He wrote:

By the word thought I understand all that of which we are conscious as operating in us. And that is why not only understanding, willing, imagining but also feeling, are here the same thing as thought. (1972, 1.222)

It thus includes all states which a subject experiences. We may call all such states conscious events. The argument then appeals to one supposed principle about what is logically possible, and one supposed principle about what is logically necessary. Descartes's supposed principle about what is logically possible is that, while he thinks and is aware of thinking, he can, compatibly with "I am thinking," "conceive that I [have] no body." For maybe Descartes just dreams that he has a body. So he claims that he can conceive and so it is logically possible that "I am thinking and I have no body." This premise certainly seems, to those who first hear it, immensely plausible. After all, in a dream anyone can be mistaken about the nature and location of their body, and the extent of their bodily powers (people often dream that they are flying); and no one can know infallibly that they are not dreaming. So surely anyone can be mistaken about whether or not they have a body, and so reflecting on what is involved in his thinking, Descartes sees no incompatibility in "I am thinking and I have no body." Yet if a proposition is conceivable, it entails no contradiction, and so – since, given the first principle, "I am thinking" entails "I exist" – "I am thinking and I have no body" entails "I exist and I have no body." So "I exist and I have no body" is also conceivable. Descartes's supposed principle about what is logically necessary is that, while he thinks, he cannot conceive that he does not exist, and takes that to show that not merely can he not conceive "I am thinking and I do not exist," but that "I am thinking and I do not exist" is inconceivable. So "I am thinking" mini-entails and so entails "I exist." That really is obvious. Only something which exists can do some act, such as thinking. If Descartes did not exist, he could not do anything at all.

Descartes claims that he is a "substance." He writes elsewhere (1972, 2.53) that "everything in which there resides . . . any property, quality, or attribute . . . is called a substance." That definition is not adequate to distinguish a substance from what Descartes regards as nonsubstances, such as properties, since properties themselves can have properties. The property of being green may have the property of being Amanda's favorite color. Descartes normally adds in other passages that a substance is "a thing which exists in such a way that it needs no other thing in order to exist" (1972, 1.239).<sup>4</sup> What he means by this is that in order for something to be a substance, it must be "conceivable" that the thing continues to exist, even if nothing else exists. Humans need oxygen in order to stay alive, but it is conceivable that – perhaps miraculously – a human might continue to live without oxygen. But that addition to the definition is unsatisfactory, because a substance could not exist without having properties – a substance can only exist if it has properties which make it that substance – my desk could not exist unless it had some particular color, some particular location, and so on. I suggest that the concepts of a "substance" and a "property" are so basic that they need to be illustrated by examples as well as defined by propositions. Rather, what Descartes needs for his argument is a definition such as "a substance is an individual component of the world which has properties," and to illustrate what he means by "an individual component of the world" by providing examples of such components. He could say, for example, that individual persons, animals, plants, planets, and so forth, are individual components of the world and have properties. So defined, the parts of substances are themselves substances – the tree outside my window is a substance, and so too are each of its branches. On this definition which I shall assume in future, Descartes is indeed a substance.

Substances have properties. A property is either a characteristic which belongs to one substance (such as being yellow, or having a mass of two kilograms) which is called a monadic or intrinsic property; or is a relation between substances (such as being taller-than

or lying-between). A property of a substance is an essential property of that substance iff (“iff” means “if and only if”) it is not conceivable that the substance could exist without having that property. Thus “occupying space” is an essential property of my desk: it is not conceivable that the desk could exist without occupying space. Descartes treats the words “property” and “quality” as synonymous, but sometimes calls an essential property of a substance an “attribute” of it (see 1.241–1.242). The essential properties of a substance taken together constitute its “nature and essence” (1.240). In the passage quoted above Descartes uses “nature” and “essence” as synonyms, and so this phrase should be read as “nature or essence.”

Descartes claims that it follows from his premise and the two principles that he is “a substance, the whole essence or nature of which is to think.” While “I think” entails “I exist,” it does not follow that I cannot exist while I am not thinking, for example, while I am in a dreamless sleep. And all that Descartes claims in the passage quoted above is the fact that he is “thinking” is his only reason for supposing that he exists. Elsewhere, however (see 2.210–2.211), he seems to suppose that we are conscious and so “thinking” all the time; really there is no dreamless sleep, it is simply that we do not remember all our dreams. But he does not need for his main argument such a dubious hypothesis. All that we need to understand by his claim that his “whole essence or nature” is to think, is that he exists as long as and only as long as he has the *capacity* for “thought” (in his sense), that is the capacity to have conscious experiences; and in this claim he is surely correct. Our conception of ourselves is that we exist while we have that capacity and cease to exist when we have lost that capacity. Admittedly, what constitutes my having that capacity is somewhat vague. Clearly, I have it if I am actually having conscious experiences, or if I am in a dreamless sleep from which I can be woken up, or if I am in a dreamless coma from which doctors with techniques superior to those available today would one day be able to wake me up. But for persons in a deep coma from which they cannot be woken up by any method which today’s doctors are able even to envisage, it is unclear whether we should say that they are alive (and so exist) so long as they can be woken by some physically possible method (that is, one compatible with the laws of nature), or whether they are dead but could perhaps be brought back to life again by some physically possible method. But it is clear enough that I do not have the capacity and so do not exist if I am not currently thinking and cannot be made to think by any physically possible (that is, non-miraculous) method, as is the case when my body begins to decay. Consequently, although what constitutes having the capacity is somewhat vague, within the limits of vagueness we should understand Descartes as claiming that he exists as long as and only as long as he has the capacity for thought. And since it is logically possible that Descartes should exist without having a body, having that capacity does not entail having a body; having a body is not part of that essence. When Descartes claims that his existence “does not depend on any material thing,” he is not denying that events in his body cause him to be alive and so to have the capacity for thought. But he is claiming that such events are not what his being alive consists in.

But while having a body is not part of his essence, Descartes believes – despite his doubts – that he does currently have a body, and so he must regard this body as an inessential part of himself, which he could cease to have. Thus, he is a substance which could exist without a certain part. But a substance cannot exist without any of its parts; and so if he were not to have a body and yet exist, then Descartes must at that time have another part distinct from his body, the essence of which part includes having the capacity to think. Descartes called such a nonbodily part his “soul.” And so Descartes concludes that he is essentially a soul,

that his “whole nature or essence” is to think. But this is where the argument, as Descartes expounds it, fails. For it is compatible with having the capacity for thought being necessary and sufficient for Descartes to be alive, and having a body not being necessary for that, that what is necessary and sufficient for Descartes to be alive is “having either a body or a soul.” Maybe the whole essence of Descartes is to “have either a body or a soul.” Maybe in fact his body does the thinking, but if he were not ever to have had a body, then in that case he would have had a soul which would then do the thinking. And that in turn is compatible with his never existing without his body and so without ever having a soul. Therefore, the conclusion that Descartes is essentially a soul does not follow.

Descartes needs a principle stronger than the first principle that “I am thinking and I have no body” is conceivable. What he needs is the principle that “While I am thinking now, my body is suddenly destroyed” is conceivable, and I suggest that this stronger principle is also correct. If it is logically possible that while I am now thinking, I have no body, it is surely logically possible that while I am now thinking, I cease to have any control over my body or to be influenced by anything that happens in it; and in that case I would not at that time “have” a body, the body would be un-owned by me. After all, some claims about “near death” experiences of patients undergoing an operation report that patients claim to have experiences – for example, an experience of floating above the operating table – at the same time as the surgeons certify that they are “brain-dead.” And while we may suspect that really the patients did not have those experiences at exactly the same time as they were brain-dead, or that surgeons can judge a patient to be “brain-dead” while there is still some activity in the patient’s brain, we can certainly understand what the reports claim, and fairly evidently they do not entail contradictions. “While I am thinking now, I lose my body” is surely conceivable. But in that case surely it is also conceivable that the body is suddenly destroyed. So “While I am thinking now, my body is suddenly destroyed” is surely also conceivable.

What, however, is not conceivable is that any substance at all can lose all its parts simultaneously and yet continue to exist. A table may continue to exist if it loses a leg, but not if it loses all its legs and the table top at the same time. And organisms can continue to exist if over time they lose all their parts, so long as those parts are gradually replaced, each part being replaced over a period of time while other parts continue to exist. A tree can continue to exist if each cell is replaced by a similar cell at different times. But what is not conceivable is that every single part of the tree should be suddenly destroyed and yet that individual tree should continue to exist. Given this further logically necessary principle, it follows that if it is conceivable that Descartes continues to think over a period while his body is suddenly destroyed, he must have during the whole of that period another part which is not destroyed. That can only be a nonbodily part which can think, that is his soul. So it follows from Descartes’s being a substance capable of thought, that he always has a soul while he is thinking – for otherwise it would not be conceivable that he should go on thinking if he loses his body. Thus Descartes thinks in virtue of his soul thinking. It follows that Descartes currently consists of two parts – an essential part (his soul) and an inessential part (his body). And, Descartes would add, what applies to Descartes, applies to all other humans. Hence substance dualism – humans consist of two substances, a soul (the essential part) and a body (the inessential part). When improved in the way which I have suggested, the argument is – I claim – valid (the conclusion follows from the premise) and (since the premise is obviously true of anyone who is currently thinking) sound.

## II

The principal objection made to Descartes's original argument,<sup>5</sup> and which – if cogent – would apply to the improved version of what I have just presented, purports to give a reason for supposing that philosophical reflection cannot reach the kind of conclusion which I claim to have reached. This is that while Descartes's first principle that it is logically possible that "I am thinking and I have no body" may indeed be true in the sense that "I am thinking and I have no body" does not entail a contradiction, Descartes does not know to which substance he is referring by "I." He might, for all he knows, be referring to his body or his brain or a substance of some other kind yet to be discovered. For that reason he cannot reach a conclusion about the nature of the *substance* which he believes himself to be. Propositions which are logically possible in that they do not themselves entail a contradiction, but which would be logically impossible if you knew fully to which substances or properties their referring terms refer are just as strongly impossible as logically impossible propositions – and in a crucial sense they do "involve" a contradiction. The appreciation that there are such propositions leads us to see that logical modality (logical possibility, necessity, and impossibility) is a species of a wider kind of modality which is now generally called "metaphysical" modality. A metaphysically impossible proposition is one which involves a contradiction in this wide sense; a metaphysically necessary proposition is one whose negation involves a contradiction in this wide sense, and a metaphysically possible proposition is one which is not metaphysically impossible. Metaphysically possible/necessary/impossible propositions which can be shown to be such a priori (that is, by mere rational reflection) are logically possible/necessary/impossible; but there are also propositions which are metaphysically possible/necessary/impossible which can be shown to be such only a posteriori (that is, not by mere rational reflection, but by empirical investigation such as scientific inquiry). The view that there are such necessary/possible/impossible propositions which are so only a posteriori was introduced into philosophy by the work of Kripke and Putnam in the 1970s, and has gained widespread acceptance. So, objectors to Descartes's argument suggest, what matters is whether "I am thinking and I have a body" is metaphysically possible, not just whether it is logically possible.

Before illustrating how there can be such propositions whose metaphysical modality differs from their logical modality, I need to introduce a technical term due to Kripke (1980, 48), "rigid designator." A "rigid designator" is a word (or longer expression) that "in every possible world designates the same object," that is it refers to the same object (substance, property, or whatever), however, the object may change in respect of its nonessential properties. Proper names usually function as rigid designators; "Richard Swinburne" picks out me, whatever my occupation or age, whereas "The Professor of the Philosophy of Religion at Oxford" referred to me only while I held that position. "Green" is a rigid designator of a property; it always refers to the same color property of a substance, whether or not that property has such nonessential properties as "being Amanda's favorite color." But "the color of my walls" is not (normally used as) a rigid designator of a particular color property, since it (normally) refers to the color green only when my walls are green; if they are repainted red, then it refers to the color red.

Kripke (1980) and Putnam (1975) made their thesis that there are metaphysically necessary (or impossible) propositions that are not logically necessary (or impossible), and so logically possible propositions which are not metaphysically possible, plausible, by means of various examples. Putnam illustrated the thesis with the example of "water is H<sub>2</sub>O."



Plausibly the word “water” was used in the early nineteenth century as a rigid designator referring to the actual transparent drinkable liquid in our rivers and seas, and to whatever had the same chemical essence as that liquid. But, in ignorance of what that chemical essence was, people in the early nineteenth century could not be sure whether or not something (liquid or solid) that was not transparent or drinkable or in our rivers or seas was water. In fact, the chemical essence in question is  $H_2O$ . And water, being what it is in virtue of having that essence, could not *not* have that essence and still remain water. “Water is  $H_2O$ ” does not entail a contradiction, and yet it is just as strongly impossible that water should not be  $H_2O$ , as that there should be a square that does not have four sides. So “water is  $H_2O$ ” is metaphysically necessary, and what makes it so is the logically contingent (that is, not-necessary) fact that our rivers and seas consist of  $H_2O$ . Yet, in ignorance of this fact, the necessity of “water is  $H_2O$ ” could be discovered only by scientific investigation – that is, a posteriori. Another example used by Kripke was “Hesperus is Phosphorus” as understood by seventh century BCE Greeks, where “Hesperus” was the name of the bright star which appeared in the evening sky, and “Phosphorus” was the name of the bright star which appeared in the morning sky. In fact, these are the same heavenly body, the planet Venus, and since these words in fact picked out the same object, and that object could not *not* be identical with itself, “Hesperus is Phosphorus” is metaphysically necessary, but this could be discovered only by scientific investigation – that is, a posteriori.

What has made these necessary propositions “necessary a posteriori” is that any sentence expressing the proposition contains at least one rigid designator of which we learn the sense by being told that it applies to certain paradigm objects (substances, properties, and their kinds) having certain superficial features, but where – we are also told – what makes an object that object is the essence, physical, chemical, or whatever (of which we may be ignorant) underlying those features. It is the underlying features which constitute the logically necessary and sufficient conditions for an object to be that object. In these cases we know what the word means, in the sense that we know what are the criteria that determine whether or not the word applies to an object – for example, the criterion that it applies to a substance iff that substance has the same chemical essence as the stuff in our rivers and seas, but we cannot apply the criteria to determine what that essence is. In ignorance of the essence, we do not fully understand what we are saying about a substance (or kind or property) when we say that it is that substance (or whatever), and may be unable to recognize the substance (or whatever) when it does not have the specified superficial features. In this kind of case the lack of understanding is something to be cured not by a priori, but by a posteriori investigation. I call designators such as “water,” “Hesperus,” and “Phosphorus,” as used in the specified centuries, “uninformative designators.”

Most designators, however (and especially designators of properties), are not uninformative. They apply to objects solely in virtue of their observable (or experienceable) properties (or are defined by words that do so) and not in virtue of any essence underlying paradigm examples of the objects. For example, we don’t need to discover some underlying essence in order to determine whether some object is “red” or “fragile,” “10 meters long” or “rough.” I have called such designators “informative designators” (see Swinburne 2013, 11–14). In a recent book, David Chalmers (2012, 470) has introduced a concept of an “epistemically rigid expression,” for which he provides two different definitions, one of which seems to give that concept the same role as my concept of an “informative designator.” This definition of Chalmers defines an “epistemically rigid expression,” as one “whose referent [that is, the object referred to] can be known *a priori*,” that is in virtue

of knowing what the expression means; but he provides no account of what it is to “know a referent *a priori*.” My suggestion for what it is for an expression to be an informative designator is that “F” is an informative designator iff any person S who knows what expression “F” means can, under ideal (i.e., best logically possible) conditions, recognize when an object (substance, property, or whatever) is “F” and when it is not; or can, under ideal conditions, recognize when an object has the properties designated by expressions in terms of which “F” is defined. Conditions are ideal for recognizing when an expression “F” applies, I suggest, iff these four criteria are satisfied:

- 1 S’s faculties are working properly. (Thus if S was sighted and goes blind, his inability to recognize a red object doesn’t count against his possession of the concept.)
- 2 S is in the best possible position for recognizing the object designated. (If S is too far away, he may not be able to recognize whether one rod is longer than another rod, but that doesn’t count against his possession of the concept of “longer than.”)
- 3 The conditions are not illusory, that is abnormal conditions which make it seem to S that an object is “F” when it isn’t (or vice versa) – for example, when abnormal lighting makes it look like a surface is red when it is not.
- 4 (a) S knows which expressions “F” mini-entails and (b) which expressions, with respect to which S satisfies criteria (1), (2), (3), and (4a), mini-entail “F.” (The point of this criterion is that S should know the logical connections of F with other expressions.)<sup>6</sup>

What more could be required than the satisfaction of my four criteria for an expression to be epistemically rigid? To insist on some deeper understanding of the nature of the designated object would take us beyond what is required in order to know *a priori* to what the expression refers. Thus “is 2 meters long” is an informative designator because anyone who knows what that expression means can under ideal conditions recognize that some object does or does not have a length of 2 meters by measuring it with a ruler on which “2 meters” is marked. Conditions are ideal if the investigator’s faculty of vision is working properly, he makes the measurement by putting the ruler alongside the object when he is close to it; and the ruler is an accurately graded ruler, not a ruler on which marks of lengths are made which bear no relation to their true lengths. Also, S needs to know such logical truths as that “it is 2 meters long” mini-entails “it is more than 1 meter long,” and – if S can count up to 3, he needs to know that “it is 1 meter shorter than 3 meters long” mini-entails “it is 2 meters long.” For some words it takes some time to apply the tests for whether they apply – to determine, for example, whether some person is “influential” or “30 years old,” or “married.” We need to look at registers of births and marriages, conduct social surveys, and question witnesses. But in none of these cases does the applicability of the word depend on some unknown underlying essence which science has not yet discovered.

A word may be an informative designator, not because it itself designates an observable (or experienceable) property, but because it is defined in terms of words that do designate such properties. While “water” (as used in the early nineteenth century) is an uninformative designator of a property, “H<sub>2</sub>O” (as used by scientists today) is – I believe – an informative designator of a property, because it can be defined in terms of atoms, atoms can be defined in terms of electrons and protons, and the latter can be defined in terms of their electric charge and mass. “Charge” and “mass” are terms of which macroscopic values can be designated by informative designators. Thus the essential property of a proton, “has mass of  $1.67 \times 10^{-27}$  kilograms” can be defined in terms of the informatively designated property “has mass of 1.67

kilograms” and the informatively designated relation of “being less massive by 1/10 than” (used 27 times). Also we (or rather scientists who know what the words mean) know the minientailments of objects having such properties – for example, that a proton is more massive than an electron, and so exerts a greater gravitational force than does an electron.

If a sentence uses only informative designators, we understand fully to what we are referring by its words, and so – if we are clever enough – we should be able to determine by mere reflection on what it is asserting – that is, by pure *a priori* reasoning, not merely whether it or its negation are inconceivable in themselves, but whether they are inconceivable when we understand fully to what its words refer, that is whether they are metaphysically impossible. It therefore follows that a proposition is metaphysically necessary (impossible or possible) if the sentence expressing it expresses a logically necessary (impossible or possible) proposition when we substitute co-referring informative designators for uninformative designators (“co-referring” in the sense that they refer to the same objects). Take the sentence “Hesperus is Phosphorus” and substitute for each of the uninformative designators “Hesperus” and “Phosphorus” an informative designator of the planet which picks out the object by the material bits which constitute it and which we can see through a telescope. In each case the informative designator will be the same word, “Venus,” or a word logically equivalent to it. And then the sentence “Hesperus is Phosphorus” reduces to “Venus is Venus” (or some logically equivalent sentence) which of course expresses a logically necessary proposition. So it is metaphysically necessary that “Hesperus is Phosphorus” because it is logically necessary that “Venus is Venus.” Similarly, because “water” in fact designates  $H_2O$ , “water is  $H_2O$ ” is metaphysically necessary because “ $H_2O$  is  $H_2O$ ” is logically necessary.

This account of metaphysical modality will capture as metaphysically necessary (impossible or possible) almost all the examples of the “metaphysically necessary” (“impossible” or “possible”) offered by Kripke, Putnam, and others. And I cannot see that any recent philosophical discussions of metaphysical necessity have given any reason for supposing that there are any “metaphysical” necessities, impossibilities, or possibilities other than ones that are logical necessities, impossibilities, or possibilities or reduce to these when informative designators are substituted for uninformative designators.

Now – to return at last to this principal objection to Descartes’s argument. This objection (in my terminology) is that Descartes’s “I” is an uninformative designator. Maybe he is referring to his body or to his brain or to some unknown substance to be discovered by a future science, and in that case while the first principle in his argument, that “‘I am thinking and I have no body,’ is conceivable,” is true, that “I am thinking and I have no body” would be metaphysically impossible, because if we substitute for “I” an informative designator of the object referred to, we would get “my body thinks and I have no body” or some other contradiction. It would only be conceivable because Descartes does not fully understand the proposition.

Chalmers (2012, 372) denies that “I” is “epistemically rigid” (in my terminology an “informative designator”), and one reason<sup>7</sup> he gives for denying this is that “the subject is not in a position to know *a priori* what the ostended entity is.” But that objection is unsound. “I” is an informative designator, by the criteria which I expounded earlier. If I am favorably positioned with faculties in working order and not subject to illusion, I cannot fail to pick out myself. I am as favorably positioned as possible when I am as close as possible to the properties by which I am identifying myself; and that is when I am identifying myself as the subject of some conscious event, for example, as the person who is having the headache

which I feel. In that case, my faculties will be in working order and I cannot be subject to illusion. I am, in Shoemaker's (1994, 82) phrase, "immune to error through misidentification." I cannot know how to use the word "I," recognize that someone is having a conscious experience, and wonder whether it is I or someone else who is having it. And, of course, I also know the relevant mini-entailments. I can make simple inferences to and from sentences concerning myself; I can infer from "I am in this room" to "I am in a room," and so on. For the same reasons my proper name "Richard Swinburne" is an informative designator; anyone who is as well placed as I am to recognize whether or not it applies would see that it applies – although, of course, no one else can be as well placed as I am to recognize whether or not it applies. And what applies to me in regard to my proper name applies to every other human with respect to their use of "I" and their proper names. Hence the objection fails because "I think and I have no body" is not merely logically but metaphysically possible. And, of course, "I think and I do not exist" is, not merely logically, but also metaphysically impossible. And the principle that a substance could not lose all its parts simultaneously and yet continue to exist, is not merely logically but also metaphysically necessary. Hence the amended version of Descartes's argument which I presented in Section I is indeed sound, because we do know to what its words refer, and so Descartes can reach a conclusion about "the whole essence or nature" of the substance which is Descartes.

### III

But what is a "soul" of the kind which Descartes's argument, duly amended, shows to constitute the essential part of each of us? To answer this question, I need to introduce some more technical terms. The definitions which I am about to give are my definitions. Alas, these terms are used in different senses by different writers, but what matters is not the particular words which I am using but the senses in which I am understanding them, and what I have to say can easily be re-expressed in different terminology. For the rest of the chapter I will use only informative designators to refer to objects or properties, and so the logical modality of a proposition will be the same as its metaphysical modality, and so such words as "necessarily" used without any limiting adverb (e.g., "physically") are to be understood as "metaphysically necessary."

I define an event as a particular substance or substances having a particular property or properties at a particular time (such as this tie being brown at 4 p.m. on May 4, 2016, or Birmingham now lying between Manchester and London), or the event of the coming into existence of a substance, or the event of the ceasing to exist of a substance. Events are of two kinds, which I will call "physical" and "mental." I define a "physical" event as one which is necessarily equally accessible to any investigator, one to which no one person necessarily has "privileged" access. A brain event, for example, is a physical event; anyone suitably located, and so forth, can find out about my brain events as well as can anyone else. By contrast, I shall define a "mental" event as an event to which necessarily there is privileged access by the substance (e.g., the human) involved in the event. Pains and thoughts and the intentions which we are trying to fulfill are thus mental events. Whatever ways others have of finding out whether I have a headache or have a thought that today is Wednesday or am intending to shut the door, I could also use; like others, I could study my behavior by watching a film of it, or inspect my brain via some instrument. But I have an additional way of finding out whether I have a headache or have a thought that today is Wednesday, by

actually experiencing the headache or thought; and necessarily no one else has that way. Therefore, these events are mental events. I define a pure mental event as one whose occurrence does not entail the occurrence of a physical event. While my seeing a desk is a mental event, because I can know better than anyone else what I am seeing, it is not what I shall call a pure mental event since its occurrence entails the existence of a desk which is a physical event. But my seeming to see a desk is a pure mental event and so is my having a headache or the thought that today is Wednesday. Then a physical property is one whose instantiation in any substance always constitutes a physical event, a mental property is one whose instantiation in any substance always constitutes a mental event, and a pure mental property is one whose instantiation always entails the instantiation of a pure mental event. Thus "having a headache" is a pure mental property, and "weighing 140 pounds" is a physical property. A physical substance is one which has only physical properties as its essential properties, a mental substance is one which has at least one mental property as an essential property, and a pure mental substance is one which has only pure mental properties as its essential properties. In future I shall ignore impure mental properties (or events), since they can be analyzed as combinations of pure mental properties (or events) and physical ones.

In this terminology all conscious experiences, that is all events of the kind, which Descartes calls "thinking," are pure mental events. So too are beliefs and desires which each of us may have while we are not conscious of them – because we can become conscious of them and then know better than anyone else that we have them, and having them does not entail the occurrence of any physical event. And Descartes's having the capacity to think is also a pure mental event, since he can know better than anyone else whether he has that capacity, and his having it does not entail any physical event. A body or a brain are physical substances. The only properties essential for something being a body are physical properties, such as having a mass and location, a particular kind of shape, and being composed of physical organs of various kinds interconnected in various ways. What Descartes's argument shows is that he and all other humans are pure mental substances, in virtue of having as our only essential part, a soul which is a pure mental substance. That does not rule out there being other kinds of pure mental substances beside humans (and so their souls), for example higher animals and inhabitants of distant planets (with their souls having capacities to form different kinds of mental events from ours).

Each of us is a particular human in virtue of having a particular soul. So how does my soul differ from your soul? Since the only essential property of humans as such is a pure mental property, the difference between humans cannot consist in having different physical properties from each other. (Of course, humans may have different physical properties and so bodies from each other, but that is not what makes each of us who we are. That we each have differently looking bodies and bodies with different histories enables us to identify each other, but it is not what constitutes our identity.) Nor can the difference between humans consist in being the subject of qualitatively different successive pure mental events. It cannot be necessary for the identity of any human who has lived a normal-length life that he had any of the later pure mental events of his life; it would still have been me who continued to live if I had been adopted immediately after my birth and had a totally different sequence of pure mental events thereafter. So if any pure mental events are necessary for the identity of some human, it must be the very earliest ones. But it does seem implausible to suppose that it wouldn't have been me who came out of my mother's womb at a certain time just because the baby who did emerge had qualitatively different aches and pains from the

ones which I had. And no particular sequence of such early pure mental events can be sufficient for the identity of a particular human baby, because many different babies could have had qualitatively identical aches and pains. Hence the identity of a particular human is not constituted by having any particular pure mental (or other) properties. And so it follows that humans (including their essential parts, their souls) have what philosophers call thisness (or *haecceity*), which makes them (including their soul) and the particular human (including the particular human soul) the ones they are.

A substance has thisness if and only if there could have existed a different substance which has all the same properties (relational as well as monadic, past as well as present) as the former substance. Thus an electron would have thisness if the world would have been different; if instead of the actual electron there had always been a different electron which had all the same properties (e.g., mass, charge, and spatial relations to other fundamental particles at all times) as the former electron. Like most physicists, I do not myself believe that fundamental particles have thisness; any electron which had exactly the same properties as the actual electron would be the actual electron. But it is a disputed issue (see French and Krause 2006), whether any fundamental particles have thisness, and so it is not a fair objection to the view that humans have thisness that it introduces an otherwise unknown category into philosophy in order to avoid incoherence. If, as I have argued, it is neither necessary nor sufficient for a human to be me, that he have any particular properties, it follows that instead of me living the life I have lived, there could have been a different human who lived that life in all detailed respects. I could have lived your life and you could have lived my life; and the history of the world would have been the same in the respect that each of these lives, identical in the physical and mental properties of the human who lived them, would have been part of the world's history. But that history would nevertheless have been very different in respect of who lived the lives.

A consequence of this is that the operation of some law of nature could not conceivably explain why a particular soul and so a particular human came into existence. Laws of nature determine which instantiations of properties are (always or with a certain probability) followed by (or perhaps simultaneous with) other instantiations of properties. Thus they tell us that (always or with a certain probability) *any* body with a certain mass will cause *any* body at a distance from it to accelerate toward it with a velocity proportional to the product of their masses and inversely proportional to the square of their distance apart. This is the case with physical substances (if any of them have thisness), and also with humans, the fundamental laws of nature could determine which kinds of substance have thisness (e.g., that babies produced by a process initiated by human sexual intercourse do and quarks do not have thisness). They could also determine which beings of kinds that have thisness have the same thisness as each other (e.g., that the same human body is always the body of the same human, that is one who has the same thisness.) "Having thisness" and "having the same thisness as" are properties (respectively monadic and relational), the instantiation and transmission of which could be determined by laws of nature. But what no law of nature could determine is which of all the many possible humans who have all the same properties as each other come into existence as the result of some kind of process (e.g., a process initiated by human sexual intercourse). This is because the difference between such humans (including their souls) is not a difference of properties. Only chance or God could determine that I rather than someone else with all the same properties as myself emerged from my mother's womb. And yet the thisness of humans, unlike that of inanimate objects, is "knowable from the inside," that is by the human whose it is. I don't merely know that there

is a difference between me having a certain experience and someone else having that experience; I – unlike anyone else – know what the difference is. I can recognize the difference between me having a pain and someone else having a pain, when I am in the best possible situation for recognizing this – which is when I have a pain, and when I know as well as is possible for anyone who is not a certain other human that it is the other human that has a pain. And thereby I also recognize that the situation is very different from what it would be if instead of me having a pain there was someone else who had all the same other properties as I in fact have and also a pain qualitatively identical to my pain. My knowledge of what is the difference between me and someone else consists not in knowing a definition, but in an ability to recognize who is having the pain.

Descartes held that not merely are souls the essential part of each of us, and that we have (in my terminology) pure mental events in virtue of our souls having pure mental events, but that the soul of each human continually interacts with their body. He held that events in his body, and more particularly in his brain, cause most of his conscious events, such as his sensations and occurrent thoughts (although not our acts of will, that is our decisions); and that our acts of will cause events in the brain which in turn cause our bodily movements. It is indeed obvious that many of our sensations are caused by brain events, and given that the sensations are events in our souls, it follows that events in the brain cause events in the soul. And it certainly seems to us that our decisions affect what we do with our bodies.

There are, however, several arguments currently adduced, which purport to show that if there are souls in Descartes's sense, they do not interact with our bodies. First there is the objection very prominent in modern philosophy books that since conscious events and physical events (including brain events) are events of such very different kinds, we would need – in order to be justified in believing that they interact – an explanation of how they interact; and that no one can produce such an explanation. Yet once the two kinds of event are defined in the way I have defined them, it must seem obvious to almost everyone at least that brain events often cause sensations and desires. Sticking a needle into almost anyone does cause some brain event which causes them pain, and this happens whoever sticks the needle in, and whenever they do it. Depriving someone of liquid for several days causes some brain event which causes almost anyone to have a very strong desire to drink. These are evident simple causal connections of this kind well known to the human race for millennia. The fact that no one can provide an explanation of why there are these connections does not provide the slightest reason for supposing that they don't occur, for example, that sticking a needle into someone does not cause a brain event which causes pain. But, the objector may say, even if we do not need to explain how brain events cause pure mental events, surely we do need an explanation of how souls come into existence in order to believe that there are such things. But given my definitions and earlier arguments, it is evident that humans do have souls, and so they must come into existence caused by the development of a fetus, itself caused by human sexual intercourse. Humans have known for centuries how to produce substances of new kinds from other substances of very different kinds, such as how to produce edible plants from seeds, without having any remotely plausible theory of how this happens. Only with the development of chemistry in the early nineteenth century did anyone begin to have a plausible theory of how this happens; and even if no one had ever discovered the underlying chemistry, that would never have cast any doubt on the obvious fact that such things happen. But if we are justified in believing that brain events often cause pure mental events, and that the development of a fetus causes the

existence of a soul even though these are events of very different kinds, then it cannot be a good objection to the claim that pure mental events, understood as events in souls, often cause brain events, that these are events of different kinds.

Next, there is the objection that interactions between pure mental and physical events would violate a very general principle known as “the causal closure of the physical” (CCP), that physical events are caused only by physical events, which seems to many philosophers and scientists as implicit in modern science (see, e.g., Kim 2005, 155–156). As I noted earlier, it seems to us that our decisions – the intentions which we form – make a difference to which bodily movements we make; and in that case since our bodily movements are caused by our brain events, it must be that our decisions cause the brain events which cause these movements; and if so, CCP must be false. But, advocates of CCP claim, we deceive ourselves in supposing that our decisions cause our bodily movements. One way to try to show that our decisions never do cause our bodily movements would be to show that when anyone makes some movement which they formed an intention to make, the same sequence of brain events which caused the movement would have happened even if the person had not formed the intention to make those movements. In the course of the last twenty years many neuroscientists have been doing experiments which, they believe, will show just this. They have tried to show that what really happens when a person forms an intention to make a movement, is that a person’s brain events cause both their bodily movement and their intention to cause it – without the intention causing the bodily movement; hence the illusion that intentions cause bodily movements.

Yet the only evidence scientists could ever produce to show that supposedly intentional bodily movements would happen whether or not intentions occur, require them to know when someone forms an intention, and so to show that the intention makes no difference to whether or not the movement occurs. And how do they know whether some intention occurs at a particular time? Because the subjects whom they are using in their experiment tell them “I formed the intention at such-and-such a time.” And then perhaps the scientist could show that a certain kind of brain sequence (beginning at a time earlier than the formation of the intention) occurs and produces a certain bodily movement, both when they did subsequently form the intention to make that movement and when they didn’t. But why should scientists believe what subjects tell them about what was the time at which they formed a certain intention? Presumably because scientists believe that the subjects are intending, that is, have the intention, to tell them when they had an intention, and that it was because they had that intention that such words as “I formed the intention at 3.00.01” come out of their mouth. So the scientists believe the subjects because they believe that the words that come out of the subjects’ mouths are caused by their beliefs and intentions. So, the only way that neuroscientists could prove that our intentions don’t make a difference to what we do on some occasion is by relying on evidence which assumes that our intentions do make a difference to what we say on another occasion. Consequently, this program for proving CCP is self-defeating. And we are then back to what obviously seems to be the case, and to what we should believe in the absence of contrary evidence, that our intentions and therefore our decisions do make a difference to which movements we make, which must happen because they make a difference to our brain states.

The alternative way of showing that pure mental events make no difference to brain events would be to show that every brain event has another physical event as its immediately prior necessary and sufficient cause; for in that case no pure mental event could make any difference to what happens. Now suppose that some scientist claims to have established a



theory such as a precise deterministic form of the theory of the conservation of energy which has the consequence that every physical event, for example, a brain event, has another physical event as its immediately prior necessary and sufficient cause. Why should anyone else believe his theory? Presumably because he tells us that he has calculated that observations made by himself or others make it very probable that the theory is true. But we believe what he tells us because we believe that the words coming out of his mouth were caused by brain events caused by his intention that these words should come out of his mouth. And so again the theory can only be believed by others than the scientist by making an assumption which entails that the theory is false. And unless the scientist could make his observations or read about them having been made by others at the same time as making the calculation which lead him to believe that they support his theory – which is implausible – he will have to believe that he correctly remembers the past observations and the earlier stages of his calculations. But his memory will only be correct if his making or reading about the past observations cause memory traces in his brain which cause him later to remember the observations. Therefore, in assuming the correctness of his memory, he is assuming that his making or reading about past observations and calculations cause his brain events – which is again to assume that pure mental events cause brain events. So even the scientist who claims to have established such a theory from which it follows that pure mental events never cause brain events could not be justified in believing his own theory. It follows that no one could show that pure mental events do not cause brain events. Thus by “the principle of credulity”<sup>8</sup> that we should believe that things are as they seem to be, in the absence of counter-evidence, since it seems that pure mental events do cause brain events and we could not have any counter-evidence to show that they don’t, we should believe that pure mental events do cause brain events. So much for the second objection to Cartesian mind–brain interactionist dualism.

A third well-known objection is the “pairing” objection, that the Cartesian dualist cannot give an account of what determines that my soul and not your soul causes my actual brain events (that is, the brain events which have in fact been occurring in the brain that is currently mine), and that actual events in the fetus which becomes my body cause the existence and pure mental events of my soul, and not those of your soul. Kim (2005, 78–85) considers an example where two guns, A and B, are fired simultaneously and cause the death of two persons, Adam and Bob. Here there is an explanation of why gun A caused Adam’s death, and gun B caused Bob’s death. The explanation is that gun A was pointing in the direction of Adam, and gun B in the direction of Bob. But, Kim claims there is no comparable relation which relates particular souls to particular bodies, which could make it the case that a particular soul was causally related to a particular body. Yet given that there is a law to the effect that the brains of human fetuses cause the existence of a soul, then that is the relation that determines to which soul a brain is related – it is related to whichever soul that it originally caused to exist. But – given my earlier arguments – no such law could determine which soul it was which a particular fetus brain caused to exist. This is inevitable, given that souls have thisness. Thisness produces an inevitable limit to scientific explanation. But exactly the same problem would arise if fundamental particles have thisness. There could not be a scientific explanation of why the initial soup of matter-energy produced by the Big Bang caused the existence of these particular electrons rather than any other ones. Physics has arguments for and against the claim that fundamental particles have thisness but the issue is not settled merely by claiming that if they did have thisness, science couldn’t explain why some process caused the existence of this electron rather than that

electron. And I have given what I suggest are strong a priori arguments in favor of the view that souls do have thisness.

## IV

There are a number of similar thought experiments, which are such that it is physically possible that they should be done, but which it is not yet practically possible to do with existing techniques of brain surgery – which bring out the force of different stages of Descartes’s argument in the context of modern neuroscience. Consider just one such thought experiment, reflection on the possible outcomes of which shows that humans have thisness. Most and probably all pure mental events (contingently) have their correlates in the brain (and on an interactionist view often cause and are caused by those brain events). Now suppose that a human P undergoes an operation in which a small diseased part of his brain (a tenth of the whole brain) is replaced by a similar part from another brain. People sometimes do have parts of the brain of that size removed, and it is normally supposed that the human after the operation is the same as the human before the operation. It is plausible to suppose that adding a replacement part after removing the diseased part wouldn’t change the identity of the human whose brain it was. So it seems at least logically possible – and plausibly true – that this operation keeps the original human alive, that the new brain is still his brain. But now suppose that the disease spreads, and that each year a different tenth of P’s brain is removed and replaced by a similar part from another brain. At the end of ten years there is a human whose brain is made of entirely different matter from that of the original human. In consequence we would expect that human to have significantly different apparent memories (of events earlier than ten years previously) from those which the original P would have had without the operations and a character (arising from the pattern of his pure mental events) significantly different from what the original P would have had, since the brain correlates of his pure mental events are so different from those which P would otherwise have had. It seems therefore at least logically possible that the resulting human is not P. It seems, however, also at least logically possible – because the process has been gradual and each new part has become integrated into the brain before a new operation is done – that the resulting human is still P. Or if this doesn’t seem logically possible, there will surely be some earlier stage of the process (perhaps after six years) when it does seem logically possible that the human then existing is still P, and also logically possible that the human is not P. And there is no good reason to doubt that what seems to be logically possible is logically possible. Yet at that stage there is surely a truth about whether or not P has survived the series of operations up to that point (in the normal sense of “continued to exist” up to then), a truth which is not entailed by any truth about the amount of brain matter and the similarity of mental life shared between the original P and the resulting human. Even if the surgeons knew everything about every atom of the brain at each stage and how these atoms would behave under all conceivable circumstances, and even if they knew all pure mental events which the then existing human was having and all such events which the human existing at different times had or would have, they would still not know who the human existing at that stage was. And even if there is some third possible outcome of one or more of the operations – for example, the outcome that “it is neither true nor false that the resulting person is P” (which I myself do not recognize as a logical possibility), it surely cannot be logically necessary that any of the operations will have that

outcome. And so even if there is such a third possible outcome, there will still be a stage at which it is logically possible that the human then existing is still P, and also logically possible that the human is not P, and yet no conceivable evidence available to humans could show whether or not P has survived. It follows that being a particular human being is not fully analyzable in terms of properties (physical or mental) possessed by any individual or of the matter of which he is made. Hence it must consist in having a certain thisness.

## V

A human being on earth is composed of body and soul. But, I claim to have shown, the body is only a contingent part; the soul is the necessary part. Yet if we are to interact with other people we need bodies, since they provide us with a public presence where others can get hold of us and we can get hold of them. Without bodies we would be solitary creatures. At death the body ceases to function, and is often totally destroyed in the crematorium. Hence the goodness of the Christian doctrine of the resurrection of the body. But what makes a body my body is its connection with my soul; and it is only the continuing existence of my soul after my death which would make possible the resurrection of a body which is mine; that would consist in a body being joined again to my soul. I have not argued that the soul continues to exist after death. I believe that we need the Christian or some other religious revelation to show this. But what I have shown is that we each have a soul as our essential part, and so that the destruction of our bodies does not entail the destruction of us. It leaves open the possibility that the soul continues to exist and will be joined again to a body.

## Notes

1. The first figure of citations from *Works of Descartes* refers to the volume, and the subsequent figures to pages of the volume.
2. Descartes's original French word translated as "conceive," "*feindre*," means "pretend"; therefore the translation in what is one standard translation of Descartes's works, is inaccurate since one can only "pretend" something to be true if one believes it to be false, whereas one can "conceive" something to be true even if one also believes it to be true. But as nothing in Descartes's argument turns on supposing that he can "conceive" only what he believes to be false, and as much modern analytic philosophy discusses what is "conceivable" in the sense which I proceed to analyze, I have retained the translation of "*feindre*" as "conceive."
3. Hence some proposition is apparently conceivable if it seems to be logically possible, if it seems to make sense to suppose that it is true. Sometimes "conceivable" is used in modern analytic philosophy to mean "apparently logically possible." I am not using it in this sense.
4. Descartes goes on to qualify this by adding to "no other thing," "except for God." But this qualification is not relevant to the present argument.
5. This second section of my chapter relies on ideas developed more fully in Swinburne (2013).
6. This developed definition of what it is to be an "epistemically rigid designator" is similar to the definition of an "informative designator" which I gave in Swinburne (2013, 11–14). But there is the difference that in the book I state the equivalent of condition (4b) simply as the condition that "S knows which expressions mini-entail F." But this earlier version of this condition is far too strong. Being "indigo" mini-entails being "colored," but I do not need to know that in order to know a priori the extension of "colored" – unless I already know what "indigo" means.
7. Chalmers's other reason for denying this is relevant only to his alternative definition of "an epistemically rigid expression," as one that "picks out the same referent in every epistemically possible scenario" (Chalmers 2012, 470). My account of an "informative designator" does not include this requirement; and does not need it. So long as we know a priori to what we are referring by the words including "I" in some sentence, we will

understand fully the proposition which the sentence expresses, even if those words pick out different referents in other sentences.

8. This principle, that it always rational to believe (which I understand as “is always probably true”) that things are as they seem to be in the absence of contrary evidence, has various names. I call it “the principle of credulity.” It applies to all propositions. If it seems to me that I am seeing a tree then it is rational to believe that I am seeing a tree – in the absence of evidence that I am hallucinating. If it seems to me that  $5 + 7 = 12$ , then it is rational for me to believe that  $5 + 7 = 12$  in the absence of evidence that I have miscalculated. For fuller discussion and justification of this principle, see Swinburne (2016, 28–30).

## References

- Chalmers, David. 2012. *Constructing the World*. Oxford: Oxford University Press.
- Descartes, René. 1972. *Collected Works of Descartes*, translated by E. S. Haldane and G. R. T. Ross, 2 vols. Cambridge: Cambridge University Press.
- French, Steven, and Decio Krause. 2006. *Identity in Physics*. Oxford: Oxford University Press.
- Kim, Jaegwon. 2005. *Physicalism or Something near Enough*. Princeton, NJ: Princeton University Press.
- Kripke, Saul. 1980. *Naming and Necessity*. Oxford: Blackwell.
- Putnam, Hilary. 1975. “The Meaning of ‘meaning.’” *Minnesota Studies in the Philosophy of Science*, 7: 131–193. Republished in *Mind, Language and Reality: Philosophical Papers*, vol. 2, 215–271. Cambridge: Cambridge University Press.
- Shoemaker, Sydney. 1994. “Introspection and the Self.” In *Self-Knowledge*, edited by Q. Cassam, 118–139. Oxford: Oxford University Press.
- Snowdon, Paul. 2014. *Persons, Animals, Ourselves*. Oxford: Oxford University Press.
- Swinburne, Richard. 2013. *Mind, Brain, and Free Will*. Oxford: Oxford University Press.
- Swinburne, Richard. 2016. *The Coherence of Theism*. 2nd edn. Oxford: Oxford University Press.

# Against Cartesian Dualism

JAEGWON KIM

A kind of mind-body dualism seems implicit in common lore about ourselves as persons, something that appears to be shared by people almost everywhere, despite the inevitable cultural and historical differences. We think, intuitively and unreflectively, that although we each have a fully material and biological body, we also have mental or spiritual dimensions, like consciousness and a rational mind, that no “mere” material things can have. When we see the word “material,” we are apt to think “not mental” or “not spiritual,” and when we see the word “mental,” we tend to think “not material” or “not physical.” The shared lore surely does not amount to anything like a clear and determinate set of beliefs, but it is fair to say that a dualism of the mental and the material is an entrenched part of our ordinary thoughts about ourselves. It is a kind of “folk theory” of our nature as creatures in the world.

But folk dualism often goes beyond a mere duality of mental and physical states, activities, and processes. It is part of the folklore in many cultures and of most established religions that each of us has a soul, or spirit, which survives bodily death and decay, and that when our bodies die we continue to exist in that our souls continue to exist. Your soul is constitutive of your identity as an individual person; as long as it exists – and only so long as it exists – you exist. Further, it is our souls in which our mentality inheres; thoughts, consciousness, rational will, and other mental acts, functions, and capacities belong to souls, not to material bodies. To have a mind, on a view of this kind, is to have a soul. Ultimately, it isn’t really that we “have” souls. It is rather that each of us *is* a soul.

Mind-body dualism was famously developed by René Descartes into a philosophical system. Its metaphysical core was the dualism of two substances, the doctrine that there are two sorts of substances, material and mental, and that minds are literally substances, things or objects in the world, on a par with material objects like telephone poles and planets. Descartes’s theory of mind is the best known, and most influential, form of mind-body dualism, and it is the subject of the present chapter. One caveat before we begin: our goal is not so much a scholarly exegesis of Descartes as it is an examination of a point of view closely associated with him. As with other great philosophers, the interpretation of what Descartes “really” said, or meant to say, continues to be controversial. For this reason, the

dualist view of mind we will discuss is better regarded as “Cartesian” rather than as the historical Descartes’s.

## 10.1 Descartes’s Substance Dualism

As noted, the dualist view of persons that Descartes defended is a form of substance dualism, the doctrine that there are substances of two fundamentally distinct kinds in this world, namely, minds and bodies – or mental stuff and material stuff – and that a human person is a composite of a mind and a body, each an entity in its own right. Dualism of this form contrasts with monism, which holds that all things in the world are substances of one kind. Materialism, or its modern descendant, physicalism, has been the only form of monism that has been on the scene for the last half century, and I believe we can expect this to continue for some time to come. No serious competition is visible on the horizon, at least in the mainstream English-language philosophical world. Today, any proposed general ontology of the world, not just views about the mind-body relation, is defined by its relationship to materialism, the position that the world consists exclusively of bits of matter and structures made up of bits of matter, all behaving in accordance with physical law. Everything is an arrangement of matter, and living organisms and minded creatures are no exceptions.

Cartesian dualism is arguably the most prominent position opposed to materialist monism, claiming that in addition to material substances, there also are those that are immaterial. But what is a substance? Traditionally, two ideas have been closely associated with the idea of a substance. First, a substance is something in which properties “inhere”; that is, it is what has, or instantiates, properties.<sup>1</sup> Consider this celadon vase on my table. It is something that has properties, like weight, shape, color, and volume; it is also fragile and elegant. But a substance is not in turn something that other things can exemplify or instantiate. This feature of substances is reflected in language: a substance is the subject of predication, something to which we can ascribe predicates like “blue,” “weighs a pound,” and “fragile,” while it cannot in turn be predicated of anything else.

Second, and this is more important for us, a substance is thought to be something that has the capacity for independent existence. Descartes wrote: “The notion of a substance is just this – that it can exist by itself, that is without the aid of any other substance” (Descartes 1984b, 159). What does this mean? Consider the vase and the pencil holder to its right. Either can exist without the other existing; we can conceive the vase as existing without the pencil holder existing, and vice versa. In fact, we can, it seems, conceive of a world in which only the vase (with all its various parts) exists and nothing else and a world in which only the pencil holder exists and nothing else. This seems like one possible sense in which substances can be thought to enjoy the capacity for independent existence. On this understanding, if my mind is a substance, it can exist without any body existing, or any other mind existing. Consider the vase again: There is an intuitively intelligible sense in which its color and shape cannot exist apart from the vase, whereas the vase is something that exists in its own right. (The color and shape would be “modes” belonging to the vase.) The same seems to hold when we compare the vase and its surface. Surfaces are “dependent entities,” as some would say; their existence depends on the existence of the objects of which they are surfaces, whereas an object could exist without the particular surface it happens to have. As noted earlier, there is a possible world of which the vase is the sole inhabitant. Compare the

evidently absurd claim that there is a possible world in which the surface of the vase exists but nothing else; in fact, there is no possible world in which only surfaces exist and nothing else. For surfaces to exist they must be surfaces of some objects – existing objects.<sup>2</sup>

Thus, the thesis that minds are substances entails that minds are objects, or things, in their own right, just like material objects – it is only that on Descartes's view, they are immaterial; they have mental properties and can engage in mental activities of all sorts, like thinking, sensing, judging, and willing. Most importantly, as substances they are capable of independent existence, and this means that there is a possible world in which only minds exist and nothing else. So my mind, as a substance, can exist apart from my body, and so of course could your mind even if your body had perished.

Let us summarize the major tenets of Cartesian dualism we have reviewed so far:

- 1 There are substances of two fundamentally different kinds in the world, mental substances and material substances – or minds and bodies. The essential nature of a mind is to think, be conscious, and engage in mental activities of other sorts; the essence of a body is to have extension and be located in space.
- 2 A human person is a composite being (a “union,” as Descartes called it) consisting of a mind and a body.
- 3 Minds are diverse from bodies; no mind is identical with a body.

What distinguishes Descartes's metaphysics of mind from the views of many rationalists of his era, like Leibniz, Malebranche, and Spinoza, is his eminently commonsensical belief that minds and bodies are in causal interaction with each other. When we perform a voluntary action, the mind causes the body to move in appropriate ways, as when my thirst causes my hand to reach for a glass of water. In perception, causation works in the opposite direction: when we see a tree, the tree causes in us a visual experience as of a tree. That is the difference between seeing a tree and imagining or hallucinating one. This gives us the thesis of mind-body causal interaction:

- 4 Minds and bodies causally influence each other. Some mental phenomena are causes of physical phenomena and vice versa.

The only way my mind can influence the objects and events around me, as far as we know, is first to move my limbs or vocal cords in certain ways and thereby start a chain of events culminating in the effects I desire – like an open window, a painted wall, or a hat retrieved from the roof. But, as we all know, it is this unexceptional thesis of mind-body interaction that in the end brought down Cartesian dualism. The question of course was not whether the interactionist thesis was acceptable in itself; rather, it was whether the thesis was compatible with Descartes's ontology – that is, whether minds and bodies, sundered apart by the dualist theses (1) and (3), could be brought together in causal interaction as claimed in (4).

## 10.2 Why Minds and Bodies are Distinct: Some Arguments

Before we look at the supposed difficulties for Descartes's dualism, it will be helpful to review some considerations that might incline us to favor the dualist thesis. Before we criticize a doctrine, we must be able to appreciate it as a rationally motivated belief—to see why a

reasonable person might find it a plausible, perhaps even compelling, view to accept. The considerations we will review below are Cartesian in that they can be traced one way or another to Descartes's Second and Sixth Meditations, and that all are at least Cartesian in spirit.

First, let us run through some of Descartes's arguments based on supposed epistemological asymmetries between knowledge of minds and knowledge of material things. The following is a typically Cartesian argument:

Argument from indubitability:

I am such that my existence cannot be doubted.

My body is not such that its existence cannot be doubted.

Therefore, I am not identical with my body.

Therefore, the thinking thing that I am, that is, my mind, is not identical with my body.

A related epistemological argument attempts to exploit an epistemological asymmetry of another kind:

Argument from transparency:

My mind is transparent to me – that is, nothing can be in my mind without my noticing that it is there.

My body is not transparent to me in the same way.

Therefore, my mind is not identical with my body.

As stated, the first premise is most plausible when applied to mental phenomena like sensations and color sensing, such as pain and seeing yellow. When it comes to intentional states, most of us would be ready to acknowledge that at least some of our beliefs, desires, and emotions are beyond our cognitive reach – that is, “unconscious” or “subconscious.” And the second premise, too, probably needs qualifications. Cases of so-called proprioception, like knowledge of one's own bodily posture and orientation of our limbs, may present difficulties.

The following could well be the strongest and most plausible epistemological argument.

Argument from subjectivity:

For each mind there is a unique subject who has direct access to its contents.

No material body has a specially privileged knower – knowledge of material things is in principle public and intersubjective.

Therefore, minds are not identical with material bodies.

This argument is based on the idea that each mental occurrence has a unique “subject” with a direct, privileged cognitive access to it, whereas this does not hold for knowledge of material bodies. In principle, more than one cognizer can be in an equal position to gain knowledge of material objects and events.

There is much to be said about these arguments, but we should move on to the metaphysical arguments. Throughout the second and sixth of Descartes's Meditations, we find constant references to the essence of mind as thinking and the essence of body as being extended in space. By extension in space Descartes means three-dimensional extension, that is, bulk. Surfaces or geometric lines do not count as material substances; only things that



have a bulk count. A simple dualist argument could be formulated in terms of essences or essential natures, like this:

Argument from essence I:

My essential nature is thinking, and it does not include being a spatially extended thing.  
For I can conceive of myself as a disembodied thing.

My body's essential nature is being an extended thing in space.

Therefore, I am not a body.

Since I am a thinking thing, the thinking thing that I am is not a body.

Generalizing, no thinking thing is a body.

How could the first premise be defended? The Cartesian might make two points. First, as the "*cogito*" argument shows, I know that I exist only insofar as I am a thinking thing, and this means that my existence is inseparably tied to the fact that I am a thinking thing. Second, an essential nature of something is a property without which the thing cannot exist; when something loses its essential nature, that is when it ceases to exist. In this sense, being a thinking thing is my essential nature; when I cease to be a thinking thing, or a being capable of consciousness, that is when I cease to be, and so long as I am a thinking thing, I exist. On the other hand, my essential nature does not include having an extension in space. For I can clearly conceive of myself as existing without a body; there is no inherent incoherence, or contradiction, in the idea of my disembodied existence. Therefore, being an extended object in space is not part of my essential nature.

Descartes's commentators have objected: It may be, as Descartes claims, that our disembodied existence is conceivable, or imaginable, but from the mere fact that something is conceivable, however clearly and vividly, it does not follow that it is actually possible. And real metaphysical possibility is what the argument requires. A body moving at a speed exceeding the speed of light is conceivable, but we know it is not possible.<sup>3</sup> Or consider this: We seem to be able to conceive how Goldbach's conjecture, the proposition that every even number greater than 2 is the sum of two prime numbers, might turn out to be true, and also to conceive how it might turn out to be false. But Goldbach's conjecture, being a mathematical proposition, is necessarily true if true, and necessarily false if false. Thus it cannot be both possibly true and possibly false. But if conceivability entails possibility, it would have to be possibly true and possibly false. This issue about conceivability and metaphysical possibility has led to an extended series of debates too complex and contentious to enter into here.<sup>4</sup> It is a live issue in modal metaphysics and epistemology. We should note, though, that unless we use reflective and carefully scrutinized conceivability as a guide to possibility, it is difficult to know what other resources we can call on when we try to determine what is possible and what is not, what is necessarily the case and what is only contingently so, and other such modal matters.

Let us say that something is "essentially" or "necessarily" F, where F is a property, just in case whenever or wherever it exists (or in any possible world in which it exists), it is F. In this sense, we are presumably essentially persons, but not essentially students or philosophers; for we cannot continue to exist while ceasing to be persons, whereas we could cease to be students, or philosophers, without ceasing to exist. Consider, then, this argument:

Argument from essence II:

If anything is material, it is essentially material.

However, I am possibly immaterial – that is, there is a world in which I exist without a body.

Hence, I am not essentially material.

Hence, it follows (with the first premise) that I am not material.

This is an interesting argument. There seems to be a lot to be said for the first premise. Take something material, say, a bronze bust of Beethoven: this object could perhaps exist without being a bust of Beethoven – it could have been fashioned into a bust of Brahms. In fact, it could exist without being a bust of anyone; it could be melted down and made into a doorstop. If transmutation of matter were possible (surely this is not metaphysically impossible), it could even exist without being bronze. But could this statue exist without being a material thing? The answer seems a definite no: the bust could not turn into something immaterial, say a number, and continue to exist. So, for any material object, being material seems like a part of its essential nature. The acceptability of the argument, therefore, hinges crucially on the second premise. Is it possible that I exist without a body? That surely is conceivable, Descartes would insist. But again, the same question arises: Is something possible just because it is conceivable? In assessing his metaphysical arguments for dualism, therefore, the transitions from conceivability, or epistemological possibility, to metaphysical possibility become critical; their legitimacy will be the crux on which the fate of these arguments depends.

### 10.3 Descartes on Mind-Body Interaction

As will be recalled, the fourth component of Descartes's dualism is the thesis that minds and bodies causally influence each other. When we act, our desires and intentions cause our limbs to move; in perception, physical stimuli impinging on sensory receptors cause perceptual experiences in us. That may be a pleonastic truism, but the point is essential to our conception of ourselves as agents and cognizers. Unless our desires, beliefs, and intentions were able to cause our bodies to move in appropriate ways, how could human agency be possible? How else could we be agents who act and take responsibility for what we do? If objects and events in the physical world did not cause perceptual experiences and beliefs in us, how could we have any knowledge of what is happening around us? How could we know that we are holding a tomato in our hand, that we are coming up on a stop sign, or that a large bear is approaching us from the left?

Descartes has something to say about how mental causation works. In the Sixth Meditation, he writes:

The mind is not immediately affected by all parts of the body, but only by the brain, or perhaps just by one small part of the brain . . . Every time this part of the brain is in a given state, it presents the same signals to the mind, even though the other parts of the body may be in a different condition at the time . . . For example, when the nerves in the foot are set in motion in a violent and unusual manner, this motion, by way of the spinal cord, reaches the inner parts of the brain, and there gives the mind its signal for having a certain sensation, namely the sensation of a pain as occurring in the foot. This stimulates the mind to do its best to get rid of the cause of the pain, which it takes to be harmful to the foot. (Descartes 1984c, 59–60)

In *The Passions of the Soul*, Descartes identifies the pineal gland as the “seat of the soul,” the locus of direct mind-body interaction. This gland, Descartes maintains, can be moved

directly by the soul, thereby moving the “animal spirits” (bodily fluids in the nerves), which then transmit causal influence to appropriate parts of the body:

And the activity of the soul consists entirely in the fact that simply by willing something it brings it about that the little gland to which it is closely joined moves in the manner required to produce the effect corresponding to this desire. (Descartes 1985 I, para. 1, 343)

In the case of physical-to-mental causation, this process is reversed: Disturbances in the animal spirits surrounding the pineal gland make the gland move, which in turn causes the mind to experience appropriate sensations and perceptions. For Descartes, then, each of us as an embodied human person is a “union” or “intermingling” of a mind and a body in direct causal interaction.

### 10.4 Princess Elisabeth versus Descartes

In what must be one of the most celebrated letters in the history of philosophy, Princess Elisabeth of Bohemia, an immensely astute pupil of Descartes’s, wrote to him in May 1643, challenging him to explain:

how the mind of a human being, being only a thinking substance, can determine the bodily spirits in producing bodily actions. For it appears that all determination of movement is produced by the pushing of the thing being moved, by the manner in which it is pushed by that which moves it, or else by the qualification and figure of the surface of the latter. Contact is required for the first two conditions, and extension for the third. [But] you entirely exclude the latter from the notion you have of the soul, and the former seems incompatible with an immaterial thing. (Quoted in Garber 2001, 172)

Elisabeth’s demand is to the point and eminently understandable. First, see what Descartes has said about bodies and their motion in the Second Meditation:

By a body I understand whatever has determinate shape and a definable location and can occupy a space in such a way as to exclude any other body; it can be perceived by touch, sight, hearing, taste or smell, and can be moved in various ways, not by itself but by whatever else comes into contact with it. (Descartes 1984c, 17)

For Descartes, minds are immaterial; that is, minds have no spatial extension and are not located in physical space. If bodies can be moved only by contact, how could an unextended mind, which is not even in space, come into contact with material things, even the finest and lightest particles of “animal spirits,” thereby causing them to move? This seems like a perfectly reasonable question.

From a modern point of view, we can put Elisabeth’s challenge perhaps like this: For anything to cause a physical object to move or change, there must be a flow of energy, or transfer of momentum, from the cause to the physical object. But how could there be an energy flow from an immaterial mind to a material thing? What kind of energy could it be? How could anything “flow” from something *outside space* to something *in space*? If an object is going to impart momentum to another, it must have mass and velocity. But how could an unextended mind outside physical space have either mass or velocity? Again, the question is not about the intrinsic plausibility of Descartes’s thesis of mind-body

interaction; the question is whether this commonsensical interactionist claim is tenable within Descartes's dualist ontology of immaterial minds and material bodies.

Descartes responded to Elisabeth in a letter written in the same month (May 21, 1643):

I observe that there are in us certain primitive notions which are, as it were the originals on the pattern of which we form all of other thoughts . . . as regards the mind and body together, we have only the primitive notion of their union, on which depends our notion of the mind's power to move the body, and the body's power to act on the mind and cause sensations and passions. (Quoted in Garber 2001, 173)

Descartes is defending the position that the idea of mind-body union is a "primitive" notion – a fundamental notion that is intelligible in its own right and cannot be explained in terms of other more basic notions – and that the idea of mind-body causation depends on that of mind-body union.

But how plausible is Descartes's reply? In calling mind-body interaction a "primitive notion," isn't he only confessing he has no account? He might be right in that; there may be no more primitive ideas and principles that account for interaction, but is Descartes himself entitled to invoke it as his defense? Doesn't his doctrine about immaterial minds outside space and material bodies in space positively invite challenges of the kind Elisabeth raised?

We may note here the fact that, though Descartes seems to take material and immaterial substances as being on an equal footing causally, there is an important asymmetry between them. My mind can exercise its causal powers – on other minds as well as on bodies around me – only by first causally influencing my own body, and nothing can causally affect my mind except through its causal influence on my body. But my body is different: It can causally interact with other bodies quite independently of my mind. Moreover, my body – or my pineal gland – serves as the necessary causal conduit between my mind and the rest of the world; in a sense, my mind is causally isolated from the world by being united with my body. To put it another way, my body is the indispensable enabler of my mind's causal powers; it is by being united with my body that my mind can exercise its powers in the world. (This should make us wonder about the causal power of minds in disembodied states; we get to this topic shortly.)

To return to Elisabeth, she is not satisfied. She immediately fires back (June 1643):

And I admit that it would be easier for me to concede matter and extension to the mind than it would be for me to concede the capacity to move a body and be moved by one to an immaterial thing. (Quoted in Garber 2001, 172)

This is a remarkable statement; it may well be the first appearance of what is now known as the causal argument for materialism.<sup>5</sup> For she is in effect saying that in order to save mental causation, she would rather embrace materialism concerning mind ("it would be easier to concede matter and extension to the mind") than accept what she regards as an implausible account offered by her mentor.

## 10.5 The "Pairing Problem": Another Causal Argument

We will now develop another causal argument against Cartesian substance dualism. If this argument works, it will show not only that immaterial minds cannot causally interact with

material things situated in space but also that they are not able to enter into causal relations with anything else, including other immaterial minds. Immaterial objects would be causally impotent and hence explanatorily useless; positing them would be philosophically unmotivated and scientifically useless.

Here is the argument.<sup>6</sup> To set up an analogy and a point of reference, let us begin with an example of physical causation. A gun, call it *A*, is fired, and this causes the death of a person, *X*. Another gun, *B*, is fired at the same time (say, in *A*'s vicinity, but this is unimportant), and this results in the death of another person, *Y*. What makes it the case that the firing of *A* caused *X*'s death and the firing of *B* caused *Y*'s death, and not the other way around? That is, why did *A*'s firing not cause *Y*'s death and *B*'s firing not cause *X*'s death? What principle governs the "pairing" of the right cause with the right effect? There must be a relation *R* that grounds the cause–effect pairings, a relation that holds between *A*'s firing and *X*'s death and between *B*'s firing and *Y*'s death, but not between *A*'s firing and *Y*'s death or between *B*'s firing and *X*'s death. What is this *R*, the "pairing relation," as we might call it? We are not necessarily supposing that there is a single such *R* for all cases of physical causation, only that some relation must ground the fact that a given cause is a cause of the particular effect that is caused by it.

Two ideas come to mind. First, there is the idea of a *causal chain*: There is a continuous causal chain connecting *A*'s firing with *X*'s death, as there is one connecting *B*'s firing with *Y*'s death, whereas no such chains exist between *A*'s firing and *Y*'s death or between *B*'s firing and *X*'s death. Indeed, with a high-speed video camera, we could trace the bullet's flight from each rifle to its impact point on the target. The second idea is the thought that each gun when it fired was at a certain distance and in appropriate orientation in relation to the person it hit, but not to the other person. That is, *spatial relations* do the job of pairing causes with their effects.

The causal chain idea does not work as an independent solution to the problem. A causal chain, after all, is a series of events related as cause to effect, and interpolating more cause–effect pairs does not solve the problem. For it begs the question: we need to explain what pairing relations ground these interpolated cause–effect pairs. So it seems we must look to spatial relations – and more broadly, spatiotemporal relations – to generate pairing relations. Space appears to have nice causal properties; for example, as distance increases, causal influence diminishes, and it is often possible to set up barriers at intermediate positions to block or impede the propagation of causal influence. In any case, the following proposition is plausible:

- (M) It is metaphysically possible for there to be two distinct physical objects, *a* and *b*, with the same intrinsic properties and hence the same causal potential or powers; one of these, say, *a*, causes a third object, *c*, to change in a certain way, but object *b* has no causal influence on *c*.

The fact that *a* but not *b* causes *c* to change must be grounded in some fact about *a*, *b*, and *c*. Since *a* and *b* have the same intrinsic properties, it must be their *relational properties* with respect to *c* that provide the desired explanation of their different causal roles. What relational properties or relations can do this job? It is plausible to think that when *a*, *b*, and *c* are physical objects, it must be the spatial relation between *a* and *c* and that between *b* and *c* that are responsible for the causal difference between *a* and *b* vis-à-vis *c* (*a* was in the right spatial relation to *c*, *b* was "too far away" to exert any influence on it, etc.). At least, there

seems no other obvious candidate that comes to mind. Later we will offer an explanation of what it is about spatial relations that enables them to play this role.

Now consider the possibility of immaterial souls, outside physical space, causally interacting with material objects in space. The following companion principle to (M) seems equally plausible, and if an interactionist dualist wishes to reject it, she would owe us a principled explanation why.

- (M\*) It is metaphysically possible for there to be two souls, A and B, with the same intrinsic properties<sup>7</sup> such that they both act in the same way at a time and a change occurs in a material object, C. Moreover, it is the action of A, not that of B, that is the cause of the change in C.

What makes it the case that this is so? What pairing relation pairs the first soul, but not the second, with the material object? Since souls, as immaterial substances, are outside physical space and cannot bear spatial relations to anything, it is not possible to appeal to spatial relations to ground the pairing. What possible relations could provide causal pairings across the two domains, one of spatially located material things and the other of immaterial minds outside space?

Consider a variation on the foregoing example: There are two physical objects,  $P_1$  and  $P_2$ , with the same intrinsic properties, and an action of an immaterial soul causally affects one of them, say,  $P_1$ , but not  $P_2$ . How can we explain this? Since  $P_1$  and  $P_2$  have identical intrinsic properties, they must have the same causal capacity ("passive" causal powers as well as "active" causal powers), and it would seem that the only way to make them discernible in a causal context is their spatial relations to other things. Doesn't that mean that any pairing relation that can do the job must be a spatial relation? If so, the pairing problem for this case is unsolvable since the soul is not in space and bears no spatial relation to anything. The soul cannot be any "nearer" to, or "more properly oriented" toward, one physical object than another. Nor could we say that there was a causal barrier "between" the soul and one of the physical objects but not the other, for what could "between" mean as applied to something in space and something outside it? It remains a total mystery what nonspatial relations there could be that might help distinguish, from the point of view of an immaterial soul, between two intrinsically indiscernible physical objects.

Could there be causal interactions among immaterial substances? Ruling out mind-body causal interaction does not in itself rule out the possibility of an autonomous domain of immaterial minds in causal commerce with one another. Perhaps that is the picture of a purely spiritual afterlife envisioned in some religions and theologies. Is that a possibility? The pairing problem makes such an idea a dubious proposition at best. For suppose you are a substance dualist who wants causation in the immaterial realm: you must allow the possibility of there being three mental substances,  $M_1$ ,  $M_2$ , and  $M_3$ , such that  $M_1$  and  $M_2$  have the same intrinsic properties, and hence the same causal powers, and yet an action by  $M_1$ , but not the same action by  $M_2$  at the same time, is causally responsible for a change in  $M_3$ . If such is a metaphysically possible situation, what pairing relation could connect  $M_1$  with  $M_3$  but not  $M_2$  with  $M_3$ ? If causation is to be possible within the mental domain, there must be an intelligible and motivated answer to this question. But what mental relations could serve this purpose? Nothing comes to mind.

Consider what space does for physical causation. In the kind of picture envisaged, where a physical thing acts on only one of the two objects with identical intrinsic properties, what

distinguishes these two objects has to be their spatial locations with respect to the cause. Space provides a “principle of individuation” for material objects. Pure qualities and causal powers do not. And what enables space to serve this role is the fact that physical objects occupying exactly the same location in space at the same time are one and the same object.<sup>8</sup> This is in effect the venerable principle of the “impenetrability of matter,” which can usefully be understood as a sort of “exclusion” principle for space: material things compete for, and exclude one another from, spatial locations. From this it follows that if physical objects *a* and *b* bear the same spatial relations to a third object *c*, *a* and *b* are one and the same object. This principle is what enables space to individuate material things with identical intrinsic properties. The same goes for causation in the mental domain. What is needed to solve the pairing problem for immaterial minds is a kind of mental coordinate system, a “mental space,” in which these minds are each given a unique “location” at a time. Further, a principle of the “impenetrability of minds” must hold in this mental coordinate system; that is, minds that occupy the same “location” in this space must be one and the same. It seems fair to say that we do not have any idea how a mental space of this kind could be constructed. Moreover, even if we could develop such a space for immaterial minds, that still would fall short of a complete solution to the pairing problem; to solve it for causal relations across the mental and physical domains we need to somehow coordinate or fuse the two spaces, the mental and the physical, to yield unitary pairing relations across the domains. It is not clear that we have any idea where to begin.

If there are Cartesian minds, therefore, they are threatened with total causal isolation – from each other as well as from the material world. Our considerations presented do not show that causal relations cannot hold within a single mental substance (even Leibniz, famous for disallowing causation between monads, allowed it within a single monad). However, what has been shown raises serious challenges for substance dualism. If we are right, we have a causal argument for a physicalist ontology. Causality requires a spacelike structure, and as far as we know, the physical domain is the only domain with a structure of that kind.

## 10.6 Immaterial Minds in Space?

All these difficulties with the pairing problem arise because of the radically nonspatial nature of minds in traditional substance dualism. Not only are minds supposed to lack spatial properties but also not to be in space at all. So why not bring minds into space, enabling them to have spatial locations and thereby solve the pairing problem? Most popular notions of minds as immaterial spirits do not seem to conceive them as wholly nonspatial. For example, when a person dies, her soul is thought to “rise” from the body, or otherwise “leave” it, implying that before the death the soul was inside the body and that the soul is capable of moving and changing its locations. Sometimes the departed souls of our loved ones are thought to be able to make their presence known to us in various ways, including in a visible form (think about Hamlet’s ghostly father). It is probably impossible to make coherent sense of these popular ideas, but is there anything in principle wrong with locating immaterial minds in physical space and thereby making it possible for them to participate in the causal transactions of the world?

But the proposal to bring immaterial minds into space is fraught with complications and difficulties and probably not worth considering as an option. First there is the question of just where in space to put them. Is there a principled and motivated way of assigning a

location to each soul? We might suggest that I locate my soul in my body, you locate your soul in your body, and so on. That may sound like a natural and reasonable suggestion, but it faces a number of difficulties. First, what about disembodied souls, souls that are not “united” with a body? Since souls are supposed to be substances in their own right, such souls are metaphysically possible. Second, if your soul is located in your body, exactly where in your body is it located? In the brain, we might reply. But exactly where in the brain? It could not be spread all over the brain because minds are not supposed to be extended in space. If it has a location, the location has to be a geometric point. Is it coherent to think that there is a geometric point somewhere in your brain at which your mind is located? Descartes called the pineal gland the “seat of the soul,” presumably because the pineal gland is where mind-body interaction was supposed to take place, although of course his official doctrine was that the soul is not in space at all.

In any case, Descartes’s thinking about the “seat of the soul” no longer makes much sense. For one thing, there is no evidence that there is any single place in the brain – a dimensionless point at that – at which mind-body interaction takes place. As far as we know, mental states and activities are distributed over the entire brain and nervous system, and it does not make sense to think, as Descartes did in regard to the pineal gland, that there is a single identifiable organ responsible for mind-body causal interaction. Second, how could an entity occupying a single geometric point cause all the physical changes in the brain that are involved in mind-body causation? By what mechanism could this happen? How is energy transmitted from this geometric point to the neural fibers making up the brain? And there is this further question: What keeps the soul at that particular location? When I stand up from my chair and go downstairs to the kitchen, somehow my soul tags along and moves exactly on the same trajectory as my body. When I board an airplane and the airplane accelerates on the runway and takes off, somehow my pointlike immaterial mind manages to gain speed exactly at the same rate and begins to cruise at the speed of 560 miles an hour! It seems that the soul is somehow firmly glued to some part of my brain and moves as my brain moves, and when I die it miraculously unglues itself from my body and migrates to a better (or perhaps worse) place in afterlife. Does any of this make sense? Descartes was probably wise to keep immaterial minds wholly outside physical space.

Moreover, giving locations to immaterial minds will not in itself solve the pairing problem. As we saw, spatial locations of physical objects help solve the pairing problem in virtue of the principle that physical objects are individuated in terms of their locations. As was noted, this is the principle of the impenetrability of matter: distinct objects exclude one another from spatial regions. That is how the causal roles of two intrinsically indiscernible physical objects can be differentiated. For the spatial locations of immaterial minds to help, therefore, we need a similar principle of spatial exclusion for immaterial minds – or the principle of the impenetrability of mental substance – to the effect that distinct minds cannot occupy exactly the same point in space. What reason is there to think such a principle holds? Why cannot a single point be occupied by all the souls that exist, like the thousand angels dancing on the head of a pin? Such a principle is needed if we are to make sense of causation for spatially located pointlike souls. But this does not mean that the principle is available; we must be able to produce independently plausible evidence or give a credible argument to show that the principle holds.

When we see all the difficulties and puzzles to which the idea of an immaterial mind, or soul, leads, it is understandable why Descartes declared the notion of mind-body union to be primitive and not further explainable in terms of more fundamental ideas. Even a



contemporary writer (Foster 2001) has invoked divine action and theology as a reply to the question how a particular mind (say, your mind) gets to be united to a particular body (your brain). Like Descartes's appeal to "primitive notions," this seems tantamount to abandoning any attempt to understand the relationship.

## 10.7 Substance Dualism and Property Dualism

It has seemed to most contemporary philosophers that the concept of mind as a mental substance is fraught with too many difficulties and puzzles without compensating explanatory gains. In addition, the idea of an immaterial and immortal soul usually carries with it various, often conflicting, religious and theological constraints and associations that many of us would rather avoid in philosophy. For example, the traditional conception of the soul involves a sharp and unbridgeable gap between humans and the rest of animal life. Even if our own mentality could be explained as consisting in the possession of a soul, what might explain the mentality of nonhuman animals? It is not surprising that substance dualism has not been a prominent alternative in contemporary philosophy of mind. But there is no call to exclude it *a priori*, without serious discussion; some highly reputable and respected philosophers continue to defend it as a realistic – perhaps the only – option.<sup>9</sup>

To reject the substantial view of mentality is not to deny that each of us "has a mind"; it is only that we should not think of "having a mind" literally – that is, as there being some object or substance called a "mind" that we literally possess. Having a mind is not like – at least, it need not be like – having brown eyes or a good throwing arm. To have brown eyes, there must be brown eyes that you have. To "be out of your mind" or to "keep something in mind," you do not have to *have* some object – namely, a mind – which you are out of, or in which you keep something. Setting aside substance dualism, you can take having a mind simply as having a certain special set of properties, capacities, and characteristics, something that humans and some higher animals possess but flowerpots and rocks do not. To say that something "has a mind" is to classify it as a certain sort of thing – as a thing with capacities for certain characteristic sorts of behavior and functions, such as sensation, perception, memory, learning, consciousness, and goal-directed action. For this reason, it is less misleading to speak of "having mentality" than of "having a mind."

Substance dualism has played a very small role in contemporary discussions in philosophy of mind. Philosophical attention has focused instead on mental activities and functions – or mental events, states, and processes – and the mind-body problem has turned into the problem of understanding how these mental events, states, and processes are related to physical and biological events, states, and processes, or how our mental or psychological capacities and functions are related to the nature of our physical structure and capacities. In regard to this question, there are two principal positions: property dualism and reductive physicalism (also called type physicalism). Dualism is no longer a dualism of two sorts of substances; it is now a dualism of two sorts of properties, mental and physical. "Property" is used here in a broad sense: mental properties comprise mental functions, capacities, events, states, and the like, and similarly for physical properties. It is a catchall term referring to events, activities, states, and the rest. So property dualism is the view that mental properties are diverse from and irreducible to physical properties. In contrast, reductive physicalism defends

the position that mental properties are reducible to, and can be identified with, physical properties. As one would expect, there are various forms of both property dualism and reductionist physicalism. However, they all share one thing in common: the rejection of immaterial minds. Contemporary property dualism and reductive physicalism acknowledge only objects of one kind in the world – bits of matter and increasingly complex structures aggregated out of bits of matter (this anti-Cartesian position is sometimes called substance physicalism) (Latham 2001). Some of these physical systems exhibit complex behaviors and activities, like perceiving, sensing, reasoning, and consciousness. But these are only properties of material systems. Contemporary debates over the status of mind have for the most part focused on the relationship between these mental features and activities on one hand and the physical properties of the systems manifesting those mental features.

It is easy enough, probably too easy, to engage in free and loose talk about immaterial souls, or nonphysical objects, in causal relations. This happens in ordinary talk as well as in philosophy, religion, and theology. The causal pairing problem should serve as a warning that such talk may carry huge, largely unexpected and undischarged metaphysical obligations.

I understand that Descartes's idea of immaterial minds arose from his personal concern with the possibility of life after bodily death. Speaking for myself, surviving as an immaterial soul would provide me with little consolation. It would be a life in which I would be totally isolated from everything else in the world, from family and friends, books and music, hills and rivers. What would such a life be like when I am in it? What perceptions and thoughts would I have? What things would I be thinking about, to begin with? And I would have to live, and suffer, this life for eternity! It seems far more preferable to bid farewell to the world once and forever.

Emily Dickinson, who thought deeply about life, death, and immortality, could not shake her sense of mystery and bafflement about afterlife; like us, she found it "hopeless to conceive":

My life close twice before its close –  
It yet remains to see  
If immortality unveil  
A third event to me

So huge, so hopeless to conceive  
As these that twice befell.  
Parting is all we know of heaven,  
And all we need of hell. (Dickinson 1960, #1732, 702–703)

Portions of this chapter derive from Kim (2011, ch. 2).

## Notes

1. Descartes writes: "Substance: this term applies to every thing in which whatever we perceive immediately resides, as in a subject . . . By 'whatever we perceive' is meant any property, quality or attribute of which we have a real idea" (Descartes 1984a, 114).
2. Many philosophers in Descartes's time, including Descartes himself, held that, strictly speaking, God is the only being capable of independent existence and therefore that the only true substance is God, all others being "secondary" or "derivative" substances.

3. One might say that this is only a case of physical possibility and necessity, not possibility and necessity *tout court*. A more standard example would be the proposition that water = H<sub>2</sub>O. It is widely accepted that this is a necessary truth (though a posteriori), but that its falsehood is conceivable.
4. See the essays in Gendler and Hawthorne (2002). Gendler and Hawthorne's "Introduction" is a good starting point.
5. Typically, this argument begins with the causal closure of the physical as a premise. This is the proposition that any physical event, if it has a cause, has a physical cause. Then, given the assumption that mental events have physical effects, it follows that these mental events are physical events. See, for example, David Papineau (2001).
6. For a fuller presentation of this argument, see Jaegwon Kim (2005, ch. 2). For a dualist response to the pairing problem, see John Foster (1989).
7. If you are inclined to invoke the identity of intrinsic indiscernibles for souls to dissipate the issue, the next situation we consider involves only one soul and this remedy does not apply. Moreover, the pairing problem can be generated without assuming that there can be distinct intrinsic indiscernibles. This assumption, however, helps to present the problem in a simple and compelling way.
8. There is the familiar problem of the statue and the lump of clay of which it is composed (the problem of coincident objects). Some claim that although these occupy the same region of space and coincide in many of their properties (for example, weight, shape, size), they are distinct objects because their persistence conditions are different (for example, if the clay is molded into a cube, the clay, but not the statue, continues to exist). We must set this problem aside, but it does not affect our argument. Note that the statue and the lump of clay share the same causal powers and suffer the same causal fate (except perhaps coming into being and going out of existence).
9. For example: John Foster (1991); W. D. Hart (1988); William Hasker (1999); Alvin Plantinga (2006).

## References

- Descartes, René. 1984a. "Author's Replies to the Second Set of Objections." In *The Philosophical Writings of Descartes*, translated by John Cottingham, Robert Stoothoff, and Dugald Murdoch, vol. 2, 93–120. Cambridge: Cambridge University Press.
- Descartes, René. 1984b. "Author's Replies to the Fourth Set of Objections." In *The Philosophical Writings of Descartes*, translated by John Cottingham, Robert Stoothoff, and Dugald Murdoch, vol. 2, 154–178. Cambridge: Cambridge University Press.
- Descartes, René. 1984c. "Meditations on First Philosophy." In *The Philosophical Writings of Descartes*, translated by John Cottingham, Robert Stoothoff, and Dugald Murdoch, vol. 2, 12–62. Cambridge: Cambridge University Press.
- Descartes, René. 1985. "The Passions of the Soul." In *The Philosophical Writings of Descartes*, translated by John Cottingham, Robert Stoothoff, and Dugald Murdoch, vol. 1, 328–382. Cambridge: Cambridge University Press.
- Dickinson, Emily. 1960. *The Complete Poems of Emily Dickinson*, edited by Thomas H. Johnson. New York: Little, Brown.
- Foster, John. 1989. "A Defense of Dualism." In *The Case for Dualism*, edited by John R. Smythies and John Belof, 1–23. Charlottesville: University Press of Virginia.
- Foster, John. 1991. *The Immaterial Self*. London: Routledge.
- Foster, John. 2001. "A Brief Defense of the Cartesian View." In *Soul, Body, and Survival*, edited by Kevin Corcoran, 15–29. Ithaca, NY: Cornell University Press.
- Garber, Daniel. 2001. "Understanding Interaction: What Descartes Should Have Told Elisabeth." In *Descartes Embodied*, 168–188. Cambridge: Cambridge University Press.
- Gendler, Tamar Szabo, and John Hawthorne, eds. 2002. *Conceivability and Possibility*. Oxford: Clarendon Press.
- Hart, W. D. 1988. *The Engines of the Soul*. Cambridge: Cambridge University Press.
- Hasker, William. 1999. *The Emergent Self*. Ithaca, NY: Cornell University Press.
- Kim, Jaegwon. 2005. *Physicalism, or Something Near Enough*. Princeton, NJ: Princeton University Press.

- Kim, Jaegwon. 2011. *Philosophy of Mind*, 3rd edn. Boulder, CO: Westview Press.
- Latham, Noa. 2001. "Substance Physicalism." In *Physicalism and Its Discontents*, edited by Carl Gillett and Barry Loewer, 152–171. Cambridge: Cambridge University Press.
- Papineau, David. 2001. "The Rise of Physicalism." In *Physicalism and Its Discontents*, edited by Carl Gillett and Barry Loewer, 3–36. Cambridge: Cambridge University Press.
- Plantinga, Alvin. 2006. "Against Materialism." *Faith and Philosophy*, 23: 3–32.

# Non-Cartesian Substance Dualism

E. J. LOWE

Non-Cartesian substance dualism is a position in the philosophy of mind concerning the nature of the mind-body relation – or, more exactly, the *person*-body relation. It maintains that this is a relationship between two distinct, but not necessarily separable, individual substances, in the sense of “individual substance” according to which this term denotes a persisting, concrete object or bearer of properties, capable of undergoing change in respect of at least some of those properties as time passes. When such an object undergoes such a change, it undergoes a change of state, for a state of an object consists in its possession of some property at a time, or during a period of time. Using a more traditional terminology, we may speak of these states as modes of the object or individual substance in question.<sup>1</sup> As we shall see, non-Cartesian substance dualism differs from its more familiar cousin, Cartesian substance dualism, with regard to the class of modes that it considers persons – as opposed to their bodies – to be capable of possessing. Therefore, it takes a different view concerning what kind of individual substance a person – or, more generally, a subject of experience – should be taken to be. More precisely, whereas Cartesian substance dualism takes subjects of experience to be necessarily immaterial and indeed nonphysical substances, non-Cartesian substance dualism does not insist on this. As we shall also see, this distinctive feature of non-Cartesian substance dualism gives it certain advantages over Cartesian dualism, without compelling it to forfeit any of the intuitive appeal that attaches to its more traditional rival.

## 11.1 The Self as a Psychological Substance

The view that I wish to defend in this essay is that a human person, conceived as a subject of mental states, must be regarded as a substance of which those states are modes – and yet not as a *biological* substance: not, that is, as a living organism of any kind, even though a human person’s body is clearly just such an organism. What sort of substance, then? Quite simply,

a psychological substance. More specifically, a person, in my view, is a substantial individual belonging to a natural kind which is governed by distinctively psychological laws, with the consequence that individuals of this kind possess persistence conditions which are likewise distinctively psychological in character. However, saying just this about persons is consistent with regarding a person as being something like a Cartesian ego or soul and this is a position from which I expressly wish to distance myself. The distinctive feature of the Cartesian conception of a psychological substance is that such a substance is regarded as possessing only mental characteristics, not physical ones. And this is largely why it is vulnerable to certain skeptical arguments to be found in the writings of numerous philosophers during the past three hundred years, including Locke and Kant. The burden of those arguments is that if psychological substances – by which the proponents of the arguments mean immaterial “souls” or “spirits” – are the real subjects of mental states, then for all I know the substance having “my” thoughts today is not numerically identical with the substance that had “my” thoughts yesterday. The lesson of this is taken to be that – on pain of having to countenance the possibility that my existence is very much more ephemeral than I care to believe – I had better not identify myself with the psychological substance, if any, that is currently having “my” thoughts, or currently “doing the thinking in me.” But if I am not a psychological substance, then it seems gratuitous even to suppose that such substances exist. Certainly, their existence cannot be established by the Cartesian *cogito*.

But why should we suppose, with Descartes, that psychological substances must be essentially immaterial? Descartes believed this because he held a conception of substance according to which each distinct kind of substance has only one principal “attribute,” which is peculiar to substances of that kind, such that all of the states of any individual substance of this kind are modes of this unique and exclusive attribute (Descartes 1984, part 1, sect. 53). In the case of psychological or mental substances, the attribute is supposed to be thought, whereas in the case of physical or material substances, the attribute is supposed to be extension. On this view, no psychological substance can possess a mode of extension, nor any physical substance a mode of thought. However, I am aware of no good argument, advanced either by Descartes himself or by anyone else, in support of his doctrine of unique and exclusive attributes. Accordingly, I am perfectly ready to allow that psychological substances should possess material characteristics – that is, that they should include physical states among their modes. It may be that there is no material characteristic which an individual psychological substance possesses essentially, in the sense that its persistence conditions preclude its surviving the loss of this characteristic. But this, of course, does not imply that an individual psychological substance essentially possesses no material characteristics: indeed, to suppose that it did imply this would be to commit a “quantifier shift fallacy” of such a blatant kind that I am loath to accuse Descartes himself of falling prey to it.

How, though, does this repudiation of the Cartesian conception of a psychological substance help against the skeptical arguments mentioned a moment ago? Well, the main reason why those arguments seem to get any purchase is, I think, that in presupposing that psychological substances would have to be wholly nonphysical, they are able to take it for granted that such substances are not possible objects of ordinary sense perception. Such arguments represent psychological substances as being invisible and intangible and, as such, perceptible, at best, only by some mysterious faculty of introspection – and hence only by each such substance in respect of itself. But once it is allowed that psychological substances have quite familiar physical characteristics and can thus be seen and touched at least as

“directly” as any ordinary physical thing, the suggestion that we might be unable to detect a rapid exchange of these substances becomes as fanciful as the skeptical suggestion that the table on which I am now writing might “in reality” be a succession of different but very short-lived tables successively replacing one another undetectably. Whether one can conclusively refute such skepticism may be an open question, but I see no reason to take it seriously or to allow it to influence our choice of ontological categories.

I believe, then, that a perfectly tenable conception of psychological substance may be developed which permits us to regard such substances as being the subjects of mental states: which is just to say that nothing stands in the way of us regarding persons precisely as being psychological substances. The detailed development of such a conception is the topic of the remaining sections of this chapter, and for the time being it must suffice to say that I conceive of psychological substances as being the proper subject-matter of the science of psychology, which in turn I conceive to be an autonomous science whose laws are not reducible to those of biology or chemistry or physics. However, it will be appropriate to close the present section with some remarks on the relationship between psychological and biological substances, that is, between persons and their bodies. I restrict myself here, thus, to the case of persons who – like human persons – have animal bodies. With regard to this issue I am, as I indicated at the outset, a substantial dualist. Persons are substances, as are their bodies. But the two are not identical substances, for persons and bodies have different persistence conditions, just as do persons’ bodies and the masses of matter constituting those bodies at different times. I should perhaps emphasize here that where a person’s body is a biological substance, as in the case of human persons, the body is to be conceived of as a living organism, not as a mere mass of matter or assemblage of physical particles. Clearly, though, my version of substance dualism is quite different from Descartes’s. Descartes, it seems, conceived a human being to be the product of a “substantial union” of two distinct substances: a mental but immaterial substance and a material but nonmental substance. How such a union was possible perplexed him and every subsequent philosopher who endeavored to understand it. The chief stumbling block was, once again, Descartes’s doctrine of unique and exclusive attributes. How could something essentially immaterial be “united” with something essentially material? But psychological substances as I conceive of them are not essentially immaterial. Moreover, on my view, human persons are themselves just such psychological substances, rather than being a queer hybrid of two radically alien substances. I should perhaps stress, though, that my criticism of Descartes here pertains solely to his doctrine of “substantial union” and not to his conception of psychophysical causation, which I consider to be far more defensible (see further, Lowe 1992).

So, as far as the relationship between a person and his body is concerned, I do not see that this need be considered more mysterious in principle than any of the other intersubstantial relationships with which the natural sciences are faced: for instance, the relationship between a biological entity, such as a tree, and the assemblage of physical particles that constitutes it at any given time. Most decidedly, I do not wish to minimize the scientific and metaphysical difficulties involved here. I do not, for example, think that it would be correct to say that a person is “constituted” by her body in anything like the sense in which a tree is “constituted” by an assemblage of physical particles.<sup>2</sup> Nonetheless, it is my hope that by adopting a broadly Aristotelian conception of substance and by emphasizing not only the autonomy but also the continuity of the special sciences, including psychology and biology, we may see a coherent picture begin to emerge of persons as a wholly distinctive kind of being fully integrated into the natural world: a picture which simultaneously preserves the

“Lockean” insight that the concept of a person is fundamentally a psychological as opposed to a biological one, the “Cartesian” insight that persons are a distinctive kind of substantial particulars in their own right, and the “Aristotelian” insight that persons are not essentially immaterial beings.

## 11.2 The Self as a Bearer of Physical Characteristics

Let us recall that we are not required to deny that a person or self has physical characteristics and recall that, although we have to regard it as being distinct from its body, we are not required to think of the two as separable – except, perhaps, purely conceptually, or purely in imagination. But what physical characteristics can we allow the embodied self to possess? All of those physical characteristics that are also ascribable to its body? Or only some of these? Or some or all of these plus others that are not ascribable to its body? What we need at this point, above all, is a principled way of distinguishing between those statements of the form “I am *F*” – where “*F*” is a physical predicate – which are more properly analyzed as “I have a body which is *F*,” and those which can be accepted at their face value as being literally true. And here it may help us to consider whether or not the self is a simple substance – that is, whether or not it has parts. For if it does not, then no statement of the form “I am *F*” can be taken at face value if being *F* implies having parts. My own view is that the self is indeed a simple substance, and I shall argue for this later.

But does not every physical predicate imply divisibility into parts as Descartes held – this being the basis of one of his main arguments for the immateriality of the self? No, it does not. For instance, “has a mass of seventy kilograms” does not imply having parts. A self could, thus, strictly and literally have a mass of seventy kilograms without it following logically that it possessed various parts with masses of less than that amount. After all, an electron has a finite rest mass, but it does not, according to current physical theory, have parts possessing fractions of that rest mass. Again, “is six feet tall” does not, I consider, imply having parts, in the relevant sense of “part.” The relevant sense of “part” is this: something is to be accounted a “part” of a substance in this sense only if that thing is itself a substance. We may call such a part a “substantial part.” Simple substances have no substantial parts. We must, then, distinguish between a substantial part of a thing and a merely spatial part of it. A spatial part of an extended object is simply some geometrically defined “section” of it – not literally a section, in the sense of something cut out from it, but merely a region of it defined by certain purely geometrical boundaries. Thus, for example, the left-hand third of my desk as it faces me is a spatial part of it. It is doubtless the case that there is also a substantial part of my desk which at present coincides exactly with that spatial part namely, the mass of wood contained within that region. But it would be a category mistake to identify that mass of wood with the left-hand third of my desk (see Lowe 1998, chs. 5 and 7). Now, “is six feet tall” certainly implies having spatial parts, but it does not imply having substantial parts. Extended things – the claims of Descartes and Leibniz notwithstanding – can be simple substances.

So far, then, I can allow that physical statements such as “I weigh seventy kilograms” and “I am six feet tall” may be taken at their face value. But a statement like “I am composed of organic molecules” cannot be so taken, but must be analyzed rather as “I have a body which is composed of organic molecules.” Even so, it is surely evident that if “I weigh seventy kilograms” is literally true of me, it will be so only in virtue of the fact that I have a body



which weighs seventy kilograms. And, indeed, it seems clear that all of the purely physical characteristics which are literally ascribable to the self will be thus ascribable in virtue of their being ascribable to the self's body – so that we can say that the self's purely physical characteristics “supervene” upon those of its body.

But what, now, *is* it for the self to “have” a certain body as “its” body? Partly, it is just a matter of that self having certain physical characteristics which supervene upon those of *that* body rather than any other – although it is clear that this fact must be derivative from some more fundamental relationship. More than that, then, it must clearly also be a matter of the self's perceiving and acting “through” that body: and this indeed must be the crucial factor which determines *which* body's physical characteristics belong also to a given self. But what *is* it to perceive and act “through” a certain body rather than any other? As far as agency is concerned, this is a matter of certain parts of that body being directly subject to the agent's – that is, the self's – will: I can, of necessity, move certain parts of *my* body “at will” and cannot move “at will” any part of any body that is not part of mine.<sup>3</sup> Here it may be conceded that someone who is completely paralyzed may still possess a certain body, although only because he *could* once move parts of it “at will” and still perceives through it. But someone who was completely paralyzed from birth – if such a condition is even possible – could only in a more attenuated sense be said to “have” a body. So much for agency. As far as perception is concerned, apart from the obvious point that one perceives the world from the position at which one's body is located – except under abnormal circumstances, as when one looks through a periscope – it may be remarked that a person perceives her own body in a different manner from how he perceives others' bodies in that her sensations of it are phenomenologically localized in the parts perceived. For example, when one feels one's foot, one locates that feeling in the foot, whereas when one feels a wall, one does not locate that feeling in the wall.

Now it is true that in a less interesting sense all action and perception is “through” a certain body, namely, in the sense that as an empirically ascertainable matter of fact I need my limbs to move and my eyes to see. But *these* facts do not as such serve to qualify my limbs and eyes as especially *mine*, that is, as parts of *my* body. For, of course, I can be fitted with various prosthetic devices for locomotion and vision, and yet these do not *thereby* become parts of my body, although they *may* do so if they enter into the more intimate relationships discussed a moment ago. What makes my body peculiarly mine, then, is not determined merely by the empirically ascertainable dependencies that obtain between its proper functioning and my ability to engage in perception and agency. Thus, for example, even though it turns out that I need a brain in order to be able to think, it does not follow that this relationship suffices to make that brain peculiarly mine. In fact, I should say that a certain brain qualifies as mine only derivatively, in virtue of being the brain belonging to *my* body, where the latter qualifies as mine in virtue of having parts related to me in the more intimate ways mentioned earlier. As far as these more intimate relationships are concerned, however, my brain is as alien to me as a stone or a chair.

My thoughts, feelings, intentions, desires, and so forth all belong properly to *me*, not to my body, and are to be associated with my body only in virtue of those intimate relationships which make it peculiarly mine. It is impossible to associate such mental states with a body nonderivatively, that is, without relying upon their ascription to the self or person whose body it is – or so I would claim. No mere examination of brain function or physical movement can warrant such an association, without a detour through a recognition of the existence of a self or person to whom the body belongs. This recognition,

in interpersonal cases, will naturally have to issue from empirical evidence – but it will be evidence of *embodied selfhood* in the first instance, not directly and independently of particular mental goings-on.

### 11.3 The Self as a Simple Substance

But what now of my crucial claim that the self is simple, or lacks substantial parts? Well, what substantial parts *could* it have, given that the self is not to be identified with the body? Parts of the body cannot be parts of the self. If the self and the body had exactly the same parts, then they would apparently have to be identical substances after all. Certainly, standard mereological theory would imply this.<sup>4</sup> Similarly, if it were urged that all and only parts of the brain, say, are parts of the self, this would imply that self and brain are identical. So, I conclude that the self can have none of the body's parts as parts of itself, unless perhaps the self could have other substantial entities in addition to bodily parts as parts of itself.

However, no other substantial entity does appear to be a tenable candidate for being a substantial part of the self, whether or not in addition to bodily parts. For instance, the self patently does not consist of a plurality of lesser “selves” acting cooperatively, despite the picturesque “homuncular” descriptions of mental functioning advanced by some philosophers (see, e.g., Dennett 1979, 122–124). Such descriptions are not intelligible if taken literally. Similarly, we should not take literally talk of “corporate persons,” that is, the idea that institutions like clubs and firms are genuinely persons in their own right (see, e.g., Scruton 1989). At neither level – neither the subpersonal nor the suprapersonal – does the concept of a person find anything other than merely metaphorical application. Nor should we regard the mind's various “faculties” – will, intellect, and appetite, or modern variants of these, such as linguistic or visual information processing “modules” – as being “parts” of the self. For, in the first place, it is a mistake to reify such mental faculties or modules, and, in any case, they certainly could not qualify as *substantial* parts, which are what are now at issue. Mental faculties or modules, unlike substances, enjoy no possibility of an independent existence, and talk of them should be interpreted as referring to nothing more than certain abstractions from the overall psychology of a person. Thus, for instance, the notion of a will without an intellect, or of a language module in the absence of belief and desire, is just plain nonsense. Finally, it will not do to speak of the self's psychological states and processes themselves – its beliefs, intentions, experiences, and so forth – as being “parts,” much less as being substantial parts, of it. For this would at best be at all appropriate only on a Humean constructivist view of the self – the so-called bundle theory – which I reject entirely as incoherent. I conclude, therefore, that if the self is a substance, then it must indeed be a simple substance, entirely lacking substantial parts.<sup>5</sup>

The simplicity of the self goes some way toward explaining its unity, including the unity of consciousness that characterizes its normal condition. Where this unity threatens to break down – as in various clinical conditions such as those of so-called multiple personality, schizophrenia, brain bisection, and so on – we are indeed inclined to speak of a plurality of selves, or of divided selves. But I think, in fact, that such talk should again not be taken literally, and that the psychological unity that most fundamentally characterizes the self is not merely to be located at the level of consciousness. A divided consciousness is, I think, in principle consistent with self-identity: what is not consistent with this is a radical disunity of beliefs and values, manifested in a radical inconsistency of thought and

action. Of course, we all display mild inconsistencies, but no one person could intelligibly be interpreted as possessing the incompatibilities of belief and value that typically characterize two different persons. Now, a *complex* entity can act in disunified ways because the various incompatible or conflicting activities can be referred to different parts of that entity. Thus, a corporate entity such as a firm or a club can act inconsistently because its members may act in conflicting ways. But the actions of the self – those that are truly predicable of it, because they are genuinely intentional, and not merely of *the body*, such as so-called reflex actions, cannot in this way be ascribed to different elements or parts within the self. So we see that the simplicity and the unity of the self are indeed intimately related, even though there must clearly be much more to the matter than these brief remarks reveal.<sup>6</sup>

Another consequence of the simplicity of the self is this. If the self is a simple substance, then it appears that there can be no diachronic criterion of identity which grounds its persistence through time.<sup>7</sup> This is not to say that there may not be some *cause* of its persistence. It may well be, thus, that the continued normal functioning of the brain is a causally necessary condition of the persistence of the self, at least in the case of embodied, human persons. But it would not follow from this that the identity of the self over time is grounded in continuity of brain function, or indeed anything else. Nor should we think it contrary to the self's status as a substance that its existence may be thus causally dependent upon the functioning of another, distinct substance – the brain or, more generally, the body. No tenable account of substance can insist that a true substance must be causally independent of all other substances. For instance, a tree provides as clear an example of a substantial entity as anyone could wish for and yet, of course, a tree's continuing existence depends upon the maintenance of a delicate balance of forces in nature, both within it and between it and its environment. However, a tree is a *complex* substance, and accordingly its persistence can be understood as being grounded in the preservation of certain relationships between its substantial parts, despite the gradual replacement of those parts through natural processes of metabolism and growth. Not so with a self, any more than with, say, an electron or other "fundamental" particle. Thus, the reason why the self – or indeed any simple substance – cannot be provided with a criterion of diachronic identity is that such a criterion, in the case of a substance or "continuant," always makes reference to the substance's constituent parts, of which simple substances have none (see further Lowe 1998, chs. 5, 7).

That the diachronic identity of simple substances, including the self, is primitive or ungrounded should not be seen as making their persistence over time somehow mysterious or inscrutable. For, in the first place, as I have already remarked, it does not preclude us from recognizing the involvement of various causal factors in their persistence. Second, we can still concede – or indeed, better, insist – that there are certain necessary constraints on the possible history of a simple substance of any given kind: that is to say, limits on the sorts of changes that it can intelligibly be said to undergo, or limits arising from empirically discoverable natural laws governing substances of this kind. Thus, in the case of the self, a possible history must have a certain internal coherence to be intelligible, not least because perception and action are possible only within a temporal framework that includes both forward- and backward-looking mental states – intention and memory. Finally, the persistence of at least some simple substances is, I consider, presumed at the very heart of our understanding of time and change in general, so that we should not expect to be able to give an exhaustive or reductive account of all such persistence (Lowe 1998, chs. 5, 7). Indeed, since the only simple substances *directly* known to us, without benefit of scientific

speculation and experimentation, are precisely ourselves, I would urge that the pretheoretical intelligibility of time and change that is presupposed by all scientific theorizing actually rests upon our acquaintance with ourselves as simple persisting substances. So, although in the *ontological* order of nature it may well be the primitive persistence of fundamental physical particles which underpins objective time-order – in other words, which makes the world one world in time – still, in the *conceptual* order of thought it is the persistence of the self that underpins our very grasp of the notion of objective time order. If this is indeed so, then it would clearly be futile to expect the concept of the self to reveal upon analysis an account of the self's identity over time which did not implicitly presume the very thing in question.

A consequence of the ungroundedness of the self's identity over time is that there is, and can be, no definitive condition that necessarily determines the ceasing-to-be or, indeed, the coming-to-be of a self. In the case of complex substances, which are governed by clearly specifiable criteria of identity, the conditions for substantial change – that is, for their coming-or ceasing-to-be – can be stated fairly exactly, even though these conditions may in some cases be infected by some degree of vagueness. But not so with simple substances. And this is not, with them, a matter of vagueness at all – not, at least, in the sense in which “vagueness” implies the existence of “fuzzy” boundaries, whose “fuzziness” may be measured in degrees. This observation certainly seems to apply in the realm of fundamental particle physics, as far as I can judge. Thus if, in a particle interaction, an electron collides with an atomic nucleus and various fission products arise, including a number of electrons, it would seem that there may be no determinate “fact of the matter” as to whether the original electron is, or is not, identical with a given one of the electrons emerging from the impact event. There is here, it would seem, a genuine indeterminateness – I prefer not to say *vagueness* – of identity.<sup>8</sup> But this should not lead us to view with suspicion the idea that electrons do genuinely persist identically through time. Note, too, that known constraints on the possible history of an electron *may* enable us to rule out *some* re-identifications as impossible in a case such as that described – so that the indeterminacy is not totally unconstrained, which would be bizarre indeed. However, the point is that, even when all such constraints are taken into account, there may still be a residual indeterminacy in a given case.

Returning to the self, we see, thus, that while we may well think that we have good scientific grounds for believing that the functioning of the brain is *causally* necessary for the continued existence of the self, nonetheless, in the nature of the case, such evidence as we possess for this is bound to be inconclusive – and not just for the reason that all empirical evidence is defeasible – since we lack any reductive analysis of what would constitute the ceasing-to-be of a self. Lacking such an analysis, we cannot really say what empirical evidence would or would not support a claim that a self had definitely ceased to be. This is why the prospects for life after bodily death must inevitably remain imponderable and unamenable to decisive empirical determination.

Against this it may be urged that, since I have insisted that perception and agency are essential to selfhood, I must allow that the cessation of these *would* constitute a decisive terminus for the self's existence. However, it is the *capacity* for perception and agency that is essential, not its perpetual *exercise*. Very well, so can we not say that the demise of this capacity – and certainly its *permanent* demise – would constitute the demise of the self? But the trouble is that saying this is not really informative. For what would *constitute* the permanent demise of this capacity? Only, as far as I can see, the very demise of the self – in other words, no genuinely *noncircular* answer to the question can be provided. It will not do

to say that the permanent cessation of brain function would constitute the demise of the capacity for perception and agency. For the most that we can really say is that there seems to be an empirical correlation between mental activity and brain function, at least in the case of human persons. But the capacity for perception and agency does not by its very nature reside in any sort of cerebral condition. Indeed, there is nothing whatever unintelligible about supposing the existence of a capacity for perception and agency in a being entirely lacking a brain.

### 11.4 Physicalism, Naturalism, and the Self

Here it may be asked: is physiological psychology, or neuropsychology, simply a contradiction in terms, then/because psychology has, in essence, nothing to do with the brain as such? Not at all, so long as this branch of science is simply seen as telling us various empirical facts about the condition of embodied human persons or selves – that is, as telling us what sorts of processes, as a matter of fact, go on in their brains and nervous systems when they think or feel or act. This is not, however, and cannot be, an account of what constitutes thought or feeling or agency in a human person. Thought can no more be, or be constituted by, a brain process than a chair can be, or be constituted by, a set of prime numbers (see Geach 1979, 134). Nor should we be tempted into saying such things as that brain processes may “realize” episodes of thinking, as more cautious modern physicalists sometimes put it for what, really, is this supposed to mean?

In answer to this last question, it will perhaps be said that what it means to say that brain processes “realize” thought episodes is that thought episodes *supervene* upon brain processes, at least in the case of human persons. But saying this sheds no real illumination either, for the notion of supervenience – however useful it may be in some contexts – is entirely out of its depth here. Suppose we ask what it means to say that thought episodes *supervene* upon brain processes. We shall be told, perhaps, that what this means is that if *A* and *B* are two human persons who share type-identical brain states at any given time – that is, whose brain structures are atom-for-atom, neuron-for-neuron, indistinguishable at that time, with all of these neurons in identical states of excitation – then *A* and *B* must be enjoying type-identical thought episodes at that time. Perhaps it will be conceded that *A*’s and *B*’s thought episodes need not be identical in content – if Putnam and Burge’s verdicts regarding so-called twin-earth cases are accepted (see especially Burge 1979) – but it may nonetheless be insisted that their thought episodes must be subjectively indistinguishable, whatever that may be exactly taken to mean. However, the empirical status of this sort of claim – and, presumably, it cannot be advertised as being anything more than a merely empirical claim, since it can have no a priori justification – is highly problematic, as I shall now try to explain.

Let us, first of all, be clear that the thesis being advanced must be that thought episodes *supervene* globally or holistically – rather than just piecemeal – upon brain processes. For it is evident that, to the extent that thought is dependent on the brain, it can be so only in a holistic way which will not permit us to make any empirically confirmable claims about individual dependencies between particular or “token” thought episodes and particular or “token” brain events and processes.<sup>9</sup> So the thesis must be that a person with a brain exactly *replicating mine* at a level of neuronal organization and excitation will enjoy a mental life – feelings, beliefs, memories, and so on – indistinguishable from mine, but *not* that any partial

replication would necessarily engender any corresponding partial similarity in mental life. Nothing short of whole-brain replication will do. But what we now need to ask is this: what causal constraints would there be upon the process of bringing two distinct brains into such a state of exact neural replication? It is irrelevant to point out that one might, in some sense, be able to imagine this being done, perhaps instantaneously, by means of a machine that we rather question-beggingly call a "brain replicator." In this imaginary scenario, I walk in through one door of the machine, the operator throws the switch, and then I and my doppelgänger walk out through another door. One might as well say that the trick could be performed by magic. So too might pigs fly. But in fact it seems clear that there is simply no non-miraculous way in which this feat could be achieved. It would not even suffice, for instance, to take identical twins from the moment of conception and attempt to submit them to exactly similar environmental and social stimuli. For, first of all, the growth of nerve cells involves a good deal of randomness (Edelman 1989, 33–37), and second, it seems likely that brains, at the relevant level of organization, constitute a class of so-called chaotic systems (see, e.g., Crutchfield *et al.* 1986, 38–49; Goldberger, Rigney, and West 1990, 34–41). Thus, it could be that because the twins are subjected to minutely different influences for brief periods during their early development – as is effectively unavoidable – neural connections end up getting laid down in quite different ways in the two brains. The more that one reflects on the matter, I suggest, the more evident it should become that the whole idea of bringing two different human brains into identical neural states is so completely fanciful that it merits no place in serious philosophical inquiry.<sup>10</sup>

It will not do for the physicalist to protest here that all that he is interested in or committed to is the bare conceptual possibility of such whole-brain replication: for even if one can really make sense of this notion, what is one supposed to do with it? Precisely because the notion of such replication is the stuff of pure fantasy, utterly beyond the realm of scientific possibility, it cannot be conjoined with any genuine scientific findings from neuropsychology in order to yield a verdict on the truth or falsehood of the supervenience thesis. Nor can we justify such a verdict by consulting our "intuitions" regarding the upshot of the imagined replication experiment – for we are simply not entitled to any "intuitions" about the matter, and any that we do have we probably owe simply to our own prejudices. So my conclusion is that even if the supervenience thesis is coherently statable and even this may be in question – we can have no possible basis, either empirical or a priori, for judging it to be true.

Now, however, it may be objected that this rejection of physicalism even in the comparatively weak form of the supervenience thesis is unacceptably at odds with a "naturalistic" view of human beings and their minds. The emergence of the human mind, it may be said, must be recognized as being a result of evolutionary processes working upon the genetic makeup of animal life-forms through wholly biochemical means. Hence, it may be concluded, a biological account of human mentality is inescapable if one has any pretense to being "scientific." There cannot – so it will be said – be anything more to thought than can be exhaustively explained in biochemical terms, for otherwise the emergence of mind seems to be an inexplicable freak or accident. But, again, this is an objection which just reflects a dogmatic prejudice. Indeed, it is thoroughly question-begging and circular. It is just assumed from the outset that any wholly adequate explanation of the *emergence* of mind must be purely biological in character, because it is already presupposed that mind or mentality is a wholly biological characteristic of biological entities – animal life-forms. But the whole burden of my position is precisely that the mind is *not* a biological phenomenon and that mentality is *not* a

property of the biological entities which constitute human bodies. That such entities should be apt to embody selves or persons can, indeed, be no accident – but why presume that the evolution of such bodies or organisms is to be explained in exclusively biochemical terms? It is the *environment* of organisms that determines the evolutionary pressures on them to adapt and change: but the “environment,” in the present instance, cannot necessarily be specified in wholly physical and biochemical terms. All that can be said is that the *proximate* causes of genetic mutation are biochemical, as are the *proximate* causal factors favoring selection. But these causal factors are themselves effects of other causes – and the chain of causation can easily take us beyond the biochemical sphere. After all, we know that minds can affect the evolution of organisms, for the intelligent activities of human beings have done so within historical time. So there is nothing miraculous or non-naturalistic in the idea that the evolution of mind and that of body are mutually interactive, just as, on my view, individual minds and bodies are themselves mutually interactive. Thus, my answer to the “evolutionary” objection is that, unless it is presumed, quite unwarrantably, that the mental must be biologically based in order to contribute to the environmental selective pressures on organisms, it cannot be held that a nonbiological view of the mental such as mine is in any way in conflict with evolutionary theory.

But we need not take a purely defensive stance on this issue. It is worth remarking that archaeological evidence points to the occurrence of a fundamental intellectual transition in the human race some 35,000 or so years ago, not apparently connected with any very radical biological or neurological development in the human organism.<sup>11</sup>

This was a rather sudden transition from a markedly primitive sociocultural condition – which had endured virtually unaltered for many millennia and in which human creativity was limited to the production of the most rudimentary and severely practical tools – to a condition recognizably akin to our own, with the flourishing of visual and plastic arts reflective of a sophisticated aesthetic sensibility. The development of this condition, we may reasonably suppose, went hand in hand with that of true language, systems of religious thought, and the beginnings of political structures. At the root of these developments, it seems, was the emergence of genuine systems of representation, without which the sophisticated level of thought, communication, and social structure essential for personal existence as we know it would be impossible. Now, as I say, it seems likely that these developments were not the upshot of any radical change in human brain structure or neural processing capacity, but arose rather through concomitant changes in patterns of social interaction and organization.<sup>12</sup> And, indeed, we can observe essentially the same phenomenon in microcosm today in the education and socialization of human infants – who, unless they are subjected to appropriate social, cultural, and linguistic stimuli at an early age, are doomed never to develop a truly human personality and character. The implication of all this, I suggest, is that selves or persons are not, in essence, created through *biological* processes but rather by means of sociocultural forces, that is, through the cooperative efforts of other selves or persons. Quite literally, *persons* create other persons.

The picture that I am sketching of self-creation and the evolution of human personality is, I believe, not at all fanciful or “unscientific.” On the contrary, what seems utterly fanciful and facile is the biological reductionism that we see so forcefully promoted by many philosophers today.<sup>13</sup> When we reflect on how much we depend for our human condition upon the artificial and social environment that we ourselves have created, it seems quite incredible to suppose that one could hope to explain the human condition as having a basis solely in the organization of the human brain. Indeed, where human brain development and

structure do differ significantly from those of the higher primates, such as chimpanzees – for instance, in connection with our respective linguistic capacities – it seems proper to regard the difference as being at least as much a product as a cause of the different lifestyles of human beings and primates. For, of course, the neural structures in these distinctive parts of the human brain develop in human infants only in response to the right sorts of educative and social influences. It is true that a chimpanzee cannot, by being treated from birth like a human child, be made to develop in the way that the latter does, and this seems to indicate some innate biological difference between them. But we cannot assume that what we possess and the chimpanzees lack is some innate propensity specifically to develop human personality, language use, aesthetic appreciation, mathematical abilities, and so forth. For it may be that what prevents the chimpanzees from benefiting by our human processes of socialization and personality-creation is not an innate incapacity to acquire the abilities which these processes confer upon us, but rather just an incapacity to engage appropriately with these particular processes, geared as they are to specifically human needs and characteristics. After all, a human being could probably never learn to swim if it had to take lessons from dolphins! But this doesn't show, of course, that it is impossible for human beings to acquire a capacity to swim only that the acquisition process must be one that is geared to distinctively human limitations. Similarly, then, it is not altogether inconceivable that chimpanzees could be successfully subjected to processes of personality-creation analogous to our own, if processes appropriately tailored to their particular limitations could be discovered and exploited for that purpose.<sup>14</sup> In partial confirmation of this, it is worth noting that, whatever one makes of the various attempts to teach chimpanzees the genuine use of language, it is clear that those attempts began to look successful only when they took into account the fact that chimpanzees have severely restricted capacities for vocalization, and substituted sign language for speech (see, e.g., Linden 1976).

Perhaps the following analogy will help to convey the general sense of my proposal. A potter takes a lump of clay – which has, as such, no special propensity to be formed into any particular type of artifact, such as a statue or a vase, even though it is *suitable* material for such a purpose, in a way that a bunch of feathers, for example, would not be – and he forms it, let us suppose, into a vase. In creating the vase, he has created a new substantial individual which is distinct from, although at the same time embodied in, the lump of clay. In a somewhat similar manner, I suggest, human persons acting cooperatively take the biological “clay” of their children and “shape” it into new persons. And this “clay” – although, of course, it has to be *suitable* to the “shaping” processes applied to it – need not be thought of as having any special propensity to receive just such a “shape.” Finally to complete the analogy – the human person emerging from this “shaping” process is a new substantial individual which is distinct from, although embodied in, the biological entity that is the “clay.” It is no accident, surely, that it is precisely this metaphor for the creation of persons that we find so often in religious and mythic literature.

Notice, furthermore, one other aspect of the analogy that is particularly apt: what constitutes “suitable” material for formation into an artifact of any given type is not purely a function of the inherent properties of that material together with the nature of the type of artifact in question, but also a function of the sorts of creative processes that the artificer is equipped to apply to the material. Clay is a suitable material to make into vases as far as *human* artificers are concerned, but only because human beings have hands with which they can shape the clay. However, it should also be remarked that many processes of artifact creation can be facilitated through – and, indeed, are sometimes made possible only by – the



use of previously created artifacts, such as, for example, the potter's wheel. In an analogous manner, then, what makes *human* biological material "suitable" for the creation of persons is not just a function of the inherent biological characteristics of that material together with the nature of the psychological capacities which need to be conferred, but also a function of the creative processes available to us given our own particular limitations – although, indeed, some of these limitations may be progressively transcended through the exploitation of previous products of our own creativity, that is, through the exploitation of our growing sociocultural, linguistic, and technological heritage.

I should perhaps stress, in conclusion, that what I have just been developing is only an analogy: I do not want to suggest that persons literally are artifacts, other than in the very liberal sense that they are products of personal creativity. Above all, unlike material artifacts, persons or selves are simple substances: parts of their bodies are not parts of them, as bits of clay are parts of a vase. Moreover, whereas it is plausible to hold that all of a vase's intrinsic properties supervene upon certain properties of its constituent clay, it is not, as we have seen, reasonable to regard the self's psychological properties as supervening upon any properties of its body, such as neurophysiological properties of its brain. As Joseph Butler, the famous Bishop of Durham, might have said, the self is what it is, and not another thing.

## Acknowledgment

This chapter first appeared in *After Physicalism*, edited by Benedikt Paul Gocke, 48–71 (Notre Dame: University of Notre Dame Press, 2012), and is here reprinted by permission of the editor and University of Notre Dame Press.

## Notes

1. For more on the ontology of substance and mode, see Lowe (2006).
2. For criticism of this suggestion, see Lowe (1989a, 119–120). The view in question is, notably, advanced by Baker (2000).
3. In another terminology, we may say that movements of certain parts of its own body can necessarily be executed as "basic" actions by the self. The *locus classicus* for the notion of a "basic" action is Danto (1965).
4. See, for example, Goodman (1977, 33–40). Standard mereological theory is possibly wrong on this score, if it is correct, as I myself believe, to differentiate between a tree, for example, and the mass of wood which temporarily composes it – for these may seem to have the same parts, at least during the period in which the one composes the other. However, while the tree and the wood arguably have the same *spatial* parts, it is much more debatable whether they have the same *substantial* parts. For instance, a certain root will be a substantial part of the *tree*, but hardly of the wood composing the tree. By contrast, a substantial part of the wood composing the tree arguably *is* also a substantial part of the tree. The issue is a complex one, which I cannot go into in further depth here. But, in any case, I think it independently reasonable to deny that substantial parts of the body are literally parts of the self – and I do not think of the body as in any sense *composing* the self.
5. For a much fuller exposition and defense of this view, see Lowe (2001).
6. I say much more about such matters in Lowe (2005a).
7. For more general discussion of persistence and criteria of identity, see Lowe (1989b; 1998, ch. 5).
8. A sizable literature related to this issue has grown out of Evans (1978), although this is no place for me to attempt to engage with it. I discuss the electron case more fully and challenge Evans's argument against indeterminate identity in Lowe (1994; see also 1998, 63–69; 2005b).
9. This appears to be an inescapable implication of Donald Davidson's well-known thesis of the "holism of the mental," for which see Davidson (1980, 217). I do not, however, accept Davidson's own view of the relations

- between mental and physical events, which is a “token-token” identity theory. See further Lowe (1989a: 113–114, 132–133).
10. It has also been pointed out that if quantum states of the brain have to be taken into account (as they will be if mental states are at all dependent on them), then exact duplication at the relevant level of organization will be ruled out by quantum mechanical principles (Penrose 1989: 270).
  11. See White (1989, 1982). See also the essays by White and others in Mellars and Stringer (1989), especially section 2.
  12. This would be consistent with much of the recent work of psychologists, anthropologists, and ethologists presented in Byrne and Whiten (1988).
  13. My opposition extends even to the most sophisticated modern proponents of the biological approach, such as Ruth G. Millikan (see Millikan 1984). However, a detailed critique must await another occasion.
  14. I should remark, incidentally, that I by no means wish to deny mentality to chimpanzees and other higher primates, although I very much doubt whether any such animal may be said to possess or embody a “self,” as I would define that term – for, as I understand it, a “self” is a being capable of rational thought and conscious self-reflection. Thus, inasmuch as mental states necessarily attach to psychological subjects which are not to be identified with their biological bodies, I am committed to the view that persons or selves are not the only species of psychological substance, and that – in an older terminology – there are “animal souls” which find a place “below” ourselves in a hierarchy of psychological substances. I hope to discuss this issue more fully elsewhere.

## References

- Baker, Lynne Rudder. 2000. *Persons and Bodies: A Constitution View*. Cambridge: Cambridge University Press.
- Burge, Tyler. 1979. “Individualism and the Mental.” *Midwest Studies in Philosophy*, 4: 73–121.
- Byrne, Richard, and Andrew Whiten, eds. 1988. *Machiavellian Intelligence: Social Expertise and the Evolution of Intellect in Monkeys, Apes, and Humans*. Oxford: Clarendon Press.
- Crutchfield, James P., J. Doyne Farmer, Norman H. Packard, and Robert S. Shaw. 1986. “Chaos.” *Scientific American*, 255: 38–49.
- Danto, Arthur C. 1965. “Basic Actions.” *American Philosophical Quarterly*, 2: 141–148.
- Davidson, Donald. 1980. *Essays on Actions and Events*. Oxford: Clarendon Press.
- Dennett, Daniel C. 1979. *Brainstorms: Philosophical Essays on Mind and Psychology*. Hassocks, UK: Harvester Press.
- Descartes, René. 1984. “Principles of Philosophy.” In *The Philosophical Writings of Descartes*, translated by John Cottingham, Robert Stoothoff, and Dugald Murdoch, 177–290. Cambridge: Cambridge University Press.
- Edelman, Gerald M. 1989. *Neural Darwinism: The Theory of Neuronal Group Selection*. Oxford: Oxford University Press.
- Evans, Gareth. 1978. “Can There Be Vague Objects?” *Analysis*, 38: 208.
- Geach, P. T. 1979. *Truth, Love and Immortality: An Introduction to McTaggart’s Philosophy*. London: Hutchinson.
- Goldberger, Ary L., D. R. Rigney, and B. J. West. 1990. “Chaos and Fractals in Human Physiology.” *Scientific American*, 262: 42–49.
- Goodman, Nelson. 1977. *The Structure of Appearance*, 3rd edn. Dordrecht, Netherlands: Reidel.
- Linden, Eugene. 1976. *Apes, Men and Language*. Harmondsworth, UK: Penguin Books.
- Lowe, E. J. 1989a. *Kinds of Being: A Study of Individuation, Identity and the Logic of Sortal Terms*. Oxford: Blackwell.
- Lowe, E. J. 1989b. “What Is a Criterion of Identity?” *Philosophical Quarterly*, 39: 1–21.
- Lowe, E. J. 1992. “The Problem of Psychophysical Causation.” *Australasian Journal of Philosophy*, 70: 263–276. Reprinted, 2003, in *Philosophy of Mind: Contemporary Readings*, edited by Timothy O’Connor and David Robb, New York: Routledge.
- Lowe, E. J. 1994. “Vague Identity and Quantum Indeterminacy.” *Analysis*, 54: 110–114.

- Lowe, E. J. 1998. *The Possibility of Metaphysics: Substance, Identity, and Time*. Oxford: Clarendon Press.
- Lowe, E. J. 2001. "Identity, Composition, and the Simplicity of the Self." In *Soul, Body, and Survival: Essays on the Metaphysics of Human Persons*, edited by Kevin J. Corcoran, 139–158. Ithaca, NY: Cornell University Press.
- Lowe, E. J. 2005a. "Can the Self Disintegrate? Personal Identity, Psychopathology, and Disunities of Consciousness." In *Dementia: Mind, Meaning, and the Person*, edited by J. Hughes, S. Louw, and S. Sabat, 89–103. Oxford: Oxford University Press.
- Lowe, E. J. 2005b. "Identity, Vagueness, and Modality." In *Thought, Reference, and Experience: Themes from the Philosophy of Gareth Evans*, edited by J. L. Bermúdez, 290–310. Oxford: Clarendon Press.
- Lowe, E. J. 2006. *The Four-Category Ontology: A Metaphysical Foundation for Natural Science*. Oxford: Clarendon Press.
- Mellars, Paul, and Chris Stringer, eds. 1989. *The Human Revolution: Behavioural and Biological Perspectives on the Origins of Modern Humans*. Edinburgh: Edinburgh University Press.
- Millikan, Ruth G. 1984. *Language, Thought, and Other Biological Categories*. Cambridge, MA: MIT Press.
- Penrose, Roger. 1989. *The Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics*. Oxford: Oxford University Press.
- Scruton, Roger. 1989. "Corporate Persons." *Proceedings of the Aristotelian Society*, 63(Suppl.): 239–266.
- White, Randall. 1982. "Rethinking the Middle/Upper Paleolithic Transition." *Current Anthropology*, 23: 169–192.
- White, Randall. 1989. "Visual Thinking in the Ice Age." *Scientific American*, 26(1): 74–81.

# Debating the Unity of Consciousness

# Substance Dualism and the Unity of Consciousness

J. P. MORELAND

The appearance of consciousness in the world is an amazing and puzzling fact in its own right. Indeed, consciousness is one of the most mystifying features of the cosmos. Colin McGinn claims that its arrival borders on sheer magic because there seems to be no naturalistic explanation for it: “How can mere matter originate consciousness? How did evolution convert the water of biological tissue into the wine of consciousness? Consciousness seems like a radical novelty in the universe, not prefigured by the aftereffects of the Big Bang; so how did it contrive to spring into being from what preceded it?” (McGinn 1999, 13–14).

However, it is not simply the existence of consciousness that is a mystery. The unity of consciousness is something that cries out for analysis and explanation as well. At any given time, we have a number of distinct experiences – feeling a pain, hearing a bird sing, seeing a chair, seeing a table. Yet these seem to be tied together and unified in some sort of deep way. This fact invites at least four questions: (1) What is the precise nature of the unity of conscious and what sort of metaphysical analysis can we give of it? (2) Why is the unity of consciousness a problem? (3) What, if anything, grounds the unity of consciousness? (4) What kind of universe must we live in for there to be entities such as states of unified consciousness that ground the unity of consciousness?

These questions focus on interesting aspects of consciousness itself. But as Thomas Nagel has recently reminded us, topics in the philosophy of mind are not local ones; they invade our understanding of the entire cosmos and its history. If irreducible consciousness and reason exist, we simply must ask what sort of reality could and did give rise to them. Thus, one must keep an eye on worldview implications of a position on a topic in philosophy of mind (Nagel 2012, ch. 1). Nagel seems correct, and question (4) above seeks to enfold this issue within the broader project of clarifying the nature and grounds of consciousness’ unity.

With this in mind, in what follows, I will, first, clarify the nature of the unity of consciousness; second, identify why it is a problem; third, offer and critique what I take to be

the best – though, ultimately, inadequate – solution as to what the unity of consciousness amounts to and what grounds that unity; I also present my own answers to these questions and show their superiority to the solution I reject; fourth, briefly gesture toward the broader worldview implications of my solution. Let us begin in earnest, then, and see what we can learn about the unity of consciousness.

## 12.1 What is the Unity of Consciousness?

Perhaps more than anyone else in the literature, Tim Bayne and David Chalmers have done very careful thinking about the unity of consciousness, although in my view, Bayne is the chief current thinker on these matters. They distinguish several different types of unified consciousness, but only three of them are relevant to my present concerns (Bayne and Chalmers 2003, 23–58). To understand these, I should point out that a phenomenal conscious state  $\phi$  is one such that there is a what-it-is-like to be in  $\phi$ .

First, there is *objectual phenomenal unity*: Two or more states are so unified if they are experienced as being of the same object. For example, the state of feeling a desk's shape and the state of seeing the desk's color, are objectually phenomenally unified just in case they are experienced as being of the same object, the desk. This type of unity generates the binding problem or what Tim Bayne calls the "feature" binding problem (Bayne 2010, 229). How is it that we experience, say perceptually, separate pieces of information as bound together in pertaining to the same object?

Second, there is *subject phenomenal unity*: This occurs when all of one's phenomenal states are had by the same subject. According to Bayne and Chalmers, this sort of unity is irrelevant for investigating the nature of consciousness' unity because it is trivially true by definition and tells us nothing about consciousness. Unfortunately, this claim is all too convenient for Bayne because without it, his own solution, as we shall see below, is not intellectually motivated. Moreover, it is just plain false. Many thinkers, including Bayne himself, believe there is no self or subject that unifies consciousness. And some thinkers like me think that a simple self or subject is the best explanation for the unity of consciousness, especially when a careful metaphysical account is given for how conscious states are in the self in the first place. Thus, it is far from trivial to employ the notion of "had by the same subject." More on this later.

Third, there is *subsumptive phenomenal unity*: Two (or more) states are subsumptively phenomenally unified just in case there is something-it-is-like to be in both states simultaneously and conjointly. All of one's phenomenal states are subsumed within a single (totalizing) phenomenal state. Bayne and Chalmers insist, rightly in my view, that one's total phenomenal field is not built up atomistically as a complex conjunction of individual phenomenal states. Rather, one's totalizing state is a whole and the various individual phenomenal states are "aspects" of that ontologically prior whole.

For example, suppose you are having two different phenomenal experiences – an awareness of  $\phi$  and an awareness of  $\psi$ . Then there will be a totalizing state  $T$  in its own right with its own what-it-is-like to be in  $T$ .  $T$  is "over and above" and not merely an atomistic conjunction of the two awarenesses of  $\phi$  and  $\psi$ , taken individually. According to Bayne and Chalmers, any subject of experience will have a single, total phenomenal state at a point in time (Bayne and Chalmers 2003, 32–33). Coming from a different direction, Cleeremans makes roughly the same point: "[There is] the intuitive idea that consciousness requires

unity of consciousness, that is, that there is no sense in which one could simultaneously have separate conscious experiences that failed to present themselves as integrated in a single phenomenal field" (Cleeremans 2003, 2).

Bayne and Chalmers go on to formulate what they call the total phenomenal unity thesis: Necessarily, the set of all phenomenal states of a subject at a time is phenomenally unified (Bayne and Chalmers 2003, 33). According to them, this thesis captures the central insight behind unity of consciousness: there is always a single phenomenal state that subsumes all of one's phenomenal states at a time, and this state is one's totalizing phenomenal state.

At this point, it is worth briefly making two observations. First, the necessity of the total phenomenal unity thesis should be explained if possible and not taken as a primitive. Second, the notion of subsumption where one state subsumes another is unclear to say the least. Bayne and Chalmers claim that it is "something of an intuitive primitive" (Bayne and Chalmers 2003, 40).<sup>1</sup> In my view, this is an unfortunate situation, and if we can ground the totalizing unity of consciousness in a clearer, metaphysically available entity, then we should seek to do that. Elsewhere, Bayne does attempt to provide more clarity to subsumption by calling it a part/whole mereological relation such that one experience subsumes another if and only if the latter is contained as a part of the former (Bayne 2010, 20). He goes on to say that his mereological account of subsumption provides a mereological account of phenomenal unity: experiences are phenomenally unified when parts of a single whole, a single experience (Bayne 2010, 45). Later on, I will try to show that, given Bayne's ontological commitments, this account of subsumption and phenomenal unity is obscure at the very least, and more likely, it is incoherent.

In sum, these three sorts of unity are all relevant to what follows. Indeed, I will try to show that proper analysis of one of these theses provides the solution to the other two. In any case, when speaking of the unity of consciousness, I will use the specific thesis that I have in mind.

## 12.2 Why are these Unity-of-Consciousness Theses a Problem?

The first reason that the unity of consciousness is a problem is the simple fact that it involves an interesting set of philosophical problems in its own right. As Cleeremans notes, "First and foremost is the fact that the unified character of conscious experience is in and of itself something that deserves explanation" (Cleeremans 2003, 18).

But there is a deeper and more pervasive problem lurking in the neighborhood. Simply put, once an adequate ontological analysis is given of the nature of and ground for the unity of consciousness, it is hard to locate that unity and ground within the constraints of the most widely accepted contemporary version of naturalism. To see this, it is important to get before us an important ontological distinction between separable and inseparable parts:

*p* is a *separable part* of some whole *W* =<sub>def.</sub> *p* is a particular, *p* is a part of *W* and *p* can exist if it is not a part of *W*.

*p* is an *inseparable part* of some whole *W* =<sub>def.</sub> *p* is a particular, *p* is a part of *W* and *p* cannot exist if it is not a part of *W*.

Inseparable parts get their existence and identity from the whole of which they are parts. The paradigm case of an inseparable part in this tradition is a (monadic) property-instance or relation-instance. Thus, if substance *s* has property *P*, the-having-of-*P*-by-*s* is (1) a

property-instance of P; (2) an inseparable part of *s* which we may also call a mode of *s*. For example, let *s* be a chunk of clay, *P* be the property of being round, and the-having-of-*P*-by-*s* be the clay's being round. The clay could exist without being round, and the property of being round could exist without there being clay (e.g., a baseball could have that property), but the clay's being round could not exist without the clay. The clay's being round is a mode or inseparable part of the clay.

For present purposes, it is important to say a bit more about criteria for naturalist ontological commitments. A good place to start is with what Frank Jackson calls the location problem (Jackson 1998, 1–5). According to Jackson, on the basis of the superiority of scientific ways of knowing exemplified by the hard sciences, naturalists are committed to a fairly widely accepted physical story about how things came to be (the Grand Story) and what they are. Given these commitments, the location problem is the task of locating or finding a place for some entity (for example, semantic contents, mind, agency) in that story.

For Jackson, the naturalist must either locate a problematic entity in the basic story or eliminate the entity. Roughly, an entity is located in the basic story just in case it is entailed by that story. Otherwise, the entity must be eliminated. Jackson correctly notes that a naturalist ought to adopt serious metaphysics: his or her ontology should start with the smallest number of different sorts of entities – those we find in our best theories in physics – and incorporate those entities that can be explained by the Grand Story's combinatorial processes, namely, mereological aggregates for individuals and structural properties for attributes. At this point, it is worth recalling that Kim and others have complained that one does not *explain* a phenomenon by labeling it supervenient.

It is hard to see how the ontology generated by the combinatorial processes at the heart of scientific explanation, for example, as seen in the atomic theory of matter and evolutionary biology, could countenance simple emergent properties or substances. In my view, “emergence” is just a name for the problem to be solved (how could simple emergent properties and substances emerge if you start with particles as depicted by physics and just rearrange them over time?). Among other things, this means that without some pretty serious, wildly ad hoc adjustments, the sort of unity possessed by consciousness (and, perhaps, its ground) cannot be located or otherwise explained, given strict naturalism.

The naturalist ontology is widely conceived as properly captured in the mereological hierarchy.<sup>2</sup> According to the hierarchy, in the category of individuals, all wholes above the level of atomic simples are mereological aggregates – aggregates of separable parts that stand in external relations to each other with a structure that is itself an aggregate of numerous relation instances – for example, causal, spatiotemporal contiguity – between and among various subgroups of parts. In the category of property, structurally supervenient properties – those that are merely a new relational combination of the parts and properties at the subvenient level – such as being H<sub>2</sub>O, fit naturally into the naturalist combinatorial depiction of how all things have developed since the Big Bang. The appearance of emergent properties – unique, *sui generis*, new, simple properties that do not characterize the subvenient base (e.g., painfulness) – appears to be a case of magic without a magician.

Now, as Tim Bayne has admitted, the unity of consciousness is not atomistic. The phenomenal field is not a structural entity constituted by separable parts standing in external relations to each other (Bayne 2010, 225–229). Among other things, this is why



panpsychism is inadequate here. For the sake of argument, we can grant it solves the problem of getting consciousness from nothing. But, as perhaps the leading defender of panpsychism has acknowledged, since panpsychism employs combinatorial processes every bit as much as strict physicalism, it is no more adequate to explain the sort of unity consciousness possesses (Skrbina 2005, 264–265). In rejecting atomism, Bayne opts for a holistic view of consciousness unity. I believe he is right about this, though as we will see below, he fails adequately to offer a ground for his holism. In a sense to be clarified below, holism seems to involve depicting the different aspects of synchronic consciousness as involving internal relations and inseparable parts or modes.

I will explain below in what sense this is the case. But for now, this is the first and fundamental problem with the unity of consciousness: it is almost impossible to locate in a standard naturalist worldview. And punting to emergence is simply to slap a label on the problem. Both the existence and unity of consciousness will have to go. As J. J. C. Smart, paradigmatically noted long ago:

It seems to me that science is increasingly giving us a viewpoint whereby organisms are able to be seen as physicochemical mechanisms . . . There does seem to be, so far as science is concerned, nothing in the world but increasingly complex arrangements of physical constituents. All except for one place: in consciousness . . . I just cannot believe that this can be so. That everything should be explicable in terms of physics . . . except the occurrence of sensations seems to me to be frankly unbelievable. (Smart 1959, 61)

Besides the location problem, there is another difficulty the unity of consciousness presents for naturalism. Given naturalism, the various entities to which a human person is identical – the organism, the brain, an object constituted by an organism, a four-dimensional physical object synchronically composed of its physical parts and diachronically composed of its stages – do not have the sort of unity needed to account for or ground the unity of consciousness because they are all mereological aggregates. Rather, the best explanation/ground for the unity of consciousness is a simple (not composed of separable parts) soul.

More than anyone else, William Hasker has championed this argument for substance dualism (Hasker 1999, 122–144; see also Hasker 2005). To remind us of what we have noted earlier, by the unity of consciousness, say, of one’s visual field, I mean (at least) two things. First, there is what Bayne and Chalmers call subsumptive phenomenal unity: all of one’s experiences are subsumed within a single, totalizing state of consciousness. This totalizing state is a conscious state in its own right, and there is a what-it-is-like to be in that state (Bayne and Chalmers 2003, 26–27). The total phenomenal unity thesis says that, necessarily, there is always a single phenomenal state that subsumes all of one’s other phenomenal states at a time.

Since the notion of “subsuming” is a bit unclear, let me state the second thing I mean by the unity of consciousness, also from Bayne, and add my own metaphysical clarification to it. According to Bayne, an atomistic theory of consciousness states that the phenomenal field is composed of “atoms of consciousness” – independent conscious states (Bayne 2010, 225–229). Among other things, this would mean that one’s field of consciousness is like a mereological aggregate – it contains and is built up by separable parts (“atoms of consciousness”) – placed into various external relation instances which constituted the field’s structure – a mereological aggregate of relation instances.<sup>3</sup>

By contrast, Bayne – and I – accept holism: The “components” of the phenomenal field are conscious only as components of that field. (It is interesting to note that diachronically, consciousness changes as a continuous flow, but the brain changes states in a discrete, atomistic way.) I add the qualification that the phenomenal field is a whole in which subsumptive components are modes or inseparable parts of something – either the whole field, or much more likely, of the grounding entity for subsumptive unity.

Now consider the following principle:

- (F) For any complex object (one with a plurality of separable parts)  $O$ , if  $O$  performs function  $F$ , then  $O$ 's performing function  $F$  consists in parts  $p_1$ – $p_n$  and sub-functions/activities  $f_1$ – $f_n$ , such that  $p_1$  performs  $f_1$  . . .  $p_n$  performs  $f_n$ .

For example, a computer performing function  $F$  just is a certain set of its parts performing their own subfunctions. Principle  $F$  can also be stated in terms of properties such that an object  $O$  having some property  $P$  consists in each part having some property or other. This is clearly the case with additive properties, for example, mass. It does not, however, rule out emergent properties. Given the reasonable assumption that supervenience for simple, emergent properties is local (the supervenient simple property obtains and is dependent on what is going on right there at the subvenient base), the principle disallows emergent properties exemplified by complex objects like  $O$  taken as an irreducible whole. But it does not disallow each of the relevant parts of  $O$  to have an emergent property as long as these parts are mereological simples.

The following argument, then, is an attempt to show that the unity of consciousness cannot be explained if one is a brain (or any of the other naturalist candidates mentioned earlier), because a brain is just an aggregate of different physical (separable) parts. It is only if the self is a single, simple subject that we have an adequate account for the unity of consciousness.

To illustrate, consider one's awareness of a complex fact, say one's own visual field consisting of awareness of several objects at once, including a number of different surface areas of each object. One's entire visual field contains several different experiences, for example, being aware of a desk toward one's left side and being aware of a podium in the center of one's visual experience of an entire classroom. Corresponding to such an experience, numerous different light waves bounce off of different objects (and off of different locations on the surface of the same object, say different areas of the desk's top side), they all interact with the subject's retinas, and they all spark signals that terminate in a myriad of locations in the brain, breaking objects down into constituents (LaRock 2013). If we add local emergence, then we could hold that each relevant part of the brain instantiates an atomistic sensory experience.

Accordingly, a physicalist may claim that such a unified awareness of the entire room by means of one's visual field consists in a number of different physical parts of the brain each terminating a different wavelength, each of which is aware only of part – not the whole – of the complex fact (the entire room). But this cannot account for the single, unitary awareness of the entire visual field. There is a what-it-is-like to have the whole visual field. If we terminate our search for an explanation for this with a holistic phenomenal field, then two problems arise. First, it is hard to see how a myriad of atomistic parts could give rise to a single, nonatomistic, holistic field; we are owed an account of this within the constraints of subject physicalism.

Second, a basic datum of our experience is not simply this or that item of awareness of the room, but that *I have and am not identical to the totalizing state*. In the history of philosophy, classic substances have served to unify things in this way, and Hasker and Eric LaRock believe this ontology provides the best answer for how we could have a totalizing, unified field of consciousness. The very same substantial soul is aware of the desk to the left, the podium at the center, and, indeed, each and every distinguishable aspect of the room. But no single part of the brain is correspondingly activated as a terminus for the entire visual field. Only a single, uncomposed mental substance can adequately account for the unity of one's visual field or, indeed, the unity of consciousness in general.

The most widely advanced physicalist rejoinder attempts to explain objectual phenomenal unity in terms of synchronicity. All the different locations of the brain processing electrical signals associated with different aspects of the object of perception (e.g., color, size, shape, etc.) fire together at the very same time, and this explains objectual unity. Unfortunately, a growing amount of empirical evidence refutes this thesis (LaRock 2015). And, philosophically, the connection between synchronicity and objectual unity is unclear. Consider LaRock's analogy: "If five chefs are located in separate kitchens and each chef is consciously aware of only part of the same recipe, it does not follow that any one chef is consciously aware of the recipe as a whole – *even if all of the chefs are consciously aware of their respective recipe parts at the same time*" (LaRock 2015, 15). The synchronicity solution, then, fails to be adequate.

Though it may not be as prominent as the synchronicity rejoinder, as far as I can tell, by far, the most sophisticated critique of Hasker's unity-of-consciousness (UOC) argument has been advanced by Warren Shrader (2006). In my view, Shrader raised three main arguments against Hasker's UOC. Accordingly, in what follows I shall state each argument followed by my reply.<sup>4</sup>

Argument #1: Hasker's UOC is supported by an unwarranted principle of reducibility (PR). As Shrader correctly points out, UOC begins with this premise:

- 1  $S_Q$  exists (i.e.,  $Q$  is exemplified by  $m$ ).

Here  $m$  is the mind,  $Q$  is an irreducibly qualitative property (e.g., being a unified visual field) and  $S_Q$  is the state of affairs of the mind's exemplifying  $Q$ . So far so good. Shrader goes on to assert that (1) is supported by PR but PR is without adequate justification:

(PR) If, strictly speaking, an object is a system of other objects, then every property of the object must consist in the fact that its constituents have such and such properties and stand in thus and so relations to each other, that is, every property of a system of objects consists of properties of, and relation between and among its constituents.

*Reply:* This is not an accurate representation of the role that PR plays in UOC. PR is not a stand-alone principle. Rather, PR is itself warranted by very fundamental, long embraced and, currently, widely shared intuitions that are very difficult to reject. And while I don't need such a strong claim for my reply to go through, it may well be that some of these crucial intuitions are properly basic. In any case, one of those intuitions is that at any moment, my state of consciousness is not built up piecemeal from little atomistic units of consciousness to form an aggregate. No, my state of consciousness is holistic and totalizing, that is, it somehow subsumes and incorporates all the current aspects of consciousness into an ontologically prior whole.

Other fundamental intuitions relevant to PR become evident when we look at Shrader's formalization of PR in his statement of premise (2) of UOC:

If Q is exemplified by p (the brain), then either:

- a)  $S_Q$  must consist in  $S_{Q_1}$  &  $S_{Q_2}$  & . . . &  $S_{Q_n}$  or
- b)  $S_Q$  must consist in  $S_{T_1}$  and  $S_{T_2}$  & . . . &  $S_{T_n}$  &  $R(p_1 \dots p_n)$  (where  $S_{T_x}$  stands for the state of affairs of  $p_x$  – a specific part of the brain – exemplifying property  $T_x$  and R stands for the (structural) relation in which all the brain's parts stand relative to one another).

Disjunct (a) amounts to the claim that the mind's exemplifying the qualitative property Q consists in the sum of the individual states of affairs in which each part of the brain exemplifies a part of Q. In this sense, Q is either just the sum of all the little Qs or in some inexplicable way a whole composed of the little Qs.

It seems to me that Hasker is correct in rejecting both disjuncts of (2). Disjunct a seems to be warranted to a small minority of philosophers, and it has been treated as such for centuries and centuries. Indeed, disjunct a appears to express some form of panpsychism. But in what may be the most authoritative exposition of the history of panpsychism and defense of the position, David Skrbina (2005) admits that the single, most difficult argument against the view is the combination problem: How can one get a holistic, totalizing unified self (and consciousness) by simply adding together little drops of sentience?

More explicitly, prior to their "unification," each unit of consciousness belonging to its own part of the brain stands in external relations to all the other units of consciousness. But as I will show in some detail later, given the holistic nature of the unity of consciousness, the different aspects (i.e., inseparable parts) of consciousness stand in internal relations to each other. How, simply by bringing the parts of the brain together in closer spatial proximity, do the external relations all of a sudden become internal relations? This seems like nothing but magic, and to claim that this fact is simply a brute one is to engage in an ad hoc assertion of the worst kind.

Disjunct (b) is even more implausible than (a). With a, at least you start with little entities that have their own unit of (attenuated) consciousness. But with (b), one starts with brute matter as described in the language of physics, chemistry and neurobiology such that each brain part exemplifies physical properties only, and those parts are placed into a relational structure (more precisely, an individual instance of a type of relational structure). The problem here lies in the fact that there are two very different types of supervenient properties: structurally supervenient properties that consist in a simple structural rearrangement of what already existed at the subvenient level, and emergently supervenient properties that consist in the appearance of a *sui generis*, new, simple quality that is not present at the subvenient level. Unfortunately, disjunct b only gives us a structurally supervenient property, but what is needed is an emergently supervenient property – being a unified property of phenomenal sentience.<sup>5</sup>

It is these fundamental insights and intuitions that ground PR, and that is why, Shrader's criticism notwithstanding, Hasker grounds his UOC by citing specific examples and counterexamples that constitute the grounds for PR.

*Argument #2: Hasker's examples of UOC are consistent with a number of unity theses some of which are trivial.* Presumably, Shrader has in mind the phenomenal unity of consciousness thesis (PUT) (for any set of phenomenally conscious states of a subject at a given time, all the members of that set are phenomenally unified) and what he calls the trivial subject unity thesis (SUT) (two conscious states are unified if they belong to the same subject at the same time). According to Shrader, Hasker is incorrect in claiming that SUT is

the only type of unity that plays a role in his argument because there is no premise of Hasker's argument that depends on SUT.

*Reply:* Shrader's two criticisms – that SUT is trivial and that it plays no role in Hasker's argument – fail to be convincing. The SUT thesis is far from trivial and, again, I must ask the reader's patience because I develop and defend this point in some detail later. But for now, let me say simply that perhaps most philosophers of mind do not believe there is a self or a subject of consciousness. So the claim that the unity of consciousness occurs (and, indeed, may well be grounded) in a substantial subject is far from trivial. Regarding the second claim, I argue later that PUT is grounded in and best explained by SUT. If this is not the case, then PUT expresses an incredibly bizarre brute fact that is utterly unique in the universe and cries out for a deeper explanation and grounding. And while I may be wrong, my reading of Hasker leads me to believe that SUT is actually a fundamental datum for him that either supports or is expressed in premise (1) of his argument ( $S_Q$  exists, i.e.,  $Q$  is exemplified by  $m$ ), a premise that Shrader himself accepts.

*Argument #3:* Hasker gives the materialist no good reason to accept (2) and, in fact, there are alternatives more conducive to materialism. Shrader asks the question, "Why should a materialist accept premise (2) of Hasker's argument?" Only if compelled? Surely this is far too strong as an epistemic requirement for accepting premise (1). Instead, says Shrader, a materialist ought to accept the principle he calls *Explain*:

*Explain:* All intrinsic properties of  $p$  (the brain) are explainable in terms of  $P_{\text{parts}}$  (where  $P_{\text{parts}}$  refers to the conjunction of the property instantiations of and relations between or among the parts of the brain).

According to Shrader, *Explain* entails that  $S_Q$  is explainable in terms of  $P_{\text{parts}}$  in that the conscious properties of the mind stand in a contingent causal relation to the physical properties of the brain or its parts. For Shrader, it is hard to see why there needs to be a conceptual or logical link between  $S_Q$  and  $P_{\text{parts}}$  as Hasker would have it. However, in point of fact, Shrader continues, Hasker is not opposed to adopting some sort of causal or nomological necessitation for the link, but he denies that this option is available to a materialist. Why? Basically, Hasker thinks this option is not available to the materialist because  $p$ , as a composite material object, is strictly speaking, a system of objects, and this implies there can be no causal link between the two.

Shrader thinks that Hasker's defense of this last claim boils down to an additional argument Shrader calls *Defense*. *Defense* includes several premises, but according to Shrader, the problematic one is (D6):  $p$  is not a concrete individual distinct from  $p_1, p_2, \dots, p_n$  (i.e.,  $p$  is no whole "over and above" its parts). Let us grant Shrader's claim that in an indirect way, (D6) commits Hasker to (D6)\*\*: The properties of  $p$  strongly supervene (with metaphysical necessity) on  $P_{\text{parts}}$ . Unfortunately for Hasker, this does not exclude  $p$  from being an individual involved in causal processes because supervenience is a synchronic, noncausal relation between  $P_{\text{parts}}$  and  $S_Q$ . Indeed, speculates Shrader, maybe the causal relation partially explains supervenience.

*Reply:* First, as I argued in my responses to arguments 1 and 2, it is misleading to say that a rejection of premise (2) is supported by (1) and (1) is supported by PR. I will not rehearse my responses here except to say that (1), the rejection of (2) and PR are ultimately supported by basic intuitions that are either the result of fundamental phenomenological awarenesses or careful consideration of the disjuncts of (2). Some of these intuitions are

embedded in Hasker's examples and illustrations which, as it turns out, actually have an argumentative function.

Second, if I understand Hasker correctly, he distinguishes between some entity *e* being more than simply the sum of its parts (and with respect to the brain, he seems to embrace this alternative) from *e* being an *object* in its own right that is "over and above" the sum of *e*'s parts (which Hasker rejects). One way both alternatives could be true is to say that *e* is composed not only by the *ps*, but also by all the relation instances between and among the *ps* (i.e., an instance of a kind of relational structure). But such a "whole" hardly constitutes an object that can sustain its own causal powers. It is very hard to see where those new causal powers would reside or how they could arise by simply placing the *ps* into new external relations.

What one needs for causal powers is an object in its own right, but to turn a relational aggregate into an object, one needs to add some metaphysical entity like a boundary, but most plausible candidates for addition are hardly the sorts of entities to sustain causal powers. And it is very difficult to see how adding a new set of external relations to the *ps* would turn the resulting aggregate into a genuine essentially characterized substance. I conclude, therefore, that Hasker seems right on this, and Shrader seems simply to help himself to contingent causal connections between the parts of the brain and the properties of consciousness without providing an adequate metaphysical grounding for such connections.

Third, Shrader seems confused about the nature of supervenience and causality. If the supervenient property is structurally supervenient, then supervenience is, indeed, constitutive and not causal. But in this case, we do not have phenomenal consciousness; there are only new structural arrangements of subvenient entities. So, this sort of supervenience is irrelevant to Hasker's argument. On the other hand, if the supervenient property is emergently supervenient, then we do have a *sui generis*, new, simply property (e.g., being an appearing of pain). But as John Searle has noted, in genuine cases of emergent supervenience, it is hard to see what else supervenience is besides a causal relation or a simple name for a metaphysical problem in need of solution (Searle 1992, 124–126). Unfortunately, in this case, Shrader faces all the problems I raised above in my defense of premise (2).

I conclude, then, that Shrader's case neither undercuts nor refutes Hasker's case for the UOC.

### 12.3 What, If Anything, Grounds the Three Types of Unity, Especially Subsumptive Phenomenal Unity?

As I mentioned earlier, the leading thinker on these matters is Tim Bayne. Because of this, and due to space limitations, I will focus my attention on Bayne's approach (Bayne 2010, ch. 12), explain why I think it is inadequate, and provide what I take to be a better solution.<sup>6</sup>

As Bayne sees it, there are three roles the unifier of consciousness (which he calls the self or subject of experience) must play: (1) Ownership – that which has conscious experiences; (2) Referential – objects of I-thoughts involved in first-person reflection; (3) Perspectival: "Selves" have a perspective, a first-person point of view. In my view, there are other key roles for the self (e.g., being the agent of libertarian acts, being that which is (metaphysically) possibly disembodied), but since, with qualifications, I accept Bayne's three, I shall not quibble with his list. However, I do want to make a few brief comments about it.

Regarding (1), there should be a clear, plausible metaphysical analysis of the sort of "having" it involves. Regarding (2), it should be expanded to say that it is not only the object

of the self-referring use of “I” but also that which employs “I” to self-refer. I also think it is question-begging at this early stage of analysis to use “I-thoughts” to characterize the nature of the role of reference the unifier must satisfy. Regarding (3), it is unclear what metaphysical notion of “having” is being used or what it is that does the having. More on this below.

Bayne begins by defending the phenomenal unity thesis on empirical grounds by beginning with the self as a biological organism, looking at specific cases of animalism and psychological views of unity to see if phenomenal unity prevails. He finds that it does. But then he shifts to a stronger a priori conceptual claim, namely, necessarily, *x* is a self if, and only if, *x* has phenomenal unity. According to Bayne, what we need is a phenomenalist (not a functionalist) conception of the self that allows us to construct selves out of streams of consciousness and affirm, as a matter of conceptual necessity, that no self can possess simultaneously two phenomenally unified streams of consciousness. Thus, if we entertain a thought experiment in which there are two functionally interactive, isomorphic streams of consciousness (i.e., the two streams exhibit functional unity) that, nevertheless, are not phenomenally unified, then we have two minds, not one.

This seems right to me. David Barnett offers a thought experiment that undergirds this intuition (Barnett 2010). Consider two people, say Fred and Ted, who have trained for years such that Fred and Ted can completely imitate the functional activities of the left and right hemisphere, respectively. Now suppose we take a third person, Joe, remove his brain, shrink Fred and Ted down to hemisphere size, and put Fred in Joe’s left hemisphere and Ted in his right. After the operation, from a third person perspective, there is only one person and one stream of consciousness present since Fred and Ted are completely functionally unified. But, says Barnett, given two streams of functionally integrated phenomenal consciousness, there are two persons present, not one.

So, Bayne’s project becomes one of finding a view that allows us to construct a self out of a phenomenally unified stream of consciousness. He considers and rejects two views – naive phenomenalism and substrate phenomenalism – and concludes by proffering his own view – virtual phenomenalism. Before we look at virtual phenomenalism, it will be instructive to see the way Bayne handles the two positions he rejects.

*Naive Phenomenalism:* According to naive phenomenalism, it is no great mystery as to why there is a 1:1 ratio between selves and streams of conscious *because selves just are streams of consciousness*. While Bayne ultimately rejects this view, he provides defeaters for a number of objections raised against it. Why would Bayne take the time to do this? In my view, it is because the metaphysics of his overall position entails that a person is synchronically identical to “his” totalizing phenomenal field and diachronically identical to (or, perhaps, the “same” as) “his” stream of consciousness. As we will see shortly, his own view (virtual phenomenalism) simply adds a Kantian-like twist to the metaphysics. In any case, in what follows I shall state the objection to naive phenomenalism, present Bayne’s response, and follow that with my own reply to Bayne.

*Argument 1:* Selves can’t *be* streams of consciousness – selves are things in their own right and streams are modifications of selves.

*Bayne’s Reply:* In some sense, streams *are* things in their own right, for example, they have their own principle of unity. The forces that knit together the components of a stream of consciousness are no less robust than those that knit together a single mind or animal.

*My Reply to Bayne:* Nowhere has Bayne demonstrated that a stream of consciousness has its own principle of unity, and the thin metaphysical framework within which he works – the totalizing phenomenal state “subsumes” its substates – employs an obscure metaphysical notion that provides no insight whatsoever as to how this is supposed to take place. Moreover, his analogy with “forces that knit together a single . . . animal” is quite revealing. When forces “knit” together independently existing entities upon which those forces work, the result is an atomistic building up of an ordered aggregate – the parts “knit” together are separable parts and the forces that bring and hold them together are external (causal) relations. You simply don’t get holism out of this; the analogy is atomistic, a view Bayne rejects.

*Argument 2:* Naive phenomenalism does not do justice to the sense in which streams are owned/had by selves.

*Bayne’s Reply:* This will be cashed out mereologically. By this, I take Bayne to mean that, not unlike bundle theories of substance, that we can give a reductive analysis of sentences like “(1): I am exemplifying the property of being-an-appearing-of-red to form a mode of me – being-appeared-to-redly” to “(2): A particular phenomenal state – being-appeared-to-redly – is a part of ‘my’ totalizing phenomenal state at that time.” Thus, ownership is a part/whole relation.

*My Reply to Bayne:* I will provide my alternative view of ownership later, but for now, I offer three brief replies. First, it seems difficult to read this view in any other way than atomistically: There is a separable part/whole relationship going on and the whole is simply a group of independent phenomenal parts standing in various external relations to each other. The problem here is that while Bayne merely and correctly points to the holistic nature of consciousness’ unity, he does not give a supporting metaphysical analysis of how this could be. Absent such an analysis, it is hard to avoid bringing to bear fairly standard metaphysical notions when it comes to evaluating his position, even when those notions entail propositions Bayne explicitly rejects. Second, a notorious difficulty for bundle or mereological theories of substance is that they seem to lack the ontological resources to ground absolute identity through standard changes. For many, this will be a problem. Finally, Bayne’s theory leaves opaque why most of us do not think our mental states are parts of us; rather, we take ourselves to be wholes that are not composed of our mental states. As we will see later, we think of ourselves as simple substances that “have” mental properties in that we exemplify them.

*Argument 3:* Naive phenomenalism can’t make sense of locutions such as “I weigh 185 pounds” or “I will die.”

*Bayne’s Reply:* Fairly obvious paraphrases are available.

*My Reply to Bayne:* I basically agree, except for one thing. I do not believe naive phenomenalism has paraphrases available that properly handle the indexical “I” (Madell 2015). For example, in handling objection 2, Bayne has to appeal to a part/whole relation between an independent phenomenal state and a totalizing state. But which totalizing state? I don’t think this can be answered without saying “mine,” or something indexically equivalent. But this is most naturally interpreted as claiming that I am one thing, and my totalizing phenomenal state is another, namely, a complex property I exemplify to form a mode of me.

*Argument 4:* Naive phenomenalism cannot account for modal properties of the self, for example, I could have had different experiences, but this would not be possible if I just am a



specific stream of consciousness. Indeed, one can conceive of a case in which there are no experiences in common between the actual and a merely possible self.

*Bayne's Reply:* Streams of events do not have their parts essentially. For example, World War II could have started as it did, but could have taken a different direction. Moreover, *any* conception of the self will have to deny at least some modal intuitions. The psychological view must reject the intuition that I could have had massive brain damage as a child. Animalism must reject the intuition that I could have been a different animal or disembodied.

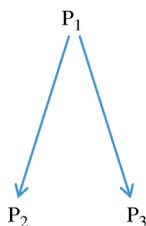
*My Reply to Bayne:* It is hard to see how a stream of events held together by external relations could, in fact, have had different events as parts. In the loose and popular sense, World War II could, indeed, have gone off in a different direction. But in the strict metaphysical sense, this war (call it World War II\*) would not be identical to World War II, but it could be treated as such for various purposes. And Bayne's claim that any view of the self must deny some intuitions is false. He only considers the psychological and animalist views. But a version of substance dualism that includes a mereologically simple, spiritual substance does not require abandoning basic intuitions in this area, including the three Bayne mentions. In fact, it seems pretty obvious that a substance-dualist conception of the person is actually the source of modal intuitions regarding the self, and advocates of alternative positions must tweak their views so as to be as close intuitively to substance dualism as possible without collapsing into substance dualism.

*Objection 5:* If your stream of consciousness fissions into two with psychological continuity, naive phenomenalism cannot tell us which stream I am.

*Bayne's Reply:* Maybe consciousness is so deeply unified that fission is impossible. Maybe the physical basis of consciousness is such that continuity is broken and neither is the original self. In any case, animalism and the psychological view have troubles here too.

*My Reply to Bayne:* Our intuitions that fission is possible are much stronger and better justified than Bayne's claim that "maybe" fission is impossible. But more importantly, Bayne admits that naive phenomenalism, animalism, and the psychological view have problems here and, in my view, Richard Swinburne has identified why this is the case (see Swinburne 1986, 147–151 and Swinburne 2013, 152–158).

Consider the following thought experiment. Suppose we perform a brain operation upon a person  $P_1$  in such a way that exactly half of his brain along with half of his body is transplanted and joined to one half body with half of a brain awaiting the transplanted parts. And the other half of  $P_1$  is likewise joined to a second half body/brain in another room awaiting the transplant. We might diagram the situation as shown in Figure 12.1.



**Figure 12.1** Brain transplant thought experiment.

Suppose further, that upon recovery, each of the two persons,  $P_2$  and  $P_3$ , manifest the same character traits and have the same memories as did  $P_1$ . Now consider the following question: There is a fact of the matter as to what happened to  $P_1$ , so where is he?  $P_1$  cannot be identical to both  $P_2$  and  $P_3$  for the very simple reason that one thing cannot be the same as two things. Thus, it seems that the following four options are our only possible ways to understand what happened: (1)  $P_1$  ceases to be and two new persons,  $P_2$  and  $P_3$ , come to be; (2)  $P_1$  survives and is identical to  $P_2$ , and a new person,  $P_3$ , comes to be; (3)  $P_1$  survives and is identical to  $P_3$ , and a new person,  $P_2$ , comes to be; (4)  $P_1$  partially survives in  $P_2$  and  $P_3$ .

Option 4 may make sense of physical objects like tables, but it is not a reasonable option with regard to persons. To see why, consider a second thought experiment. Suppose a mad surgeon captures  $P_1$  and announces that he is going to transplant his left brain hemisphere and body half into one half body/brain and the right hemisphere and body half into another half body/brain. After surgery, he is going to torture one of the resultant persons and reward the other one with a gift of a million dollars and a trip to Hawaii. You can choose which of the two persons, A or B, will be tortured and which will be rewarded. It is clear that whichever way you choose, your choice would be a risk. Perhaps you will cease to exist and be neither A nor B. But it is also possible that you will be either A or B. However, one thing does not seem possible – your being partially A and partially B. For in that case you would have reason to approach the surgery with both a feeling of joyous expectation and horrified dread! But it is hard to make sense of such a mixed anticipation because there will be no person after the surgery who will experience such a mixed fate. Partial survival, at least when it comes to persons, does not seem to make sense. Persons are mereological unities, not collections or combinations of things that admit of partial survival like physical objects.

Swinburne concludes that since there is a fact of the matter as to what happened to  $P_1$ , and since the thought experiment provides us with exhaustive knowledge of all the physical and psychological facts with respect to  $P_1$ , there must be additional facts that constitute the identity of  $P_1$ , and given the situation, the most plausible candidate for those facts are facts about  $P_1$ 's substantial soul/self.<sup>7</sup>

Returning to Bayne, he concludes that naive phenomenalism is too naive, and we must look elsewhere for a more adequate model. Bayne very briefly considers and rejects a second view – substrate phenomenalism – the view that the self is identical to the material substrate that underlies and generates consciousness. But this won't work, says Bayne, because there is no a priori guarantee that a single generative, underlying mechanism will produce only one stream of consciousness, and he is looking for a view that makes the unity of consciousness and its 1:1 relationship with a self a matter of a priori necessity. We now turn to a presentation of Bayne's own view.

*Virtual Phenomenalism:* For Bayne, the "self" is merely an intentional entity, one whose identity is determined by the cognitive architecture underlying the stream of consciousness; a sort of brain architecture that generates a fictitious entity like a character in a novel. So, the "self" is a virtual center of phenomenal gravity. In *de se* reference, the "subject" represents itself as itself; conscious states are automatically *de se*. Streams of consciousness are constructed around a single intentional object like a narrative is unified around the novel's main character. So the cognitive architecture underlying your stream of consciousness represents that stream as if it were had by a single self – the virtual object that is "brought into being" by *de se* representation.

The cognitive architecture underlying consciousness creates a unifying single subject/center of consciousness as a projected, virtual reality due to the *de se* nature of the constructed conscious states. A unified field projects one and only one virtual self. Other approaches go wrong in thinking that there must be a real entity that plays the role of the self, but the self is a mere intentional object (like Zeus?). The self isn't really real, but self-talk is still "legitimate" like we talk of a character in a novel.

What should we say about Bayne's virtual phenomenalism? To begin with, I must confess that when I first read it, I thought that Bayne must be kidding. In my view, his position amounts to a bunch of mere assertions that allow him to avoid an obvious solution: substance dualism. How could we ever tell whether virtual phenomenalism is true? What possible evidence could be marshaled for it? Ontologically speaking, I think his view is really just a version of naive phenomenalism (thus, Bayne's defense of the view) with a Kantian-style just-so "as if" story added to it.

Second, we don't start with a role that needs to be filled – if we did, we would have no idea what roles to choose (what about the role of digestion or circulating the blood) – and posit a self to fill it. No, we start with knowledge by acquaintance of our own simple self, and then upon reflection, we see that it plays various roles, for example, the unifier of consciousness (Moreland 2011). This approach is why we selected the brain as an organ of special importance in the first place. Bayne has it backward. Moreover, why do people throughout human history and all over the world take themselves to be indivisible, disembodiable souls? I think the answer is that people are simply able to be aware of themselves.

Third, if the phenomenal field just is unified in and of itself, what would be the need for the brain to project a (virtual) self? Why would conscious states automatically and of necessity be *de se*? I think Bayne just posits these as brute facts, but surely they are so odd, given his view, that we would be better off trying to find a different position that makes more sense of these and related issues.

Fourth, on Bayne's view, *de se* reference is systematically false. I-thoughts have no real, veridical intentional object. But surely this is far too extreme and skeptical. Part of what allows Bayne to get away with his view of *de se* reference is his inadequate characterization of the second role for the self, namely, to be an object of reference for I-thoughts. A more adequate characterization is this: The self is the object of the self-referring use of "I" but also that which employs "I" to self-refer. Once we see this, it becomes clear that there are no irreducible I-thoughts. Rather, there are substantial souls, selves, Is with the power of self-awareness and self-reflection that can be expressed indexically.

Finally, I believe Bayne's view gives inadequate analyses for the other two roles he claims the self must play: ownership and having a perspectival point of view. Regarding ownership, there is no real self that owns anything. There is simply the totalizing, holistic phenomenal field that has individual states as parts. Given his rejection of atomism, it is hard to see what kind of parts these are. At the very least, this aspect of his view is in need of considerable clarification.

Regarding the next role, what exactly is it that has this perspectival point of view? Given the arguments by Hasker and LaRock considered earlier, it can't be the brain because it is not a simple. Nor can it be the totalizing phenomenal field because for Bayne, that field just is the perspectival point of view in and of itself. In my view, for two reasons, there is no such thing as the property of being a (first-person) perspectival point of view that something exemplifies.

For one thing, if there is such a property, it is an impure one. An impure property, for example, being identical to Socrates or to the left of a desk, require reference to a particular to be described. Such a property cannot constitute such a referent without being circular – the property presupposes and, therefore, cannot constitute the particular – Socrates, the desk, or an individual person – to which reference is made. Similarly, being a first-person perspectival point of view presupposes the I.

For another thing, there most likely is no such property. In general, one may give a reductive analysis of the first-person perspective as follows: S has the property of being a first-person perspectival point of view if, and only if, S is a personal, viewing kind of point, that is, S is a kind of substance (point), a sentient (viewing) substance, with the properties (including ultimate potentialities) characteristic of persons (e.g., self-awareness and so on). The first-person perspective is not a property persons have, it is the thing persons are – centers of a personal kind of consciousness. Persons qua substantial, unified centers, exemplify ordinary mental properties – being-a-thought-that-P, being-a-sensation-of-red, being painful. But they do not have in addition to these the property of being a first-person perspectival point of view. When a substantial personal ego exemplifies an ordinary mental property that is *ipso facto* a first-person perspectival point of view. There is no additional fact that needs grounding in a superfluous property – being a first-person perspective. The “first-person perspective” is just a way of describing/referring to an ontologically prior substantial, sentient person with ordinary mental properties to which that perspective can be reduced.

Our discussion of Bayne’s view of the “self” is closely related to his depiction of a holistic view of consciousness. And while we have touched on this topic here and there, I want to finish my evaluation of Bayne’s overall position by examining more fully his account and defense of holism.

Bayne admits that we are a long way from having a theoretical account of what it is for consciousness to be unified (Bayne 2010, 45). Moreover, holism needs no mechanism to explain phenomenal binding (Bayne 2010, 244). We will see shortly why he thinks this to be the case. Bayne’s holism includes a metaphysical and a neurological aspect. I will focus on his metaphysical account and include his neurological ideas as they relate to metaphysical issues.

Metaphysically, Bayne says that in forming a view of consciousness we must avoid atomism and he uses a quilt analogy for consciousness (Bayne 2010, 244). The atomist says that we must glue together independent experiences to form a total phenomenal state. But this is wrong. The total state is basic. And like a quilt, the many coalitions of content are woven together to form an overarching state of consciousness.

As previously noted, subsumption is a mereological part/whole relation that explains phenomenal unity (experiences are phenomenally unified when parts of a single whole). To my knowledge, Bayne nowhere mentions inseparable parts, so by a part/whole relation it seems that he must mean separable parts standing in external relations to each other and the whole of which they are parts. He also says that phenomenal unity is a unitary relation between particular token mental states, and at least some experiences are parts of a single composite experience and, therefore, are phenomenally unified (Bayne 2010, 21, 31).

Finally, experiences come into existence *as* constituents of the whole phenomenal field. Thus, the existence of experiences depends on the ontologically prior totalizing field (Bayne 2010, 236, 244). This is why no account of binding is needed for the holist. It

seems that this is just a brute fact about consciousness. Unfortunately, “experiences come into existence *as* constituents of the whole phenomenal field” merely provides conditions for the coming-to-be of experiences. We can also provide certain conditions among subatomic parts that are such that atoms come into existence as entities requiring these conditions to be present. But Bayne’s statement says nothing about the nature of the experiences themselves, and I think his assertion is consistent with both an atomistic and holistic view of consciousness.

Before we turn to Bayne’s neurological views about holism, I want to raise some problems with his metaphysical position. First, how could it be that the components of the phenomenal field are conscious only as components of that field? How could it be that individual experiences come into existence as aspects of the unified field? Up until these events take place, everything that exists – for example, the brain and nervous system – are aggregates composed of separable parts and external relations. All of a sudden, *presto*, a conscious experience comes into existence and it stands in internal relations to an ontologically prior, non-aggregated whole of which the experience is a mode? This seems like a magical brute fact without a magician! And it is a bizarre brute fact, given the explanatory resources of a naturalist world view as we have already seen. As we shall see, my account solves this problem easily. An experiential state’s coming-to-be amounts to the soul’s exemplifying a phenomenal property to form a mode of the soul, and the unity of consciousness is explained by the fact that all the united conscious experiences are exemplified by the same simple (not composed of separable parts; see below) of which the experience are modes internally related to that soul.

Second, his account is, despite protests to the contrary, atomistic. Atomism satisfies his account of subsumption as a mereological part/whole relation in which subsumption involves one experience being a part of the whole (and he seems to know only of separable parts which stand in external relations; thus subsumption and the totalizing field would be atomistically composed). He says that phenomenal unity is when one particular mental state token/event stands in a unitary relation with another. I have never heard of such a relation and have no idea what it is or how it would be located in a naturalist ontology. It seems to be an ad hoc solution that allows Bayne to avoid substance dualism. Finally, his quilt analogy is telling. The pieces of a quilt exist prior to their attachment to other pieces to form the quilt. And a quilt is just a collection of independent quilt patches externally connected to each other.

Third, he does give a neurological account that he claims solves the unity of consciousness. A state’s total neural correlate can be divided into two components: a differentiating correlate (accounts for the state’s content) and a non-differentiating component (that part of the correlate that remains when the differentiating component is removed). So visual experiences of motion and tactile experiences of our feet have different differentiating correlates but the same non-differentiating components. Different parts of the brain generate via their cooperative integration a unified visual experience, especially with regard to the non-differentiating correlates. The neural basis of consciousness involves reciprocal interactions between the thalamic and cortical processing. This dynamic core is grounded in subsystems and reaches out to various domain-specific processing nodes. The very nature of this process guarantees that any features that are made conscious are made conscious together.

It is hard to see how Bayne could know that the very nature of this process unifies consciousness as stated in the last sentence above. This is just a brute assertion, and,

indeed, an assertion that he needs to avoid substance dualism. Moreover, the description of the whole process is atomistic to the core. Bayne's problem is this: It is hard to believe that an atomistic brain – even one with different atomistic components standing in intimate external causal relations with other components – could give rise to a non-atomistic whole whose “components” are internally related, especially if emergence is local.

Bayne seeks to shore up his view by citing certain analogies he believes make his view more intuitively acceptable. Just as undetached regions of water molecules in a cloud are not typically regarded as clouds themselves, an undetached component of the field of consciousness is not itself an experience in its own right. And just as undetached regions of clay are not themselves statues, so undetached components of the field of consciousness are not themselves experiences.

Unfortunately, these are cherry-picked analogies and bad ones at that. There are a number of examples where the parts of a whole are, in fact, the same type of individuals as the whole. A solid chunk of salt has parts that are themselves chunks of salt. The undetached pieces of clay are themselves pieces of clay. A group of 4-dimensional time slices have 4-dimensional time slices as parts. And so on. By picking favorable analogies, Bayne fails to address the problem of how to determine which analogies – the helpful or unhelpful ones – are most relevant to consciousness and why.

Second, the analogies are themselves poor ones. Take the cloud analogy. When one draws very close to a “cloud,” it is hard not to take an eliminativist line regarding them: a cloud does not exist as whole object. Rather, it is a group of water vapor (or atomic simples) arranged cloudwise. The reason undetached regions of clouds are not themselves clouds is that there are no clouds in the first place. And if clouds are real objects, it is hard to see why undetached regions of clouds are not clouds. If you remove all the surrounding water molecules of a specific region without doing anything to that region itself, it is a cloud. Why would it change its classification by not doing anything to it and just changing an environment to which it is externally related?

What about the statue analogy? A statue is a lump of clay (marble, etc.) arranged in a certain way that is taken to be a statue by its maker or the community's agreement in which the statue exists to take it that way. The reason that undetached regions of clay are not statues is that no one takes them to be such. And that's it. These regions are still clay and that is a better analogy with consciousness than statues. For these reasons, I take it that Bayne's account of holism does not work and we need a different model.

Before I briefly present my alternative ontology of the unity of consciousness, it would be helpful to spell out some ontological notions I shall use to cash out my position.

*Properties.* A property is a universal, that is, something that can be nonspatially in, exemplified, possessed by many things at the same time. And as I have defended elsewhere, constituent realism is the best view of how properties relate to the ordinary particulars that “have” them (Moreland 2013a). According to constituent realism, properties are universals that, when exemplified, become constituents of the ordinary particulars that have them. Thus, if the mind exemplifies a mental property, say, the property of being-a-thought-of-London, then that property enters into the very being of the mind as a metaphysical constituent (Willard 1999).

*Parts.* Earlier I have discussed the difference between separable and inseparable parts, so I refer the reader to that discussion.

*Substances.* A *substance* =<sub>def.</sub> an essentially characterized particular that (1) has (and is the principle of unity for its) properties but is not had by or predicable of something more basic than it; (2) is an enduring continuant; (3) has inseparable parts but is not composed of separable parts; (4) is complete is species.<sup>8</sup>

*Spiritual Substances.* A *spiritual substance (self or soul)* is =<sub>def.</sub> (1) a substance; (2) metaphysically indivisible in being (though it may be fractured in functioning); (3) not spatially extended (though some characterizations hold that it may be spatially located); (4) essentially characterized by the actual and potential properties of consciousness.

*Internal Relations.* If something, A (say the color yellow) stands in an internal relation (brighter than) to B (say the color purple), then anything that did not stand in that relation to B could not be A. So if any color were not brighter than purple, it could not be the color yellow. If a thing X stands in an internal relation to another thing Y, then part of what makes X the very thing it is, is that it stands in that relation to Y.

Given this framework, it is fairly straightforward to spell out the nature of the unity of phenomenal consciousness and the nature of its ground. The substantial self (soul, I) is spatially unextended and not composed of separable parts. The self's having a mental state, say, an awareness of a table, occurs when the self exemplifies the property being-an-awareness-of-the-table – call this property P, and this forms a mode of the self, namely, the-having-of-P-by-the-self. This mode may be described as the self's being-appeared-to-tablely, it is an inseparable part of the self and it stands in an internal relation to the self. This is what a particular phenomenal state is. Synchronically, the various phenomenally conscious modes of the self are unified into one totalizing phenomenal mode (state) by being modes of the same simple self and by being internally related to that self. Finally, the self is a unique kind of substance in that it has the power of self-awareness and self-reference.

Curiously, in a few places Bayne expresses a view similar to mine. He offers a tripartite conception of the unity of consciousness according to which experiences are individuated by (1) the subject of experience, (2) time, and (3) phenomenal properties (Bayne 2010, 24). Bayne adds that background states of consciousness are ways the subject's overall phenomenal field is modified (Bayne 2010, 238).

Unfortunately, it is hard to see how Bayne can consistently embrace these statements, given other items central to his overall position. For one thing, Bayne does not believe in a real self or subject. He repeatedly says that the subject is “constructed” out of the stream of consciousness, it is a projected, virtual “reality” that is not actually real, though we act “as if” it is real. There is no “homunculus” or self that inspects (is aware of?) images (conscious states?) projected onto the screen of subjectivity (Bayne 2010, 230). If there is no real subject, it cannot be part of what individuates experiences.

Second, as we have already seen, it is the totalizing field that “has” experiences, not a subject. That's why he says that it is one's overall phenomenal field that is modified by background states of consciousness. But this does not seem consistent with the tripartite analysis because on that view, as I make explicit in unpacking the straightforward ontological implications of this analysis, it is the subject that is modified by experiences, background or not, not the totalizing field.

Finally, Bayne's tripartite view is most naturally taken to imply that a subject has a conscious experience by exemplifying the relevant phenomenal property. But as we saw

earlier, Bayne's mereological view of subsumption implies that it is the subsuming state that "has" the subsumed conscious state in a part/whole way. As I have argued, not only does this view make it hard to avoid an atomistic view of consciousness, it implies that it is the totalizing phenomenal field that has various experiences, not the subject, and the "having" is a part/whole relation and not the nexus of exemplification between a substantial subject and a phenomenal property.

Returning to my own view, it is not my purpose here to argue directly for substance dualism. I have done that elsewhere (Moreland 2009, ch. 5; Moreland 2013b). Rather I want to show how it addresses and solves the problems that I have claimed Bayne's view does not.

- 1 It solves the binding problem and underscores Hasker's argument considered earlier by employing a simple, substantial self.
- 2 It employs clear, standard ontological notions that have constituted the heart of ontology for a long time.
- 3 It provides a solid analysis and correction/clarification of Bayne's three roles for a self: *Ownership*: that which has conscious states is the simple self, and the "having" amounts to the self's exemplifying various properties of consciousness to form modes of the self. *Referential*: substantial, simple selves simply have the power of self-awareness and self-reference. Thus, the I both employs "I" in linguistic acts of self-reference and is the object of those acts. *Perspectival*: the substantial, simple self exemplifies and unifies various properties of consciousness that, in turn, have the property of intentionality. In this way, the self/I has an irreducible, unified totalized conscious state/mode that is about intentional objects. Because modes are inseparable parts internally related to its self, this totalizing state is unique to one self and cannot be shared.
- 4 It provides a way of relating the three types of unity: (i) *Objectual phenomenal unity*: Two or more states are so unified if they are experiences as being of the same object. (ii) *Subject phenomenal unity*: this occurs when all of one's phenomenal states are had by the same subject. (iii) *Subsumptive phenomenal unity*: two (or more) states are subsumptively phenomenally unified just in case there is something-it-is-like to be in both states simultaneously and conjointly. Subject phenomenal unity occurs when all of one's phenomenal states are modes of the same, simple I. This view is neither irrelevant nor true by definition. Rather, it is a substantive (!) thesis with a developed ontology. It is because of subject phenomenal unity that objectual phenomenal unity obtains: Two or more states are experiences as being of the same object (e.g., the color and shape of a table) because they belong to the same object and the substantial I is simply aware of the table as a whole, including its various aspects. Finally, the reason subsumptive phenomenal unity obtains is because it is the same self that exemplifies the conscious property constituting each phenomenal mode/state and the self's unification of these into a totalizing mode is due to the self's simplicity.
- 5 My model gives a simple explanation for why Frank Jackson's observation is correct: "I take it that our folk conception of personal identity is Cartesian in character – in particular, we regard the question of whether I will be tortured tomorrow as separable from the question of whether someone with *any* amount of continuity – psychological, bodily, neurophysiological, and so on and so forth – with me today will be tortured"



(Jackson 1998, 45). Moreover, people don't have to be taught to be dualists like they must if they are to be physicalists. Indeed, little children are naturally dualists. Summing up research in developmental psychology, Henry Wellman states that "young children are dualists: knowledgeable of mental states and entities as ontologically different from physical objects and real [nonimaginary] events" (Wellman 1990, 50).<sup>9</sup> The reason human persons all over the world and throughout history have overwhelmingly believed in a substantial self is because they *are* substantial selves and they have the ability to be aware of themselves.

## **12.4 What Kind of Universe Must We Live in for there to be Entities such as States of Unified Consciousness that Ground the Unity of Consciousness?**

At the beginning of this chapter I noted with approval Thomas Nagel's observation that proffered solutions to problems in philosophy of mind and personal identity have broad worldview implications. Space limitations mean that I can only gesture at some points here. But I have developed and defended them in considerable detail elsewhere (Moreland 2008). Simply put, the existence of consciousness, the precise nature of its unity and the entity that grounds its unity provide evidence for a theistic worldview and against a naturalist one.

For example, naturalist William Lyons argues that "[physicalism] seem[s] to be in tune with the scientific materialism of the twentieth century because it [is] a harmonic of the general theme that all there is in the universe is matter and energy and motion and that humans are a product of the evolution of species just as much as buffaloes and beavers are. Evolution is a seamless garment with no holes wherein souls might be inserted from above" (Lyons 1995, lv). Lyons's reference to souls being "inserted from above" appears to be a veiled reference to the explanatory power of theism for the existence of substantial souls/selves.

Similarly, Crispin Wright notes:

A central dilemma in contemporary metaphysics is to find a place for certain anthropocentric subject-matters – for instance, semantic, moral, and psychological – in a world as conceived by modern naturalism: a stance which inflates the concepts and categories deployed by (finished) physical science into a metaphysics of the kind of thing the real world essentially and exhaustively is. On one horn, if we embrace this naturalism, it seems we are committed either to reductionism: that is, to a construal of the reference of, for example, semantic, moral and psychological vocabulary as somehow being within the physical domain – or to disputing that the discourses in question involve reference to what is real at all. On the other horn, if we reject this naturalism, then we accept that there is more to the world than can be embraced within a physicalist ontology – and so take on a commitment, it can seem, to a kind of eerie supernaturalism. (Wright 2002, 401)

Again, Wright's reference to an "eerie supernaturalism" is obviously a recognition of the presence of a theistic explanation for various anthropocentric subject matters (e.g., consciousness, the unity of consciousness, a soul as its ground) if we do not reduce or eliminate them.

Further, John Searle's observation is worth pondering:

How is it that so many philosophers and cognitive scientists can say so many things that, to me at least, seem obviously false? . . . I believe one of the unstated assumptions behind the current batch of views is that they represent the only scientifically acceptable alternatives to the antisecularism that went with traditional dualism, the belief in the immortality of the soul, spiritualism, and so on. Acceptance of the current views is motivated not so much by an independent conviction of their truth as by a terror of what are apparently the only alternatives. That is, the choice we are tacitly presented with is between a "scientific" approach, as represented by one or another of the current versions of "materialism," and an "unscientific" approach, as represented by Cartesianism or some other traditional religious conception of the mind. (Searle 1992, 3–4; also see 31)

Finally, speaking of the fear of religion (i.e., the fear that theism may be true), Thomas Nagel frankly admits:

I speak from experience, being strongly subject to this fear myself: I want atheism to be true and am made uneasy by the fact that some of the most intelligent and well-informed people I know are religious believers. It isn't just that I don't believe in God and, naturally, hope that I'm right in my belief. It's that I hope that there is no God! I don't want there to be a God; I don't want the universe to be like that. My guess is that this cosmic authority problem is not a rare condition and that it is responsible for much of the scientism and reductionism of our time. One of the tendencies it supports is the ludicrous overuse of evolutionary biology to explain everything about life, including everything about the human mind. Darwin enabled modern secular culture to heave a great collective sigh of relief, by apparently providing a way to eliminate purpose, meaning, and design as fundamental features of the world. (Nagel 1997, 130–131)

In sum, I have tried to clarify the nature of the unity of consciousness and identify why it is a problem. In addition, I stated and critiqued what I take to be the best – though, ultimately, inadequate – solution as to what the unity of consciousness amounts to and what, if anything, grounds that unity. Finally, I presented my own answers to these questions and, briefly gestured toward the broader worldview implications of my solution.

## Notes

1. They go on to show that, under certain conditions, the subsumptive unity thesis (for any set of phenomenal states of a subject at a time, the subject has a phenomenal state that subsumes each of the states in that set) is materially equivalent with the logical unity thesis (for any set of phenomenal states of a subject at a time, the subject has a phenomenal state that entails each of the states in the set). But material equivalence is not identity and subsumption is still left as an alleged intuitive primitive.
2. For a defense of the idea that the hierarchy goes out, not up, and that so-called top-down causation is of no use in providing for free will, see Moreland (2016).
3. According to constituent realism, when an object exemplifies a property or a collection of objects exemplify a relation, the property or relation (universals) constitutes an immanent essence of the individuating property or relation instances, which in turn, particularize the individual structure in view, keeping it from being an abstract universal. I have defended this view elsewhere (Moreland 2013a).
4. I wish to thank William Hasker for his helpful comments about the weaknesses in Shrader's critique.
5. Later, I provide epistemic and metaphysical reasons for taking the modality of a and b to be a metaphysical necessity.

6. The other significant attempt to solve the unity of consciousness problem is offered by Baker (2000, especially part I). I have criticized her view elsewhere (Moreland 2009, 131–137). It may be worth mentioning that split-brain issues have been raised as a significant defeater for advocates of the sort of unity of consciousness like Bayne and me. In my opinion, Bayne (2008) has provided an adequate response to this problem.
7. Due to space limitations, I set aside Bayne's objection 6. Interacting with it is not essential to my project.
8. Two helpful treatments of substances and related entities are Hoffman and Rosenkrantz (1997) and Brown (2005).
9. I owe this reference to Stewart Goetz and Mark Baker.

## References

- Baker, Lynne Rudder. 2000. *Persons and Bodies: A Constitution View*. Cambridge: Cambridge University Press.
- Barnett, David. 2010. "You Are Simple." In *The Waning of Materialism*, edited by Robert Koons and George Bealer, 161–174. Oxford, Oxford University Press.
- Bayne, Tim. 2008. "The Unity of Consciousness and the Split-Brain Syndrome." *The Journal of Philosophy* 105(6): 277–300.
- Bayne, Tim. 2010. *The Unity of Consciousness*. Oxford: Oxford University Press.
- Bayne, Tim, and David J. Chalmers. 2003. "What is the Unity of Consciousness?" In *The Unity of Consciousness: Binding, Integration and Dissociation*, edited by Axel Cleeremans. Oxford: Oxford University Press.
- Brown, Christopher M. 2005. *Aquinas and the Ship of Theseus*. London: Continuum.
- Cleeremans, Axel. 2003. "Introduction." In *The Unity of Consciousness: Binding, Integration and Dissociation*, edited by Axel Cleeremans, 1–20. Oxford: Oxford University Press.
- Hasker, William. 1999. *The Emergent Self*. Ithaca, NY: Cornell University Press.
- Hasker, William. 2005. "On Behalf of Emergent Dualism." In *In Search of the Soul*, edited by Joel B. Green and Stuart L. Palmer, 75–100. Downers Grove, IL: InterVarsity Press.
- Hoffman, Joshua, and Gary S. Rosenkrantz. 1997. *Substance: Its Nature and Existence*. London: Routledge.
- Jackson, Frank. 1998. *From Metaphysics to Ethics*. Oxford: Clarendon Press.
- LaRock, Eric. 2013. "From Biological Naturalism to Emergent Substance Dualism." *Philosophia Christi*, 15(1): 97–118.
- LaRock, Eric. 2015. "Emergent Dualism is Theoretically Preferable to Reductive Functionalism." Unpublished manuscript.
- Lyons, William. 1995. "Introduction." In *Modern Philosophy of Mind*, edited by William Lyons, xlv–lxviii. London: J. M. Dent.
- Madell, Geoffrey. 2015. *The Essence of the Self*. New York: Routledge.
- McGinn, Colin. 1999. *The Mysterious Flame*. New York: Basic Books.
- Moreland, J. P. 2008. *Consciousness and the Existence of God*. New York: Routledge.
- Moreland, J. P. 2009. *The Recalcitrant Imago Dei*. London: SCM Press.
- Moreland, J. P. 2011. "Substance Dualism and the Argument from Self-Awareness." *Philosophia Christi*, 13(1): 21–34.
- Moreland, J. P. 2013a. "Exemplification and Constituent Realism: A Clarification and Modest Defense." *Axiomathes*, 23(2): 247–259.
- Moreland, J. P. 2013b. "A Conceptualist Argument for Substance Dualism." *Religious Studies*, 49 (March): 35–43.
- Moreland, J. P. 2016. "Why Top-Down Causation Does Not Provide Adequate Support for Mental Causation." In *Neuroscience and the Soul*, edited by Thomas M. Crisp, Steven L. Porter, and Gregg A. Ten Elshof, 51–73. Grand Rapids, MI: Eerdmans.
- Nagel, Thomas. 1997. *The Last Word*. New York: Oxford.
- Nagel, Thomas. 2012. *Mind & Cosmos*. Oxford: Oxford University Press.

- Searle, John. 1992. *The Rediscovery of the Mind*. Cambridge, MA: MIT Press.
- Shrader, Warren. 2006. "The Unity of Consciousness: Trouble for the Materialist or the Emergent Dualist?" *Faith and Philosophy*, 23(1): 33–44.
- Skrbina, David. 2005. *Panpsychism in the West*. Cambridge, MA: MIT Press.
- Smart, J. J. C. 1959. "Sensations and Brain Processes." *Philosophical Review*, 68(2): 141–156.
- Swinburne, Richard. 1986. *The Evolution of the Soul*. Oxford: Clarendon Press.
- Swinburne, Richard. 2013. *Mind, Brain & Free Will*. Oxford: Oxford University Press.
- Wellman, Henry. 1990. *The Child's Theory of the Mind*. Cambridge, MA: MIT Press.
- Willard, Dallas. 1999. "How Concepts Relate the Mind to its Objects: The 'God's Eye View.'" *Philosophia Christi*, 1(2): 5–20.
- Wright, Crispin. 2002. "The Conceivability of Naturalism." In *Conceivability and Possibility*, edited by Tamar Szabo Gendler and John Hawthorne, 401–439. Oxford: Clarendon Press.

# Problems with Unity of Consciousness Arguments for Substance Dualism

TIM BAYNE

The brain constructs a tenement for mind but fails to disclose the tenant.

Sherrington, *The Endeavor of Jean Fernel* (1946)

Consciousness is often taken to be an impediment to materialism and to provide a motivation for some form of dualism. Contemporary arguments for dualism typically focus on the qualitative character of consciousness. Materialists, it has often been argued, face insuperable objections in explaining why there is “something that it’s like” to be conscious, and why what it is like to be in one kind of conscious state is different from what it is like to be in other kinds of conscious states (see, e.g., Chalmers 1996). Although some of those who endorse arguments of this kind are substance dualists, more commonly they espouse a dualism only of properties, and either remain silent on the nature of the subject of experience or identify subjects of experience with purely material entities, such as brains or organisms. This chapter sets arguments from the qualitative character of consciousness to one side, and focuses instead on arguments for dualism that appeal to the unity of consciousness. Unlike arguments from the experiential nature of consciousness, unity of consciousness arguments are explicitly designed to establish subject dualism rather than property dualism; indeed, advocates of unity of consciousness arguments are often silent about the nature of conscious properties themselves.

Objections to materialism that appeal to the unity (or “simplicity”) of consciousness have a venerable history; in fact, they seem to predate objections to materialism that focus on the qualitative character of consciousness. In the early modern period one can find unity of consciousness arguments in the writings of Descartes and Leibniz, and in the recent literature they have been defended by David Barnett, William Hasker, and Richard

Swinburne (among others). The idea that the unity of consciousness is an impediment to materialism – and by the same token supports dualism – clearly has a deep and abiding appeal. I will argue that this appeal is not warranted.

### 13.1 Descartes on the Unity of Consciousness

Descartes's unity of consciousness argument for dualism is to be found in the sixth of his *Meditations on First Philosophy*. Although this argument has been overshadowed by the other two arguments for substance dualism that Descartes gives in the *Meditations* – the argument from doubt and the argument from the conceivability of disembodiment – Descartes claims that his unity of consciousness argument was itself sufficient to establish substance dualism. The argument runs as follows:

there is a great difference between the mind and the body, inasmuch as the body is by its very nature always divisible, while the mind is utterly indivisible. For when I consider the mind, or myself insofar as I am merely a thinking thing, I am unable to distinguish any parts within myself; I understand myself to be something quite single and complete. Although the whole mind seems to be united to the whole body, I recognize that if a foot or arm or any other part of the body is cut off, nothing has thereby been taken away from the mind. (Descartes 1996, 59)

Descartes's claim that the mind is "utterly indivisible" is on its face puzzling, for it seems evident that the mind can be divided in multiple ways. We can divide the mind into its various faculties, distinguishing action from perception, and distinguishing both of these faculties from the faculty of thought. We can divide the mind into its various acts, distinguishing making an inference from retrieving a memory; and we can distinguish both of these acts from the act of making a decision. And we can divide the overall stream of consciousness into its various components, distinguishing between experiences associated with distinct perceptual modalities, and between perceptual experiences, bodily sensations, and affective experiences. Rather than the mind being "utterly indivisible," it would seem to be more accurate to say that there is no end to the ways in which the mind is divisible.

Descartes was aware of this general line of response, but he denied that it undermined his argument. As he put it,

As for the faculties of willing, of understanding, of sensory perception and so on, these cannot be termed parts of the mind, since it is one and the same mind that wills, and understands and has sensory perceptions. By contrast, there is no corporeal or extended thing that I can think of which in my thought I cannot easily divide into parts; and this very fact makes me understand that it is divisible. (Descartes 1996, 59)

At first glance it might be puzzling what Descartes has in mind here, for it is not clear why he thinks that his critic need deny that it is one and the same mind which wills, senses and understands, or indeed why they need deny that it is one and the same mind which makes inferences, retrieves memories, makes decision, and has experiences of different kinds. Why couldn't the critic agree with Descartes that although it is indeed I myself who am the subject of willing, sensing, and understanding, it is nonetheless the case that I engage in these activities in virtue of the fact that one part of me wills, another senses, and a third

understands? Descartes seems to assume that the mind as a whole is the basic (non-derivative) subject of willing, sensing, and understanding. That claim might indeed be correct, but it is not at all clear what entitles Descartes to assume its truth. Claiming that it is “one and the same mind which wills, which senses, which understands” certainly fails to provide it with any support.

A second problem with Descartes’s argument is that it assumes that any parts that the mind might have would need to be structured in ways that are introspectively accessible. Even if none of the distinctions that folk psychology recognizes fail to demarcate distinct parts of the mind, it is possible that the mind might decompose along other lines. It is possible that Descartes is at this point leaning on his views about the transparency of the mind, and assuming that if the mind has parts then the divisions between these parts would need to be introspectively apparent. But if this is indeed the assumption on which Descartes is leaning then so much the worse for his argument, for few contemporary theorists are persuaded that the mind is transparent to introspection in the required sense.

But perhaps Descartes’s central point is not that the mind is indivisible, but rather that the parts into which it can be divided are not capable of independent existence. To use E. J. Lowe’s (1996) useful phrase, perhaps Descartes is denying only that the mind has “substantial parts.” This interpretation of the argument perhaps receives some support from its capacity to illuminate Descartes’s puzzling reference to the fact that “if a foot or arm or any other part of the body is cut off, nothing has thereby been taken away from the mind” (Descartes 1996, 59). Consider a human body. Although it forms a unity, its parts can exist in isolation from each other. (Amputated limbs can survive in freezers and decapitated heads can be preserved in formaldehyde.) By contrast, the “parts” of a mind cannot exist independently of each other. An act of willing cannot exist independently of a mind that wills; the making of a decision cannot exist independently of a mind that decides; and neither perceptual experiences nor propositional attitudes can exist in isolation from the minds to which they are attached. Moreover, these claims appear to be conceptual truths rather than empirical generalizations that might be subject to revision in light of future findings. Thus understood, Descartes’s argument can be understood as an implicit rejection of an atomistic conception of mental phenomenon, according to which mental states, acts and episodes can exist independently of the particular mind in which they are located – or indeed independently of any mind at all.

I will return shortly to the question of why the rejection of atomism might be thought to put pressure on the materialist, but let us first explore in more detail the claim that the mind lacks substantial parts. A critic might suggest that even if individual mental items (experiences, thoughts, and so on) are incapable of independent existence, perhaps the results of the split-brain experiments show that the mind has substantial parts of *some* kind. In the words of one of the leading split-brain surgeons, perhaps the split-brain data show that “when you divide the brain surgically by midline section of the cerebral commissures the mind also is correspondingly divided” (Sperry 1984, 661). And if that is the case, wouldn’t it show that the mind has substantial parts?

Let us first note that there is disagreement about precisely how to interpret the split-brain data. Sperry holds that split-brain patients have two, independent, streams of consciousness, one of which is associated with left-hemisphere activity and one of which is associated with right-hemisphere activity. Although this “two-streams” view dominates the literature, other accounts of the split-brain data have been offered. One alternative to the two-streams view is that the split-brain procedure creates a fragmented stream of consciousness, in

which patients have pairs of simultaneous experiences that are each unified with a third experience but not with each other (Lockwood 1989; Schechter 2014). I myself have defended a third view of the split-brain, according to which consciousness in the split-brain remains unified and the appearance of disunity is created by rapid switches between left-hemisphere activation and right-hemisphere activation (Bayne 2008). Evaluating the respective merits of these accounts would take us too far away from our present concerns; suffice it to note that Sperry's claim is not uncontroversial.

But let us assume – as the majority of commentators do – that split-brain patients have two streams of consciousness. Would this show that the mind has substantial parts? No, for even if the split-brain procedure brings two minds into being, further argument is needed in order to show that the descendent minds (or parts thereof) were parts of the ancestor mind. Consider what happens when a single nation gives rise to two nations, as when, for example, Czechoslovakia gave rise to the Czech Republic and Slovakia. The two resulting states derive from the ancestor state, but it is a further question whether either of the descendent states was a proper part of the ancestor state. (Typically they won't have been, for nation states don't ordinarily contain states as proper parts.) Another analogy: one can create two houses from the materials provided by the destruction of one house, but it is a further question whether either of the resulting houses were parts of the ancestor house. (Typically they won't have been, for houses don't ordinarily contain houses as proper parts.) The split-brain operation might show that it is possible to create two minds from the constituents of a single mind, but this doesn't show that the original mind was composed of substantial parts. Indeed, the holism of the mental undermines the suggestion that any of the mental items that are associated with the two descendent minds might also have been associated with the ancestor mind.

I have suggested that Descartes's claim that the mind lacks substantial parts is left unscathed by the split-brain data. This obviously doesn't vindicate Descartes's claim, but it is *prima facie* plausible, and I am happy to accept that minds lack substantial parts. What implications might the mind's lack of substantial parts have for the prospects of materialism? Why might Descartes have thought that the mind's lack of substantial parts is inconsistent with materialism?

I suspect that Descartes was reasoning as follows: "If materialism is true then the mind must be a (nonfundamental) physical entity – it must be an organism or a (nonfundamental) part thereof. But all nonfundamental physical entities have substantial parts. Organisms clearly have substantial parts, as do brains and all of their macroscopic parts. (Indeed, the divisibility of the majority of brain parts is manifest in their bilateral nature, with the pineal gland functioning as a noteworthy exception to this general rule.) So if the mind lacks substantial parts then materialism is false."

But if this is how Descartes reasoned then his argument was fallacious. The argument might pose an objection to identity theorists, but the materialist need not be an identity theorist, and indeed most materialists are functionalists rather than identity theorists. From the functionalist perspective, the identification of a mind with a brain is a category mistake – it reifies minds. Minds are not substances, but are systems that emerge from the appropriate functioning of an organism (or a part thereof). The existence of a mind is grounded in and supervenes on the activity of an organism (or a part thereof), but minds are not things in the way in which organisms (and their parts) are things. From the functionalist perspective, it would be preferable to say that an organism is minded or has mental properties than to refer to "its mind." Crucially, the fact that the material basis of a mind has substantial parts no



more entails that minds themselves have substantial parts than the fact that the economy has a material basis entails that an economy has substantial parts.

### 13.2 Leibniz on the Unity of Consciousness

Let us turn now to Leibniz's "unity of consciousness" argument. In one of the most celebrated passages in the *Monadology* Leibniz presents the following objection to materialism:

we must confess that the perception, and what depends on it, is inexplicable in terms of mechanical reasons, that is, through shapes and motions. If we imagine that there is a machine whose structure makes it think, sense and have perceptions, we could conceive it enlarged, keeping the same proportions, so that we could enter into it, as one enters into a mill. Assuming that, when inspecting its interior, we will only find parts that push one another, and we will never find anything to explain a perception. And so we should seek perception in the simple substance and not in the composite or in the machine. (Leibniz 2000, para. 17)

Although this is clearly an argument for substance dualism rather than an argument for property dualism, at first glance it is difficult to see why it qualifies as a unity of consciousness argument, for it contains no obvious appeal to the unity of consciousness. However, it is reasonable to treat this as a unity of consciousness objection to materialism (as it often is) on the grounds that the arguments involve the claim that the conscious subject must be a unity, not just in the sense that it must be a single thing but in the more fundamental sense that it must be simple. On the basis of that claim Leibniz argues that the subject cannot be a material entity, for none of the material substances that might plausibly be identified with the subject of experience is simple.

How does Leibniz attempt to establish that the self is simple? His argument is curious. He begins with the relatively plausible claim that the operations of a mind that is constructed in the manner of a mill could never explain consciousness. Leibniz's worry here is best understood as a version of the explanatory gap objection to materialism (Levine 1983; Chalmers 1996). The idea, in a nutshell, is that mechanistic explanation can account only for structural-cum-functional phenomena, and because mechanistic explanation is the only form of explanation that is available to the materialist when it comes to accounting for mental phenomena, it follows that the materialist can account for mental phenomena only if such phenomena can be fully analyzed in structural-cum-functional terms. But of course there are good reasons to doubt whether mental phenomena *can* be fully analyzed in structural-cum-functional terms. Thus, we will never find anything to explain a perception by appealing "only to parts that push one another," as Leibniz puts it.

So far, one might think, so good – but how do these considerations motivate the claim that the self must be simple? Leibniz might indeed have shown that the materialist faces a serious (and potentially unbridgeable) explanatory gap, but how would positing a simple subject of experience help? Leibniz moves directly from the failure of mechanical explanation to the conclusion that the self must be simple – "so we should seek perception in the simple substance and not in the composite or in the machine" – but I fail to see the motivation for this inference. It might be justified if it were obvious how the activity of a simple substance could explain consciousness in all its myriad manifestations but that is

patently not obvious, and Leibniz fails to provide even a sketch of how such an explanation might go. Indeed, my hunch is that the dualist is forced to treat the relationship between the subject of experience and its states of consciousness as primitive. But doing that, it seems to me, is no advance at all on versions of materialism that posits metaphysically brute relations between physical-functional states and states of consciousness. If Leibniz had offered us an account of how the operations of a simple substance explained consciousness then we might be in a position to compare the merits of that account with those of the accounts offered by materialists, but he didn't and so we can't.<sup>1</sup>

### 13.3 Barnett on the Unity of Consciousness

With the arguments of Descartes and Leibniz in the background, let us now turn our attention to three of the unity of consciousness arguments that have been defended in the recent literature. Might contemporary versions of the unity of consciousness objection represent improvements over the versions developed by Descartes and Leibniz?

One of the more complex unity of consciousness objections to materialism is due to David Barnett. The starting point of Barnett's argument is what he calls *The Datum* – the intuition that a pair of people cannot be conscious:

You might pinch your arm and feel a pain. I might simultaneously pinch my arm and feel a qualitatively identical pain. But the pair we form would not feel a thing. Pairs of people *themselves* are incapable of experience. (Barnett 2010, 161; emphasis in original)

Barnett argues that the only plausible explanation for *The Datum* is that in order to be conscious an entity must be simple. He then uses that conclusion to reject materialism, for – echoing Descartes and Leibniz – he claims that no version of materialism is consistent with the simplicity of the self.

Barnett's route from *The Datum* to the simplicity of the self is not straightforward. As I read it, his argument begins with the claim that there are only four features that a materialist could reasonably appeal to in order to explain *The Datum*. She could invoke the fact that pairs of people lack: the right number of immediate parts (*number*); immediate parts standing in the right kinds of relations to each other (*relations*); immediate parts of the right nature (*nature*); immediate parts that have the right kind of structure (*structure*). Barnett argues that none of these features can account for *The Datum*, even when they are considered collectively. But if these features cannot account for *The Datum*, then – Barnett concludes – the human body is “no better a candidate for being a subject of experience than a pair of people” (Barnett 2010, 168). But if the human body is not a plausible candidate for being a subject of experience then something that is ordinarily associated with a human body must be. Barnett says little about what that something might be, but the obvious candidate for this role is an immaterial self.

The crucial step in this argument is clearly the claim that none of the four features just mentioned can account for *The Datum*. Barnett argues for that claim by considering a number of entities that are in some sense “intermediate” between a pair of human beings on the one hand and a normal human body on the other, suggesting that in each case it is “absurd” to suppose that the entity in question might be conscious. The evident absurdity of ascribing consciousness to each of these intermediate entities is meant (I take it) to help

motivate the (far-from-evident) absurdity of ascribing consciousness to a normal human body, since there is no relevant difference between the intermediate entities and a normal human body. So, at least, I take Barnett to be arguing.

One of the intermediate entities that Barnett considers is Ned Block's (1978) "miniature men in the head" creature. In this scenario we are to imagine

that the head of an otherwise normal human is filled with a group of little men. . . . Also inside the head is a bank of lights connected to inbound sensory neurons, a bank of buttons connected to outbound motor neurons, and a bulletin board on which a symbol (designating the current state of the system) is posted. Each man is given a simple set of instructions: if a given symbol is posted, then if certain lights are illuminated, press a given button. Together, the billions of men function, on a relevant level, just as a normal human brain functions. Yet the idea that this collection of tiny men might also be conscious is absurd. (Barnett 2010, 168)

Barnett concludes that since a normal human brain functions "in just the way in which a system of miniature men functions," and since it would be "absurd" to ascribe consciousness to the system of miniature men, it follows that we cannot appeal to functional (that is, structural and relational) considerations to explain why it is not appropriate to ascribe consciousness to a pair of human bodies. And since, Barnett claims, the same point can be made with respect to both of the other features that materialists might plausibly appeal to in accounting for *The Datum* (that is, *Number* and *Nature*), it follows that the materialist cannot justify the intuition that individual human beings, but not pairs of human beings, can be conscious.

It seems to me that this line of argument is wholly unconvincing. For one thing, it is highly doubtful that the various "intermediate" entities that Barnett consider really do elicit the kind of intuitive response that Barnett takes them to or – more importantly – that his argument requires. I certainly don't share his view that it is absurd to ascribe consciousness to the system of miniature men. I agree that it is not *obvious* that the system of miniature men would be conscious, but little of significance follows from that claim, for it is equally true that it is not obvious (a priori, that is) that *anything* should be conscious. This point is really nothing more than a restatement of the claim that facts about consciousness are not a priori entailed by any other kinds of facts – they are conceptually primitive. (Those who embrace behaviorism or analytical functionalism regarding consciousness will reject this claim, but it is pretty much common ground between everyone else.) Crucially, Barnett's argument doesn't require that it is not obvious that the miniature men system is conscious; instead, it requires that it is obvious that it is *not* conscious. And that claim is highly controversial. Why should we assume that consciousness would be absent if indeed the billions of tiny men function "just as a normal human brain functions"? To assume that is to assume that functionalist accounts of consciousness are false. Functionalism might indeed be false, but rejecting it would be question-begging in the present context. Barnett's argumentative strategy requires that our intuitive response to the tiny men scenario should not differ from our intuitive response to the scenario involving a pair of individuals, and that it is as "absurd" to attribute consciousness to the system of tiny men as it is to attribute it to a pair of human beings. I don't find that claim at all intuitively plausible, and the fact that these two scenarios differ from each other along dimensions which materialists reasonably take to be relevant to the presence of consciousness surely motivates treating them very differently.<sup>2</sup>

We might also note that the very structure of Barnett's own argument seems to presuppose that our intuitive response to the pair of people is not on a par with our intuitive response to the miniature men scenario, for if one thought that it was equally "absurd" to ascribe consciousness to a system composed of miniature men as it is to ascribe it to a pair of human beings why discuss the latter scenario at all? Why not simply begin with the miniature men scenario? It seems clear that Barnett begins by asking his reader to reflect on whether a pair of people could be conscious because he assumes (rightly, I think) that it is much more obvious that a pair of people couldn't be conscious than it is that a system of miniature men couldn't be conscious. But this fact not only undermines Barnett's argument, it actually provides positive motivation for materialism, for the materialist has the resources to explain why we are more inclined to think that a system of miniature men could be conscious than we are to think that a pair of individuals could be conscious. By contrast, it is not obvious that the dualist has the resources to explain why our intuitions about the miniature men differ from our intuitions about a pair of people, or indeed why they differ from our intuitions about ordinary human beings. After all, why should we suppose that immaterial selves cannot attach themselves to systems of miniature men or pairs of human beings just as easily as they can attach themselves to individual human beings?

There is a second problem with Barnett's objection to materialism. He writes:

One way to show that no combination of Number, Relation, Nature, and Structure can explain *The Datum* is to consider the human body, not as we ordinarily do, as a solid, human-shaped object, but rather as a structure of many organs, or of billions of cells, or of quadrillions of particles. We need to make salient the composite aspect of the body. The more salient we make this aspect, the less comfortable we will be ascribing consciousness to the body itself, until, at the limit, the whole idea will seem absurd. (Barnett 2010, 167)

I am happy to concede that many people might feel slightly "queasy" about ascribing consciousness to human beings when they consider them not as whole, unified, organisms but rather as complex structures composed of billions of tiny particles. I have no idea just how common such feelings might be, but I wouldn't be surprised to discover that they are relatively widespread; I myself have had such feelings on occasion. But why we should assume – as Barnett's argument requires us to – that such intuitions are to be trusted? More precisely, why should we privilege these intuitions over those that are prompted by considering the human body as a "solid, human-shaped object" – intuitions that clearly motivate the ascription of consciousness?

In fact, there are good reasons to think that intuitions that are generated in the context of adopting the "mereological stance" toward an object – that is, treating it as nothing over and above the sum of its parts – are generally untrustworthy. Consider a painting as nothing more than a structure of quadrillions of particles and one might be "less comfortable" ascribing aesthetic properties to it; consider a society as nothing more than a structure of quadrillions of particles and one might be "less comfortable" ascribing political properties to it; consider a nation state as nothing more than a structure of quadrillions of particles and one might be "less comfortable" ascribing economic properties to it. In each case we seem to be subject to what we might call "the mereological illusion," in which adopting the mereological stance toward an object undermines one's willingness to attribute "high-

level” properties to it.<sup>3</sup> Barnett’s argument, I submit, will be compelling only to those who fail to recognize the mereological illusion for what it is.

### 13.4 Hasker on the Unity of Consciousness

A second theorist who has defended unity of consciousness objections to materialism is William Hasker, who takes as his point of departure Leibniz’s analogy of the mechanical mill. The crucial feature of the mill, says Hasker, lies in the fact that it is:

made up of many distinct parts, coupled with the fact that *a complex state of consciousness cannot exist distributed amongst the parts of a complex object*. The functioning of any complex object such as a machine, a television set, a computer, or a brain consists of the coordinated function of its parts, which working together produce an effect of some kind. But where the effect to be explained is a *thought*, a state of consciousness, what function shall be assigned to the individual parts, be they transistors or neurons? Even a fairly simple experiential state – say, your visual experience as you look around this room – contains far more information than can be encoded in a single transistor, or a single neuron. Suppose, then, that the state is broken up into bits in such a way that some small part of it is represented in each of many different parts of the brain. Assuming this is to be done, we have still the question: *who or what is aware of the conscious state as a whole?* For it is a fact that you *are aware* of your conscious state, at any given moment, as a unitary whole. So we have this question for the materialist: when I am aware of a complex conscious state, what *physical entity* is it that is aware of that state? This question, I am convinced, does not and cannot receive a plausible answer. (Hasker 2010, 181–182; emphasis in original)

This passage would benefit from some unpacking, for it contains a number of quite distinct lines of argument.

One line of argument should already be familiar, for it is in effect a simple restatement of Leibniz’s explanatory gap objection. Hasker’s version of this objection adds nothing to Leibniz’s, and we needn’t repeat the points that we have already made concerning it.

A second line of argument appeals to the idea that the contents of consciousness contain more information than can be encoded by individual neurons. Precisely what sorts of representational contents can be encoded in single neurons is still something of an open question. The popularity of “grandmother cells” – cells whose contents are fine-grained enough to represent particular individuals such as one’s grandmother – has been invigorated in recent years due to discovery of neurons that appear to represent particular buildings (the Sydney Opera house; the Taj Mahal) and people (e.g., Halle Berry; Jennifer Aniston) (Kreiman, Koch, and Fried 2000; Quiroga *et al.* 2005). Whether or not these neurons really do have the selectivity that is claimed for them is something of an open question (see Loosemore and Harley 2010 for some salutary skepticism), but even if these neurons are indeed “Grandmother cells” it is highly implausible to suppose that all conscious content is encoded by individual cells. Instead, much neural representation involves population codes that are distributed across many neurons. On these models, complex contents are not represented by particular neurons, nor is it the case that “the state is broken up into bits in such a way that some small part of it is represented in each of many different parts of the brain.” Instead, content as a whole is represented by the state of the

representational system as a whole. The details of how the brain manages to represent the contents of consciousness have yet to be worked out, but there is no principled reason for thinking that they can't be worked out, and thus no principled objection to materialism here.

A third line of argument that is implicit in the above passage involves the claim that a complex state of consciousness cannot be distributed between the parts of a complex object. The argument clearly assumes that the materialist is committed to the claim that complex states of consciousness must be distributed among the parts of a complex object.

Hasker doesn't say what it would be for a complex state of consciousness to be distributed between the parts of a complex object, but I suspect that he has something like the following in mind. Suppose that auditory experiences were fully located in the auditory cortex and visual experiences were fully located in the visual cortex. In that case, the complex state of consciousness consisting of a visual experience and an auditory experience would be distributed between the visual and the auditory cortices, and strictly speaking we should ascribe the auditory experience to one entity (the auditory cortex) and the visual experience to another entity (the visual cortex). But if these two experiences are ascribed to different entities then – the argument continues – we have failed to secure the unity of consciousness, for in its robust form the unity of consciousness requires that unified experiences are both states of the same entity (that is, they are co-subjective) and that they are co-conscious (that is, that they possess a conjoint phenomenal character). But if the auditory experience is fully located in one region of neural real estate and the visual experience is fully located in another region of neural real estate then it is hard to see how either of these two conditions could be met. The experiences might bear various kinds of external relations to each other (such as being supported by the same brain), but they would be no more internally related than *my* auditory experience and *your* visual experience are. Most importantly, they would be neither co-subjective nor co-conscious.

I regard this as the strongest unity of consciousness objection to materialism. That said, however, I don't think that it succeeds. Although an atomistic conception of consciousness of the kind just sketched does raise questions about how consciousness could be unified, I am not persuaded that the materialist is committed to atomism. Surely it could turn out that the fundamental unit of consciousness is the entire conscious stream, such that it is not possible for modality-specific experiences to be fully located in discrete regions of neural space? Indeed, one might well argue (as I have – see Bayne 2010) that the evident unity of consciousness is itself a reason to reject conscious atomism. But what of the science of consciousness?

Assuming that materialism is consistent with the rejection of conscious atomism, might Hasker nonetheless argue that the science of consciousness indicates that materialists should also be atomists?

If he were to make such a claim he certainly wouldn't be alone, for a number of theorists have argued that the science of consciousness supports atomism (see, e.g., O'Brien and Opie 1998; Zeki 2008). But although he wouldn't be alone he would be mistaken, for the science of consciousness is perfectly consistent with phenomenal holism – the claim that a subject's overall conscious state cannot be broken down into independent units of consciousness. So, at any rate, I argue.

The dominant argument for atomism begins with the claim that the neural mechanisms underpinning consciousness – the “neural correlates of consciousness” (NCCs), as they are often described – are not to be found in any one location but are instead scattered throughout the brain.

The multiplicity of cortical loci where correlations with awareness have been found provides some evidence against one of the oldest ideas about consciousness, that the contents of awareness are represented in a single unitary system . . . Instead, the data described above seem more consistent with a view in which the contents of current awareness can be represented in many different neural structures. . . . In contrast to the idea of a unitary and content-general Cartesian theatre of awareness, the data . . . fit more naturally with the following simple hypothesis: *the neural correlates of awareness of a particular visual attribute are found in the very neural structure that perceptually analyses that attribute.* (Kanwisher 2001, 97; emphasis in original)

The data that Kanwisher has in mind concern the fact that different perceptual attributes seem to be associated with activity in particular cortical areas. For example, the visual experience of motion is associated with activity in the middle temporal visual area of the brain (MT). Lesions to MT will produce deficits in the capacity to visually experience motion, and the artificial stimulation of MT will produce hallucinations of visual motion. We might think of MT as the locus of the analysis of visually represented motion, as Kanwisher puts it. It is an open question whether *all* conscious contents are represented in this localized manner – I myself doubt it – but let us assume for the sake of argument that the localization that seems to hold of visual motion holds more generally. Would it follow that the NCCs are distributed across many different neural structures, as Kanwisher suggests? I don’t think so.

In order to see why not we need to distinguish different types of NCCs. The kind of NCCs on which the science of consciousness has focused are *differentiating* NCCs (Bayne 2010; Hohwy and Bayne 2015).<sup>4</sup> A neural event functions as a differentiating NCC for conscious state C if and only if its occurrence is typically sufficient for the occurrence of C in a conscious creature. MT seems to be a differentiating NCC for visual experiences of motion, for the evidence suggests that MT activity is typically sufficient for conscious states of that kind in conscious individuals. Differentiating NCC must be distinguished from total NCCs. Whereas differentiating NCCs abstract away from the domain-general mechanisms that are implicated in conscious states of all kinds (what some authors misleadingly refer to as “enabling NCCs”), these mechanisms are built into total NCCs. Thus, it is total NCCs rather than differentiating NCCs that correspond most closely to the intuitive notion of an NCC – that is, the neural event that is minimally sufficient for the occurrence of the corresponding conscious event.

We are now in a position to see why Kanwisher’s discussion is potentially misleading. Although her comments are naturally understood as suggesting that MT functions as a total NCC for visual experiences of motion, the available evidence suggests only that it is a differentiating NCC for visual experiences of motion, for all of the data linking visual experiences of motion with MT activation presupposes a conscious subject. Moreover, one wouldn’t expect a slice of MT that had been excised from a brain and placed in a petri dish to generate visual experiences, no matter how robustly it was zapped. It is plausible to suppose that in order to generate experiences of any kind, MT activity needs to be suitably

integrated with non-differentiating neural activity. Of course, none of the foregoing establishes that holistic approaches to consciousness are more plausible than atomistic approaches. That is not my point. Rather, my point is that the neuroscience of consciousness doesn't establish consciousness atomism.

Let us return to Hasker's argument. Although I have argued that the neuroscience of consciousness isn't committed to the claim that complex states of consciousness must be distributed among the parts of a complex object, it is obviously possible that the materialist might be committed to this claim for other reasons. However, if the materialist is so committed to that claim then surely the onus is on Hasker to show that this is the case. As far as I can see, Hasker has not done so.

Let me turn finally to Hasker's fourth unity of consciousness argument, which concerns a problem that the materialist allegedly faces in accounting for the awareness of complex conscious states. Hasker lays out this argument as follows:

- 1 I am aware of my present visual field as a unity; in other words, the various components of the field are experienced by a single subject simultaneously.
- 2 Only something that functions as a whole rather than as a system of parts could experience a visual field as a unity.
- 3 Therefore, the subject functions as a whole rather than as a system of parts.
- 4 The brain and nervous system, and the entire body, is nothing more than a collection of physical parts organized in a certain way . . .
- 5 Therefore, the brain and nervous system cannot function as a whole; it must function as a system of parts.
- 6 Therefore, the subject is not the brain and nervous system (or the body, etc.).
- 7 If the subject is not the brain and nervous system then it is (or contains as a proper part) a non-physical mind or "soul." . . .
- 8 Therefore, the subject is a soul, or contains a soul as a part of itself. (Hasker 2010, 182)

Premise (1) seem plausible, although I myself would want to distinguish between the various components of a visual field being experienced "as a unity" and their being experienced by the same subject of experience, for there is at least a conceptual gap between these two dimensions of the unity of consciousness. But the real puzzle here concerns not (1) but (2): what precisely does it mean for something to function "as a whole" rather than [just?] "as a system of parts"? Does something function as a whole in virtue of doing things that none of its parts do? In that case, photocopiers function as wholes, for only the photocopier itself produces photocopies. The functions of a photocopier's parts are obviously essential to the functioning of the photocopier as a whole – what it is for a photocopier to make copies *just is* for its parts to be appropriately related and for them to perform their various functions – but it is nonetheless true that only the photocopier makes photocopies. So, if this is what Hasker means by something having a function "as a whole" rather than as a "system of parts" then I see no reason to deny that an organism (or indeed the parts thereof) cannot function "as a whole" – indeed, there is every reason to think that organisms (and the parts thereof) *can* function as wholes. (There are clear echoes of the mereological illusion here.) I conclude that Hasker's unity of consciousness arguments for substance dualism are no more persuasive than the previous arguments that we have considered.



### 13.5 Swinburne on the Unity of Consciousness

In a number of places Richard Swinburne has argued that substance dualism is “forced upon anyone who seriously reflects on the fact of the unity of consciousness at a time and over time” (Swinburne 1997, 160). Let us consider Swinburne’s arguments for this claim, beginning with the unity of consciousness at a time.

The heart of Swinburne’s argument from the synchronic unity of consciousness is contained in the following passage:

neuroscience seems to indicate that the immediate causes of conscious events of different kinds (e.g., visual sensations, auditory sensations, or olfactory sensations, occurrent thoughts, etc.) include events in different parts of the brain; and also that the immediate causes of different properties (e.g., the colour and the shape) of what we must regard as one conscious event (e.g., perception of a coloured shape) include events in different parts of the brain. So we would fail to tell the whole history of the world if we traced only the history of each part of the brain, regarded as a separate substance, and the instantiations of mental properties most immediately causing or caused by events in that part; for there would then be truths about properties (such as co-experienced sensory properties) which we would have to attribute – falsely – to two different substances. (Swinburne 2013, 143)

Taking certain exegetical liberties, we might formalize Swinburne’s argument as follows:

- 1 It is a necessary truth that co-conscious experiences are had by the same substance.
- 2 The neuroscience of consciousness indicates that if materialism were true, then many co-conscious experiences would be assigned to distinct substances.
- 3 Therefore, materialism must be false (and the substance of human experiences must be an immaterial entity rather than the organism or any part thereof).

Most materialists – at least, those who hold that conscious states are states of substances in the first place – will agree with (1). But few, I think, will find (2) plausible. Swinburne’s sole argument for (2) involves the paragraph quoted above, in which he claims that “neuroscience seems to indicate that the immediate causes of conscious events of different kinds (e.g., visual sensations, auditory sensations, or olfactory sensations, occurrent thoughts, etc.) include events in different parts of the brain.”

There are obvious echoes of Hasker’s third unity of consciousness argument here, and clearly everything turns on what Swinburne means by “the immediate cause of a conscious event.” His claim has some plausibility if we are to understand this notion in terms of a differentiating NCC, but I see no reason to equate the substance to which an experience belongs with its differentiating NCC. MT activation might indeed function as the differentiating NCC for visual experiences of motion, but it wouldn’t follow that MT was the basic subject of such experiences. The prospects of equating the substance of an experience with its total NCC seem to be more promising, but for the reasons outlined above there is little reason to think that total NCCs are distributed throughout the brain. If the neuroscience of consciousness showed that the total NCC of an auditory experience was fully located in auditory cortex and the total NCC of a visual experience was fully located in visual cortex then *perhaps* Swinburne’s argument would have some bite,<sup>5</sup> but I know of no

reason to think that the total NCCs of visual and auditory experiences are located in visual or auditory cortices.

I turn now to what Swinburne's describes as his diachronic unity of consciousness argument for substance dualism.<sup>6</sup> Swinburne's central line of argument involves deploying a number of puzzle cases, and arguing that our intuitions about these cases can be accommodated only by adopting a "further fact" view of the self, according to which the continued existence of a person over time consists in the continued existence of a mental substance, and "it is metaphysically possible that the substance acquires a totally new body, totally new apparent memories, and character" (Swinburne 2013, 163). I have some sympathy with the idea that the continuity of the self is relatively independent of both bodily continuity and psychological continuity, and in previous work I have explored various ways of developing this idea (Bayne 2010; Dainton and Bayne 2005), but I am not convinced that our intuitions regarding these puzzle cases justify a further fact view of the self.

Swinburne's central puzzle case involves a twist on the familiar split-brain procedure. Here, Swinburne imagines that the two hemispheres of one person ( $P_1$ ) are split, with  $P_1$ 's right hemisphere then being transplanted into the head of an individual ( $P_2$ ) who has only a left hemisphere and  $P_1$ 's left hemisphere being transplanted into the head of a third individual ( $P_3$ ) who has only a right hemisphere. Each of  $P_1$ 's half-brains are then connected in the appropriate ways to the half-brains of  $P_2$  and  $P_3$ , so that  $P_2$  and  $P_3$  function as ordinary subjects of experience. Swinburne stipulates that  $P_2$  and  $P_3$  would exhibit equal degrees of neural and psychological continuity with  $P_1$ . The central question concerns what we are to say about the survival of  $P_1$  in this scenario.

Swinburne argues that  $P_1$  cannot be identical to both  $P_2$  and  $P_3$ , for they are not identical to each other, and identity is of course a transitive relation. That leaves, he suggests, only three possible accounts of what has happened: (1)  $P_1$  has survived only as  $P_2$ ; (2)  $P_1$  has survived only as  $P_3$ ; and (3)  $P_1$  has not survived. Swinburne then proceeds to argue against complex (or "reductive") accounts of personal identity – that is, views which deny that facts about personal identity outstrip fact about physical and psychological relations – on the grounds that they cannot accommodate the intuition that there is a fact of the matter as to which of these three scenarios is correct. Swinburne does not himself state which of these scenarios is correct – nor, for that matter, does he offer any guidance as to how we might possibly determine which of these scenarios is correct – but he insists that one of these three scenarios must in fact be correct.

Call the intuition that there is a determinate fact of the matter about what happens to  $P_1$  in this case "the determinacy intuition." I think that Swinburne is right in thinking that complex accounts of personal identity cannot accommodate the determinacy intuition, but I'm not convinced that we should follow him in assuming that this is an intuition that *ought* to be accommodated. Why shouldn't we simply jettison it?

Swinburne motivates the determinacy intuition by appealing to a device that will be familiar to many: Williams's (1970) famous thought-experiment in which one of the descendants will receive a million dollars and an enjoyable life while the other will be subjected to a life of torture.

The surgeon asks  $P_1$  to choose whether the person who will receive such-and-such parts will be rewarded and the other person tortured, or the person who will receive the other parts will

be rewarded and the first person tortured; and the surgeon promises to carry out  $P_1$ 's wishes . . . Being selfish,  $P_1$  wishes to be rewarded and not tortured. So how is  $P_1$  to choose? Whether someone's future life will be happy or painful, or whether they will continue to exist at all after the operation . . . do seem very clearly to be factual questions. Yet, as  $P_1$  awaits the transplant and knows exactly what will happen to his or her brain, they are in no position to know what will happen to them, and so in no position to know how to choose which subsequent person will be rewarded. . . . When we know everything about which planks in the ship of Theseus have been replaced or reassembled when, then we know all there is to know about what is the same and what is different about the subsequent ships; although there are different ways in which we can describe what has happened, they are logically equivalent to each other. But when we know everything about the extent to which later persons have the same brains and the same apparent memories and other mental life of earlier persons in the half-brain transplant experiment, it does look very strongly that there is still something all-important to know – as the mad surgeon's addition to the story brings out: it is the all-important fact about who survives the operation and what happens to them. (Swinburne 2013, 153–154)

I agree that the mad scientist scenario is effective in eliciting the determinacy intuition, but I don't think it provides us with reason to take that intuition seriously – that is, to afford it the kind of warrant which Swinburne's argument requires. Indeed, skepticism about the robustness of this intuition is motivated by the very article in which the mad surgeon scenario was introduced, for one of the signal lessons of that article is that intuitions about personal identity are vulnerable to framing effects.<sup>7</sup>

That being said, it is worth reflecting on the "all-important thing" that  $P_1$  wants to know in contemplating the mad surgeon's proposal, for it seems to me that reflection on this matter motivates the idea that  $P_1$  survives as both  $P_2$  and  $P_3$ . What  $P_1$  wants to know concerns the first-person perspective.  $P_1$  currently has a first-person perspective, and we can assume that each of the descendant individuals will also have a first-person perspective.  $P_1$ 's question is whether either (or indeed both) of these future first-person perspectives qualifies as a continuation of his or her current first-person perspective. To answer these questions we need to understand the mechanics of first-person reference, and the ways in which the continuity of the "I" is transferred across times, both retrospectively in the form of memory retrieval and prospectively in the form of the planning (that is, in the formation and execution of intentions). Although the splitting of the self complicates these relations in various ways (see Hirsch 1991), there seems to be no principled reason to deny that  $P_1$ 's first-person perspective isn't inherited by both descendants, for both descendants have first-person access to  $P_1$ 's experiences in the form of autobiographical memory, and both descendants will inherit, and will be disposed to implement,  $P_1$ 's intentions.<sup>8</sup>

Should we say that there is one person who survives as both descendants (and thus that the descendants are identical to each other – Dainton 1992), or should we say that there are two persons in this scenario, albeit individuals who share a common temporal part (Lewis 1976)? I see little to choose between these alternatives, for they agree on all the essential facts – namely, the ways in which experiences are distributed between first-person perspectives.

### 13.6 Concluding Thoughts

I bring this chapter to a close by taking a step back from the details of particular unity of consciousness arguments and reflecting on some of the general features of this family of objections to materialism.

As we have seen, unity of consciousness arguments focus on the alleged shortcomings that materialists face in attempting to account for the unity of consciousness. From this point, theorists move quickly – and typically without comment – to the claim that dualism is true. This line of argument is essentially a negative one. It assumes that dualism and materialism exhaust the theoretical alternatives, and that it must be possible for an immaterial substance to have a unified consciousness, since (the theorist assumes) the kind of consciousness that we enjoy is obviously unified, and (the theorist claims to have established that) our experience could not possibly be unified were we purely material beings.

What this approach manifestly fails to do is to provide any positive account of the relationship between substance dualism and the unity of consciousness. Not only do theorists make no attempt to show how dualism explains the unity of consciousness, they don't even make any attempt to show that it is consistent with the unity of consciousness.<sup>9</sup> One might well argue that the obstacles that dualists face in accounting for the unity of consciousness are no less pressing than those that materialists face in this regard. In fact, it seems to me that there is an important sense in which the obstacles facing the dualist are *more* pressing than those which face the materialist.

Unity of consciousness arguments concern the challenges posed by accounting for the contents of complex states of consciousness. Whether they focus on the components of a total experiential state that subsumes experiences associated with distinct perceptual modalities (Bayne and Chalmers 2002), or whether they consider instead the distinct perceptual attributes drawn from a single sensory modality, the dualist's interest is with the question of how the various components of consciousness are bound together to form a unified conscious state that is had by a single subject of consciousness. That is all well and good, but how is the dualist to account for the diversity of these experiential elements in the first place? How is the dualist to explain why one subject enjoys an experience of the sound of trumpets at one time and the smell of roses at another, or why another subject enjoys an experience of the sound of bell-birds together with an experience of the smell of roasting coffee? Substance dualists offer no answers to these questions, and given their insistence on the simplicity of the self it is difficult to see what answers they could offer to them. By contrast, materialism at least offers the prospect of accounting for the diversity-within-unity that consciousness exhibits, for the materialist associates consciousness with the operations of a complex system. In focusing on the challenges posed by accounting for the unity of consciousness, substance dualists have been guilty of overlooking those posed by accounting for its diversity.<sup>10</sup>

## Notes

1. As Angus Menuge has reminded me, dualists who are also theists might argue that a dualist conception of consciousness is overall simpler than a materialist conception, since the theist will need to posit basic relations between states of an immaterial substance and consciousness in order to account for God's consciousness.
2. There is reason to think that philosophical intuitions don't operate in a theoretical vacuum, but are instead theory-dependent. Given this fact, one could argue that an individual's willingness to ascribe consciousness to the system of miniature men is dependent on his or her prior (and perhaps unarticulated) commitments regarding such issues as the plausibility of functionalism. But if that is right then the intuitive absurdity of these scenarios is not pre-theoretical in the sense in which Barnett takes it to be (Barnett 2010, 169), and his argument would be robbed of much (if not all) of its dialectical force.

3. The mereological illusion should be distinguished from what Bennett and Hacker (2003) call “the mereological fallacy,” which is the fallacy of thinking that the proper parts of a conscious subject can possess the kinds of properties that only conscious subjects can possess (such as consciousness).
4. Differentiating NCCs are often referred to as a “core NCCs” (see, e.g., Block 2005; Chalmers 2000; Koch 2004).
5. Although even here materialists (about our identity) would presumably argue that (1) can be met by simply denying that cortical regions are substances. The only genuine substance in this ballpark (they might claim) is the organism of which the relevant neural region is a part.
6. I’m not convinced that this argument really deserves to be described as a unity of consciousness argument, but nothing of note turns on the label.
7. Williams presents two versions of what is essentially the same scenario. One version (typically) elicits the intuition that personal identity follows bodily continuity, and the other version (typically) elicits the intuition that personal identity follows psychological continuity. The explanation for these contrasting intuitions is provided by the different context (or “frame”) associated with the two versions of the scenario. Many theorists take Williams to have shown that thought experiments are not reliable ways of adjudicating between rival accounts of personal identity. Although that conclusion is perhaps premature, there is little doubt that Williams’s paper shows that our intuitions regarding personal identity are highly malleable. For some discussion see Dainton and Bayne (2005).
8. This entails either that the descendants are the same person (and in effect that the person is “scattered”), or that two individuals can have the same first-person perspective. Both descriptions are counter-intuitive, but given how unusual the situation is it is hardly surprising that there is no intuitive way of describing its results. Thanks to Angus Menuge for prompting me to say more here.
9. Some theorists might be tempted by this thought that simplicity of the self would entail that consciousness is unified, but I myself see little reason to endorse this thought. The simplicity of the subject of experience does not itself ensure – let alone “explain” – the unity of consciousness, and I see no obstacle in the idea that a “simple” entity could be in two conscious states at one and the same time without those two states being co-conscious with each other.
10. I am grateful to Angus Menuge for his very helpful comments on a previous draft of this chapter.

## References

- Barnett, D. 2010. “You Are Simple.” In *The Waning of Materialism*, edited by Robert Koons and George Bealer, 161–174. New York: Oxford University Press.
- Bayne, T. 2008. “The Unity of Consciousness and the Split-Brain Syndrome.” *The Journal of Philosophy*, 105(6): 277–300.
- Bayne, T. 2010. *The Unity of Consciousness*. Oxford: Oxford University Press.
- Bayne, T., and D. Chalmers. 2002. “What is the Unity of Consciousness?” In *The Unity of Consciousness*, edited by A. Cleeremans, 23–58. Oxford: Oxford University Press.
- Bennett, M. R., and P. M. S. Hacker. 2003. *Philosophical Foundations of Neuroscience*. Malden, MA: Blackwell.
- Block, N. 1978. “Troubles with Functionalism.” *Minnesota Studies in the Philosophy of Science*, edited by C. Wade Savage, vol. 9, 261–325. Minneapolis: University of Minnesota Press.
- Block, N. 2005. “Two Neural Correlates of Consciousness.” *Trends in Cognitive Sciences*, 9(2): 46–52.
- Chalmers, D. 1996. *The Conscious Mind*. Oxford: Oxford University Press.
- Chalmers, D. 2000. “What Is a Neural Correlate of Consciousness?” In *Neural Correlates of Consciousness: Empirical and Conceptual Questions*, edited by T. Metzinger, 17–39. Cambridge, MA: MIT Press.
- Dainton, B. 1992. “Time and division.” *Ratio* (n.s.), 5: 102–128.
- Dainton, B., and T. Bayne. 2005. “Consciousness as a Guide to Personal Persistence.” *Australasian Journal of Philosophy*, 83(4): 549–571.
- Descartes, R. 1996. *Meditations on First Philosophy*, edited and translated by John Cottingham. Cambridge: Cambridge University Press.

- Hasker, W. 2010. "Persons and the Unity of Consciousness." In *The Waning of Materialism*, edited by Robert Koons and George Bealer, 175–190. New York: Oxford University Press.
- Hirsch, E. 1991. "Divided Minds." *Philosophical Review*, 100: 3–30.
- Hohwy, J., and T. Bayne. 2015. "The Neural Correlates of Consciousness: Causes, Confounds and Constituents." In *The Constitution of Phenomenal Consciousness*, edited by S. Miller, 155–176. Amsterdam: John Benjamins.
- Kanwisher, N. 2001. "Neural Events and Perceptual Awareness." *Cognition*, 79(1–2): 89–113.
- Koch, C. 2004. *The Quest for Consciousness: A Neurobiological Approach*. Englewood, CO: Roberts and Company.
- Kreiman, G., C. Koch, and I. Fried 2000. "Category-specific Visual Responses of Single Neurons in the Human Medial Temporal Lobe." *Nature Neuroscience*, 3: 946–953.
- Leibniz, G. W. F. 2000. "Leibniz, *Monadology*." In *Readings in Modern Philosophy. Volume 1: Descartes, Spinoza, Leibniz and Associated Texts*, edited by Roger Ariew and Eric Watkins, 285–293. Indianapolis, IN: Hackett.
- Levine, J. 1983. "Materialism and Qualia: The Explanatory Gap." *Pacific Philosophical Quarterly*, 64: 354–361.
- Lewis, D. 1976. "Survival and Identity." In *Philosophical Papers*, edited by David Lewis, vol. 1, 55–77. Oxford: Oxford University Press.
- Lockwood, M. 1989. *Mind, Brain and the Quantum*. Oxford: Blackwell.
- Loosemore, R., and T. Harley. 2010. "Brains and Minds: On the Usefulness of Localization Data to Cognitive Psychology." In *Foundational Issues in Human Brain Mapping*, edited by Stephen José Hanson and Martin Bunzl, 217–240. Cambridge, MA: MIT Press.
- Lowe, E. J. 1996. *Subjects of Experience*. Cambridge: Cambridge University Press.
- O'Brien, G., and J. Opie. 1998. "The Disunity of Consciousness." *Australasian Journal of Philosophy*, 76(3): 378–395.
- Quiroga, R., L. Reddy, G. Kreiman, C. Koch, and I. Fried 2005. "Invariant Visual Representation by Single Neurons in the Human Brain." *Nature*, 435(June): 1102–1107.
- Schechter, E. 2014. "Partial Unity of Consciousness: Evidence and Implications." In *The Unity of Consciousness and Sensory Integration*, edited by D. Bennett and C. Hill. Cambridge, MA: MIT Press.
- Sherrington, C. 1946. *The Endeavour of Jean Fernel*. Cambridge: Cambridge University Press.
- Sperry, R. 1984. "Consciousness, Personal Identity and the Divided Brain." *Neuropsychologia*, 22(6): 661–673.
- Swinburne, R. 1997. *The Evolution of the Soul*, 2nd edn. Oxford: Clarendon Press.
- Swinburne, R. 2013. *Mind, Brain, and Free Will*. Oxford: Oxford University Press.
- Williams, B. 1970. "The Self and the Future." *Philosophical Review*, 79: 161–180.
- Zeki, S. 2008. "The Disunity of Consciousness." *Progress in Brain Research*, 168: 11–18.

## Further reading

Hasker, W. 1999. *The Emergent Self*. Ithaca, NY: Cornell University Press.

# Debating Near-Death Experiences

# Evidential Near-Death Experiences

GARY R. HABERMAS

The popular subject of near-death experiences (NDEs) occupies a potentially crucial place in scholarly discussions of topics such as human nature and the possibility of an afterlife. For instance, there is definitely some hypothetical relevance for serious contributions to the chief metaphysical issue discussed in this volume, namely, the nature of human persons. In this essay we will investigate primarily one key subject: the topic of whether NDE observations provide any potential evidence for the existence of a conscious human self during a ND state, such as when neither the heart nor the brain (especially the upper brain) register any known activity. Our dialogue partner is Michael Marsh, whose 2010 volume and other writings since then have hypothesized that NDEs do *not* signify or manifest any other reality beyond death.

Most researchers on both sides of the NDE issue largely agree that the chief means of adjudicating the relevant issues, either as a potential pointer to the mystery of human nature, or as a hint of an afterlife, is to ascertain whether or not NDEs are veridical events. Experiencers (NDErs) have often reported accurate observations that they apparently made during their NDE, including ones that they could otherwise not have known or observed from their location, perhaps while they were even unconscious. It is to this issue of potential evidence for the NDE that we turn for the majority of this chapter.

## 14.1 Michael Marsh's Hypothesis

For Marsh, the available evidence warrants “a brain-based origin” (2010, xvii) for NDEs. On his thesis, NDEs are subjective events, the reactions of brains that are “recovering from various antecedent clinical crises” (p. 241), especially as the brain “is rapidly gaining conscious-awareness” (March 2010, 252, also xx, 29, 105; 2016, 10). This is why, in order to hold this view, Marsh realizes that he must argue against the veridical NDE claims of opposing researchers, since at least in part, “Such a critique underpins the foundational tenets of the present book” (Marsh 2010, 29, 97; 2016, 1–2, 9).



To advance his hypothesis, Marsh makes use of his medical training and additional scientific studies to get to the heart of his thesis, which is developed through several discussions (2010, chs. 4–9; 2016, 6–14). He maintains that neurophysiological explanations serve us better than otherworldly subjectivity (2010, 96), that brain-based illusions can explain the evidential claims (2010, 97), and that “the brain can replicate all of the experiential facets of OB phenomenology” (2010, 127).

Further, phantom limb phenomenology helps us to understand our self-illusions (2010, 102–105), while “dream-state pathology” is compared to the “quasi-dreamlike phenomenology of both NDE and OBE” – out-of-body experiences (2010, 156–157). Pro-NDE researchers are then taken to task for not paying sufficient attention to the latest studies in “disturbed temporal lobe function” (2010, 158–159, 168–169). Lastly, Marsh adds the specific challenges posed by factors like endorphin metabolism, hypoxia, anoxia, hypercapnia, and ketamine (2010, ch. 9). His conclusion is that human consciousness never leaves the human body (2010, 52, 97).

Beyond the neurological aspects, Marsh also attributes some of the problems with NDEs to cultural and personal interpretations (2010, 34–38, 92–94, ch. 12; 2016, 19–20), the related intrusion especially of Eastern as well as other unwarranted and far-fetched metaphysical ideas (2010, 38–52, 249–252, 256; 2016, 16, 20), and the lack of any reason to suspect divine revelation here (2010, 241, 256; 2016, 1, 20). Yet, Marsh is immensely impressed with the real transformations that seem to affect NDErs (2010, 244–256, 263; 2016, 20–21).

While Marsh does comment on a few evidential claims reported by some NDE researchers, this is done rather disappointingly in a hit-and-miss manner. He examines nowhere near the number or sort of cases one might expect, especially given that the success of his thesis may actually depend on the data here, nor does he treat the best-evidenced examples (2010, 15–27, 29–32).

For instance, Marsh critiques some of the very earliest, intriguing cases produced by cardiologist Michael Sabom and lodges four criticisms: Sabom’s examples are too trivial, they may be accounted for by the brain, health care providers need to attempt rescues rather than reading numbers on machines for the sake of NDE corroboration, and placing concealed numbers near the ceiling of emergency rooms simply does not help us (2010, 123–136). Plus, the NDErs could always have learned the information before or after their experience (p. 31).

Admittedly, questions like the brain’s potential involvement in NDEs, or learning information from outside the ND window are crucial, but NDE researchers have long recognized these possibilities. Yet, Marsh’s challenges to these NDE data just miss the mark. Asserting simply that observing actual numbers on machines is a trivial endeavor, or that one’s attention should rather be directed to saving lives, or by declaring the stupidity of testing NDEs diverts attention and fails to address the crux of the matter. Yet, such maneuvers certainly do not address the major NDE issues.

Marsh should direct more serious, detailed efforts to ascertaining the possible presence of potential confirmation concerning the available plethora of NDE evidence, or to developing more solid critiques with which to confront the NDE researchers who produce substantive evidential claims. Time and again, he comments briefly that the evidence produced by NDErs is insufficient or inadequate, but he rarely either attempts to counter specific instances or to deal with the best evidential NDE cases.<sup>1</sup> In fact, it might be charged that the opposite is true: he appears to choose cases that are addressed more easily. Perhaps

this lack is due to Marsh's personal conviction that the physiological considerations noted above are simply sufficient to prove his thesis. Whatever the reason, we will argue below that his responses comprise a severe miscalculation of the amount and quality of the available corroborative data.

In a last, rather strange, move noted by reviewers like Edward Kelly (2010), Marsh switches gears quite abruptly and unabashedly in chapters 10–12 of his text, moving to what are apparently his own Christian convictions (Marsh 2010, 219). Several times he mentions key theological themes like the resurrection of Jesus (pp. 195–200), as well as Christian doctrines such as bodily resurrection (pp. 200–211), the Trinity (p. 218), and the nature of revelation (pp. 219–222, 241), holding them to be true. In fact, his major problem with NDEs seems to be that these reported events often conflict with Christian views (2010, 218; 2016, 5–6).

The problem here is not so much that Marsh holds these Christian doctrines to be true, in the sense that everyone accepts overarching philosophical positions and convictions in life. In this sense, we all have tinted glasses through which we view life and reality. That is not the issue here *per se*. Rather, perhaps the major problem with this theological angle is that Marsh seems to simply and rather naively postulate the truth of these Christian doctrines without providing any specific evidence in their favor, then in turn posits them as the gold standard to which the NDE reports must somehow conform. However, specific evidence and reasons are needed in support of our worldviews, especially if we are then going to employ them as the grounded, established paradigm for judging other positions!

But this seemingly odd move by Marsh, namely to critique NDEs according to his own Christian presuppositions which he has not even defended as true, leads to an additional serious issue. Since he attributes NDEs to natural causes, chiefly to the experiencer's waking consciousness, his natural hypothesis does not allow him to truly account for the incredible personality changes exhibited by NDErs. These changes are acknowledged by Marsh, without issue. But he wants subjective ND experiences, with life-changing results regularly arising from them. But can he have his thesis both ways?

The problem is that he now seems to have wedged himself into a very tough spot. Since he considers that NDEs are only subjective, confused experiences inside the brains of NDErs, why do they regularly produce such incredible, often life-long changes? Marsh acknowledges these phenomena and is clearly nonplussed by the transformations, and he concludes that he needs to respond adequately. While he deserves good grades for creativity, the more he discusses this question, the more his view reveals another flaw that shows his position quite clearly to be inadequate.

Marsh postulates that in no sense should NDEs be considered as any sort of revelatory visions or messages from God, let alone even being objective experiences. Rather, these internal, subjective sensations result from "metabolically disturbed or temporally hypoxic brains." However, while addressing another researcher who postulates that NDEs could indicate humankind's next evolutionary stage, Marsh objects that such an option would be "illogical, if not absurd" on the grounds that NDEs proceed from subjective roots in a brain that is "recovering from various antecedent clinical crises" (2010, 240–241).

Yet, in an effort to still explain adequately the quite extraordinary personal transformations that come quite frequently from what he takes to be confused, disturbed brains, Marsh concedes that these profound changes are quite likely "moments of divinely directed grace towards each individual" (p. 256). Really? This is nothing short of amazing! In other words, Marsh is clearly torn between his thoroughly subjective view of NDEs, as confused and

disturbed brain events, while still being confronted with the genuine and positive life changes that often occur in the NDEr. Here Marsh feels like he must resort to some quasi-objective, divine infusion of grace that blesses what he has already declared to actually be an over-wrought, muddled, and chaotic mental state!

Then while Marsh judges that the evolutionary perspective mentioned directly above is “absurd” due to this same impaired condition of the NDEr’s mind, yet he wants us to accept that his own, even more exalted view of divine grace working through these equally impaired minds hits the mark exactly? Here we observe that what is good for Marsh’s goose clearly is not extended or applied by him equally to the other man’s gander!

Even worse, this is not the only spot where Marsh’s thesis runs aground on the NDE data itself. As we will observe next, Marsh’s chief problem turns out to be his inattention to and inability to explain the literally hundreds of corroborated NDE cases that have been documented carefully in recent literature. As scholars on both sides recognize, if the NDE evidence obtains, then the authenticity of NDEs can stand on their own.

## 14.2 Evidential NDEs

Given the foregoing, Marsh’s physiological explanations are incapable in themselves of building sufficient groundwork that automatically nullifies or eliminates any actual veridical evidence for NDEs. For starters, an interdisciplinary cadre of specialists is involved in NDE research and there are certainly alternative ways to evaluate and interpret the same medical data that Marsh employs (Beauregard and O’Leary 2007). As Kelly (2010) asserts, the physiological and neurological material raised by Marsh is well known and has been researched for many years, but it fails to eliminate any potential ND corroboration. It might be said that whereas Marsh’s thesis seeks to explain what *might* be occurring subjectively, evidenced cases more directly indicate that much more apparently *did* occur objectively.

While no one objects to Marsh raising cautionary questions, objections, and possible alternative hypotheses, yet he likewise needs to exercise due diligence not to reject real veridical data with comments like accurate observations on medical monitors during NDEs concern only trivial matters, or that researchers need to pay more attention to other things besides NDE corroboration. Quick and superficial dismissals like these point to an apparent lack of seriousness or perhaps even an overall uneasiness concerning the many dozens of veridical claims.

With so many researchers across the theoretical landscape agreeing on the import of treating the veridical NDE cases, where is the discussion on this absolutely crucial issue in recent years? Arguably the most insightful dialogue occurred from 2007 to 2008 over four full issues in the *Journal of Near-Death Studies*. A vigorous debate took place largely between a skeptical philosopher Keith Augustine (specializing in NDEs), and several distinguished NDE scholars. In these spirited discussions, Augustine proposed many specific challenges, aimed especially at the evidential NDE reports. Augustine’s respondents usually addressed his challenges in areas where they were the lead investigator. Augustine provided both the lead essay as well as the final rejoinder in each of the first three issues. It may have been the most rigorous investigation of claimed NDE corroboration.

In his lead installment of three major articles, Augustine (2007a) questioned several well-known NDE episodes, such as Kimberly Clark Sharp’s tennis shoe reportedly observed on the roof of a Seattle-area hospital, Pam Reynolds’s brain surgery, and NDE reports among

blind persons. It also provided a brief overview of prospective experiments which attempt to ascertain whether NDErs were able to identify random visual targets placed high overhead.

Augustine's second main article (2007c) examined numerous cases of *false* perception in NDE accounts, including discrepancies with the physical world, encounters with living persons that never occurred, imagery like mythological creatures or fictional characters, and so on. He argued that these sorts of perceptions are much more compatible with hallucinations.

Augustine's final main essay (2007d) argued that other ND features, like psychophysiological and cultural correlates suggested that NDE imagery is solely the product of an individual's mind rather than a supernatural reality. Throughout, Augustine argued against NDEs operating either beyond or independently of the human brain.

Augustine also allowed that the most impressive species of NDE data would probably be more indicative of an afterlife thesis than that of his hallucinatory hypotheses, if they could stand up to scrutiny. For instance, citing positively the work of Michael Potts (2002), Augustine postulated that the following would be among the most helpful in establishing an otherworldly interpretation of NDEs: specific clothing details of those health care personnel who resuscitated the patient, the precise order of events during the resuscitation, emergency room details that could have been learned only by being present, and so on.

One general response to Augustine pointed out that arguing for an afterlife was seldom the goal of NDE research (for example, Greyson 2007). But Augustine (2007b), again following Potts, replied that while technically these studies may not have aimed at establishing an afterlife, it would be very difficult to deny that such a conclusion would follow rather naturally if human consciousness were capable of functioning after death, especially if natural explanations failed to suffice, as a large number of studies seemed to indicate.

Augustine (2007b, 269 [original emphasis]) concludes: "If there were evidence of the sort Potts outlined, *then* the data would contradict my critique of near-death veridicality studies; but, as Potts also noted, anything of the sort has yet to happen." It is noteworthy that Augustine is another skeptic who agreed repeatedly that evidential NDEs are necessary in order to indicate a case for an afterlife or to refute naturalism. Naturalist Susan Blackmore (1993, 113, 125, 128, 262–263) similarly asserts the need for NDE evidence and allows that such could even show her view to be mistaken.

Accepting the evidential challenges laid down especially by Augustine and earlier by Blackmore (1993), both of which were precursors to some of Marsh's 2010 remarks, the *Journal of Near-Death Studies* respondents buttressed their particular areas of study. For instance, Kimberly Clark Sharp (2007) defended her account with the various confirmed specifics of Maria's tennis shoe spotted during the patient's heart attack. Other evidential NDE cases of many different sorts were also argued.

#### 14.2.1 Evidential NDE corroboration within the room

Before moving to an evidential case for NDEs, a brief response needs to be made regarding the rather distinct issue of criticisms by Augustine (2007a) and other researchers concerning the prospective experiments that have attempted to determine whether any NDErs were able to detect random visual targets such as numbers placed high overhead. The critics enjoy pointing out that while some partial information has been collected here and there, no

full identifications of the random numbers by NDErs have yet occurred. These critical scholars may hope that these issues could serve to cast doubt on all NDEs.

In response, as will be noted more specifically below, some half-dozen or more evidential NDE cases *have* included either the successful identification of specific numbers, or the recognition of other casually placed or thrown objects that happened to be found overhead, that could only be seen from a position nearer to the ceiling. These include the successful identification and repetition of a random 12-digit numeral on the top of a medical device, a four-figure number, and two of the author's own cases where smaller numbers were visible only to the NDEr. In a somewhat related case, a five-figure number was also correctly identified.

This list would be increased significantly if we also included the correct identification of unlikely or strange objects not seen from the ground, but found by NDErs up near the ceiling. For instance, coins have been found and retrieved after being spotted during NDEs. These examples indicate that skeptics should be a bit more reserved in their criticisms of the prospective cases, given that comparable results have already been correctly observed by NDErs.

Augustine suggested (also via Potts) that evidential reports be produced regarding the specific clothing worn by resuscitation team members, the precise order of events that transpired, and other emergency room details that could have been known only by one who was present at the time. Although Augustine and Potts declared that such details were nonexistent, their comments were clearly mistaken, and on several fronts, as will be shown in this chapter. On many occasions, it appeared to be more a matter of disbelieving or disregarding the details when they were actually produced. Certainly this would apply in the cases of correctly cited numbers and other odd objects that could only be observed from overhead!

Pertaining to accounts where emergency room clothing was identified by NDErs, Ken Ring and Madelaine Lawrence (1993) reported the intriguing account of Joyce Harmon, an ICU hospital nurse. On her first day back at work after vacation, she was a member of the medical team that successfully resuscitated a female patient whom she did not know. The very next day she saw the patient, who responded, "Oh, you're the one with the plaid shoelaces!" and explained that she observed them while watching the resuscitation from overhead.

Intriguingly, Harmon had just purchased the plaid shoelaces during her vacation and had worn them to the hospital for the very first time. Though casual or mundane conversations often occur in a hospital setting even during stressful times, the color of shoelaces does not appear to be the *most likely* scenario that would be discussed or even noticed during a frantic resuscitation attempt!

In another case,<sup>2</sup> a nurse practitioner of my acquaintance in a Midwestern hospital rushed to the scene involving a patient experiencing an emergency situation due to cardiac arrest, where she assisted in a successful procedure. A couple of days later an unknown patient introduced herself as the resuscitated patient. The latter explained that, during the resuscitation, she witnessed the process from above her body and had observed a rather unique object worn on the nurse practitioner's clothing, which the patient described in minute detail. The object had been borrowed the previous day and had already been returned.

The nurse practitioner was stunned most by the patient's intricate description of the odd object even during the cardiac arrest. Even if a conversation were overheard during the

resuscitation and recounted *so precisely* (quite a stretch in itself, if not impossible for someone in her condition), it is quite unlikely that all the details would have been described precisely during a cardiac arrest. If the story were heard later, how would the unconscious patient be able to identify the “owner”?

What about Augustine’s and Pott’s request for the knowledge of a precise *sequence* of events during a resuscitation? While there are many examples from which to draw, one of the most detailed was reported by emergency room pediatrician Melvin Morse (1990), who recounted the case of a girl (“Katie” – actually Kristle Merzlock) who had nearly drowned. A physician was present at poolside, and it was documented that Kristle was without a pulse for at least 17 minutes, as well as having no gag reflex, with fixed and dilated pupils, and was “profoundly comatose.” Morse noted that this condition most probably indicated that at least her upper brain was not functioning at that time, and that reversible brain damage had probably occurred.

Three days later, Kristle inexplicably revived. About three weeks later, in a follow-up exam, she took almost an hour to tell her entire story of that time. Morse was incredibly impressed with the precise sequencing of her emergency room events and descriptions. She knew that a tall physician without a beard was the first one to enter the emergency room. Then she recounted that Morse, shorter and sporting a beard, had come in next and was chiefly responsible for resuscitating her. She also recounted that she had first been brought into a larger room, and then was moved into a smaller one, for X-rays. She knew that she had been intubated through her nose, although this procedure is more commonly done orally.<sup>3</sup>

Given that Kristle was unconscious the entire time, with her eyes closed, requiring mechanical ventilation in order to breathe during the next three days, this is an incredible report. Even if it were thought that Kristle somehow could have heard certain snippets of emergency room conversation, details like the physicians’ physical characteristics and the sizes of the hospital rooms certainly seem to require sight, although that is certainly not the best explanation here. The clear and confirmed sequence is also beyond typical jumbled memories. Morse declared that this experience changed his life, including his religious agnosticism.

Morse (1990) relates other evidential NDE cases that also included a sequence of events, including an eight-year-old girl who nearly drowned in a swimming pool after her hair was caught in the drain. Her parents, then an emergency medical team that arrived, and lastly, physicians in the emergency room, all administered CPR for more than 45 minutes before her heart began beating once again. A short time later, she exhibited “full recall of the event” and was capable of recounting the entire extended process of resuscitation (pp. 32–33).

Another sequential case (Holden 2016) involved a woman who experienced a cardiac arrest that lasted a few minutes. During that time, she described the conditions and resuscitation attempts performed in the ambulance the entire way to the hospital. Perhaps most crucially, the paramedic crew members attested that several corroborated observations that she recalled occurred *precisely during her cardiac arrest and were reported immediately thereafter*, in which cases (as documented below regarding ventricular fibrillation) both heart and at least upper brain activity would be eliminated in just seconds, and then lower brain activity slightly afterwards.

Lastly, Augustine and Potts also requested emergency room descriptions that could only have been known by someone actually present. Again, many examples on record can be

drawn from here. We have already mentioned cardiologist Sabom's (1982) *ten* documented, specific examples where patients, sometimes without any measurable heartbeat, detailed many items regarding the layout of furniture and other objects in the hospital emergency room, specific observations regarding the instruments that were located there, and even specific readings on the dials, precisely *during* resuscitation. Their observations also included some uncommon medical practices that took place. Sabom concluded that several of the verified observations were such that they could only have been observed visually, beyond any potential hearing.

Greyson (2007) recalled another account that he investigated personally. During open-heart surgery, the patient later described watching the scene from above his body, noting that, during the time while his chest cavity was open, his surgeon began "flapping his arms as if trying to fly." When Greyson interviewed the surgeon, the latter explained his "peculiar habit." After washing and at other times during the surgery in order to keep his hands from any contamination, he often placed them against his chest and pointed out various things to his assistants, by using his elbows. The appearance was that of "flapping" his elbows up and down. The participating cardiologist who was also in the room confirmed that the patient had accurately reported these details just "shortly after he regained consciousness following the surgery" (Cook, Greyson, and Stevenson 1998).

The sorts of cases requested by Augustine and Potts, exhibiting accurate descriptions drawn from specifics in the NDEr's immediate vicinity, are hardly rare. Holden (2009) provides a list of more than 100 specific, evidenced NDE cases reported by many authors, the majority of which were of the NDEr's immediate surroundings. Many of these accounts are recorded by major researchers, such as eleven cases from Ring and Valarino (1998), ten cases from Cook, Greyson, and Stevenson (1998), the ten mentioned from Sabom (1982), seven from Fenwick and Fenwick (1995), and seven from Morse (1990, 1994). To these we might add others, such as three more cases from Morse (1992) along with a list from my own research, collected over a period of 35 years.

These examples appear to address clearly the specific evidential requirements of Potts and Augustine, with details being reported during surgery, where the best explanation is that the patient actually recalled events that were observed during that time, confirmed by someone actually present. The repeated testimony of patients that they were positioned above their bodies counts for something, too. The "arm-flapping" episode illustrates the many testimonies where the patient reported the information very quickly after regaining consciousness, which contributes to the overall veridicality of this conclusion. Altogether, just from the narrated accounts or sources mentioned above, the total stands at over *120 evidenced NDE cases from the immediate vicinity* of the patient.

#### 14.2.2 Evidential NDE corroboration at a distance

Other skeptics prefer different sorts of corroboration. Despite the requests from Augustine and Potts to produce very specific types of visual data *within* the emergency or operating room, or from the immediate vicinity of the NDEr, other skeptics like psychologist Susan Blackmore (1993) prefer visual information reported by the NDEr from a distance away, outside the room, that was corroborated later. Such data would be helpful in order to rule out the patient having learned the information from normal sense data drawn from the immediate proximity, even given the unlikelihood that such precise knowledge should be gained in that manner.

For Blackmore, two specific varieties of cases would be the most evidential, and have the most potential to disprove her naturalistic hypothesis: “distant vision” where the NDER could not have obtained the reported data from their presence in the resuscitation room, and accurate testimony from NDEs in blind persons. However, she attested that, search as she could, she was never able to locate good examples of either sort (Blackmore 1993, 125–133).

It should be agreed that verifiable cases like the ones Blackmore requested, especially those from beyond the NDER’s immediate line of physical sight or hearing, could definitely provide yet another highly evidential angle that could help to establish a more multifaceted NDE case. We will mention several such cases here. A few skeptics have criticized Kimberly Sharp Clark’s tennis shoe example, although subsequent research has buttressed her account (Holden 2016; Greyson 2016, personal communication). Blackmore (1993) is even positive here, pointing out that verified cases such as Sharp’s could constitute strong evidence for the disembodied NDE thesis. The naturalistic thesis, on the other hand, predicts that such information must be derived by normal means.

Besides Clark’s example, another case also involving a shoe found on a hospital roof was reported from all the way across the country, by Kenneth Ring and Madelaine Lawrence (1993). The resuscitated patient claimed to have experienced a NDE, floated above her body, and then watched the resuscitation attempt going on beneath her. Then she experienced being “pulled” through several floors of the hospital until she emerged near the building’s roof, where she viewed the Hartford, Connecticut, skyline. Looking down, she then observed a red shoe. When nurse Kathy Milne heard the story, she reported it to a resident physician, who mocked the account as a ridiculous tale. However, in order to ascertain even the possibility of an accurate report, he enlisted a janitor’s assistance, and was led onto the roof, where he found the red shoe just as it was reported! This occurred in 1985 and Milne was unfamiliar with Maria’s tennis shoe account, which was published just shortly before.

In the case of Kristle Merzlock, the young girl who nearly drowned and was resuscitated by Morse (1990), she reported more than the specifics of the resuscitation attempt and the sequential details from the emergency room. Upon regaining consciousness three days later, her intensive care nurses initially heard her recollection of having visited heaven, guided by an angel. Kristle also testified that, although she was unconscious and hooked up in the hospital, she was “allowed” to observe her parents and siblings some distance away, at home for the evening. She provided exact details regarding where each person was located in the house, identifying the specific things they were doing, as well as the type of clothes that they were wearing. For instance, she identified that her mother was cooking roast chicken and rice for dinner. All of these particulars were subsequently confirmed.

Certainly one of the most evidential, detailed distance cases involved a patient in Milwaukee who experienced a cardiac arrest that lasted 30 minutes. While in a coma, he watched the local medical proceedings from above his body. Then he realized that he was able to “move” to his home in Florida, over 1,200 miles away! Here he watched the person who was house-sitting for him and his wife while they were in Wisconsin. Coming out of the coma, he reported to his wife a sequence of many amazing, very specific, as well as out-of-the-ordinary details that he had observed in their home. These reports were confirmed later, both by discussions with the house-sitter as well as by their own



observations upon returning to Florida, including the unique events that the patient had narrated beforehand to his wife and house-sitter. In two other instances, the distances between the patient and their observed details were 20 and 30 miles away, respectively (Holden 2016)!

In addition to these examples, many other evidential NDE accounts have also included verification at a distance, with a number of them being substantiated by subsequent interviews. Greyson, Kelly, and Kelly (2009, 230) note from their NDE research that “60 people reported being aware of events occurring outside the range of their physical senses.” Greyson (2010) reported that a number of distance cases were researched and confirmed, including examples by Van Lommel *et al.* (2001) and Sartori, Badham, and Fenwick (2006).

Two journal articles by Cook, Greyson, and Stevenson (1998) and Kelly, Greyson, and Stevenson (2000) recorded a total of 15 more NDEs with observations viewed at a distance, each of which were subsequently confirmed (Long 2010). A number of documented distance NDEs also occurred in the absence of heart or brain activity (Holden 2016). To these, we could also add still more confirmed distance testimonies that were reported by Morse (1992), Sabom (1982), and Ford (1978). Just these accounts alone combine to *total more than 100 evidenced NDEs reported a distance away from the experiencer!* This testimony must be viewed as among the most convincing corroborated reports of all.

#### 14.2.3 *Evidential NDE corroboration concerning previously deceased persons*

Another exceptional category of data concerns evidential accounts where it was unknown that particular loved ones or friends had died recently, often somewhere else in the country or even the world, until they reportedly appeared in the NDE. Greyson (2010) notes three species of these experiences: (1) where the death of the deceased individual was previously unknown to the NDEr, who subsequently learned the relevant information. (2) On other occasions, the deceased person had died at the same time as the NDE experience, or “immediately beforehand,” thus precluding previous knowledge by those present. (3) The deceased was sometimes a person whom the NDEr did not even know.

Common to all three species (especially in this last category), data were purportedly learned from the deceased individual that was sometimes unknown to anyone beforehand, though it was confirmed subsequently (Kübler-Ross 1983; Habermas and Moreland 1998; Morse 1992; Alexander 2012). Over two dozen such examples were collected by Greyson (2010). Besides Greyson’s cases and the others cited above, we might add two other type 1 cases (Kübler-Ross 1983; Alexander 2012), along with six more type 2 cases (two by Kübler-Ross 1983; two by Morse 1992; along with others in this category reported by Van Lommel 2010; and John Myers 1971).

Regarding the three subtypes in this category, Greyson (2010) notes that since these accounts cannot be attributed to the NDEr’s expectations, subjective hallucinations should be ruled out. Further, Greyson asserts that “cases such as these provide some of the most persuasive evidence for the ontological reality of deceased spirits. Recent medical and societal advances in end-of-life care offer favorable opportunities for the further investigation of these cases” (2010, 169).

### 14.2.4 *Evidential corroboration from shared NDEs*

Still another type of corroboration is supplied by *simultaneous or shared NDEs* reported by at least two different people, including accounts from physicians or nurses, among others (Morse 1992, 1994; Houlberg 1992; Miller 2012). In one of Morse's cases, a boy died in a traffic accident while his deaf sister "observed" the entire accident process, though she was a distance away at home during the event (Morse 1992). She reportedly was even able to accompany her deceased brother to heaven, though she alone had to return while he remained there. Upon her "return," she brought back information that no one else knew, but was verified subsequently!

Moody (2010) described another shared NDE where five family members in Atlanta were present during the last moments of their relative's life. They were stunned as each of them simultaneously witnessed a bright light appear inside the room where they were all gathered, which then morphed into an entranceway through which they watched as their family member's image/spirit apparently left her body and appeared to enter joyfully! The family members agreed that the entry reminded them of Natural Bridge in Virginia's Shenandoah Valley!

In yet another example (Holden 2016), a woman who had experienced cardiac arrest for several minutes cited a string of corroborated items precisely *during* her arrested state, which is highly evidential, given her lack of heart and brain function during that time. But to compound the matter further, her husband shared the beginning of her NDE, witnessing her immaterial self actually rising above her body!

Another strange example from the author's NDE collection concerns two family members who were both in hospitals in different portions of the country. They reported simultaneous NDEs where they were with each other in the clouds! When they awoke, they each reported the experiences to those in the room around them.

### 14.2.5 *Evidential corroboration from blind NDErs*

Returning to the earlier challenge from Blackmore to produce NDE accounts from blind persons, yet another type of NDE evidence was produced by Ken Ring and Sharon Cooper (1997), who provided a detailed report of 31 blind NDErs. These cases produced several accurate testimonies, both from inside the room occupied by the patient, as well as outside. Even though there are fewer cases here so the evidence seems to be less than what other sorts may provide, it still would have been very difficult for many of the specific items to have been known previously by the patients, through any of their physical senses.

For example, during a NDE, one woman who had been blind from birth reported color images, including a rendezvous with two close friends from her youth, both of whom were also blind, and both of whom had died previously. She reported that two other deceased friends and a deceased relative were also present. She provided accurate physical descriptions of each one, even though she had never seen any of them before. In another case, a blind man correctly described the pattern of colors and designs on a neck tie that he received. Another man who also had been blind since birth described the scene outside of his home by providing many specific details, including snow on the ground and the appearance of a streetcar driving nearby.

Ring and Cooper conducted extensive interviews with these blind individuals, including attempting to track down their stories with others who were present. Their conclusion is

that “the blind persons in our study saw what they certainly could not possibly have seen physically. Our findings in this section only establish a putative case that these visions were factually accurate, and not just some kind of fabrication, reconstruction, lucky guess, or fantasy” (p. 124). Other authors have also included further data on NDEs in the blind (Miller 2012; Anderson 1980).

#### 14.2.6 *Additional evidential NDE corroboration*

Holden’s 2016 (edited) volume especially makes it exceptionally difficult for NDE naysayers, for this text contains more than 100 confirmed NDE reports, adding many noteworthy cases to our total! According to the text’s researchers, the major criterion for the NDE accounts included in this volume is that each episode *must* have been “directly confirmed by at least one other person” (p. xxvii). Such a strict order makes this text all the more valuable!

This new research contains 14 corroborated cases from the vicinity of the NDEr, 18 more accounts from a distance beyond the NDEr’s senses (sometimes *very* far away!), plus 36 additional reports from cardiac arrest patients. Add to these testimonies a few dozen other cases that include shared NDEs, meeting loved ones and friends in the ND state, along with learning previously unknown information during NDEs. Additional “compound” cases are also included in this text, where NDEs were accompanied by physical healings, the acquisition of messages unaccompanied by spoken words, or other evidential occurrences. A total of 66 accounts in this work were not included in our earlier tallies (Holden 2016, personal correspondence). Along with several additional cases not previously mentioned in this chapter, this would bring our total evidenced count from this chapter alone to *over 300 corroborated NDE cases* across a wide variety of types and subtypes!<sup>4</sup>

There is much variety among the 100+ corroborated cases in this last volume, too, including some of the most evidential cases on record. For examples, the NDE reports range from an NDEr experiencing cardiac arrest who peered down from the ceiling and observed a dated quarter perched on top of a tall medical machine underneath him, which was retrieved and verified. Another NDEr with cardiac arrest repeated correctly a long serial number on top of another tall medical device, and the number was recorded as she gave it. A few days later the medical unit was removed from the room. In the process, the exact number was confirmed, exactly as it had been given by the NDEr days before. Still another NDEr found a penny located up above the heads in the room.

Additional examples from this volume are simply a matter of choosing from among dozens of cases. For instance, a patient without heartbeat or brain activity for fully 20–25 minutes during hospital surgery correctly reported details concerning an anesthesiologist who came rushing back into the room after having left, as well as a collection of post-it notes attached to the physician’s monitor that had accumulated during the surgery, but were not present prior to the beginning of the operation.

In another case, after a severe heart attack left a patient in a state of clinical death, she reported being up above her body. While looking down, she correctly observed a number of items, including an IV bottle that accidentally smashed on the operating room floor, plus watching as an especially prized hair clip fell onto the floor, was stepped on, and broken. The patient even identified properly the particular physician who had accidentally broken the clip.

One more example involved a patient in a coma who had an NDE but was still able to watch her family members who were in the hospital cafeteria. Her father, a smoker, announced that he was going to light up a cigarette. Then she watched her two grandmothers assert that they were also going to smoke with him, even though, incredibly, one of them had never smoked in her entire life, and had always proclaimed loudly over the years that no one would *ever* see her with a cigarette! Yet the patient witnessed the incident correctly, as was confirmed to her just two weeks later.

Another aspect of the overall state of the evidence needs to be mentioned here, too. While we have seen many of the different directions and angles from which the NDE corroboration emerges, we dare not miss the fact that literally dozens of individual NDE cases presented above were confirmed by the presence of *more than one species or kind of evidence each*. The examples can be multiplied almost at will: two drowning girls and other examples without heartbeat while later providing complete *sequential* descriptions over a half hour or more; several cases of *very long-distance* observations from many miles away, at least three of which were also cardiac arrest cases, with one of these NDEs additionally being verified by a healthy person.

Moreover, a few cases of blind NDErs (some from birth) have emerged, where they “saw” items that were corroborated, before reverting back to their previous blind state. Then, many cardiac arrest patients still presented plentiful examples of accurately observed data, with at least one of these NDErs being reportedly detected by another healthy person; the corroboration in other cardiac arrest cases came *before* resuscitation ever began; while still other NDE testimonies came from occurrences *during* the state of cardiac arrest. Then we have the accounts where numbers on objects located above the heads of those in the room down below were accurately reported by cardiac arrest patients, and many other accounts of meeting deceased loved ones or friends whose deaths were unknown, but where accurate knowledge was also communicated. The prospect of explaining all these combinational accounts seems *especially* troublesome for NDE deniers.

### 14.3 Naturalistic Rejoinders

Of course, there also have been many attempts to argue for various natural alternate theses against such NDE testimonies, but these actually come up quite short in comparison to the overwhelming details drawn from the available hundreds of NDE evidential cases. Many of the reasons should be apparent here.

Miller mentions a large number of these natural efforts that seek to explain NDEs in completely normal terms (2012, 31–48, 105–108). Then he lists a wide variety of rejoinders directed at these natural efforts, containing almost 200 total pages of responses as researched by the most able NDE defenders today. In other words, while critics attempt to disprove these NDE reports as nothing more than subjective experiences, hundreds of pages of rejoinders and counter-critiques have been offered by the NDE researchers themselves. Further, it would not be at all difficult to amass dozens of pages more in refutation of the natural line-up, such as contained for example in the four-issue dialogue in the *Journal of Near-Death Studies* in 2007–2008 mentioned above.

At any rate, such debates are certainly not solved by page-counts of competing data! The real question is the ideological outcome of this clash. Which position make the best retorts

and established a clear advantage? Some questions will help to point the way through the major issues.

Did the events recounted in the ND reports occur *during* the NDE itself, or at some other time? It is undisputed that many of these corroborated ND reports clearly *concern* incidents that had occurred *during* the distressed medical states. But were every last one of these testimonies derived from naturally acquired information that was learned by some means *other than* via an actual NDE (perhaps through natural senses, or just prior to the resuscitation, or sometime *after* the particular occurrences took place, as Marsh concludes)?

Quite clearly, many of the *more than 120 evidential NDE accounts* mentioned above from inside the hospital room where the patient was positioned contained many observations that were both specific as well as quite often being rather mundane, as opposed to the attention-grabbing, easily remembered reports. These evidential and correctly reported accounts cover a very wide variety of stories: from readings on machine dials, to an instance of an unplugged medical device, to a “bird-like,” arm-flapping surgeon in the operating room, to the proper location of misplaced dentures, to a nurse wearing plaid shoelaces. Other times, the nature and wording of jokes told to relieve tension inside the room were recounted, or the reporting of embarrassing incidents that transpired in the room, while some NDEs were reportedly shared by other healthy persons in attendance, and so on. Together, a simply staggering array of hypothesizing would have to identify how these items were not observed precisely during the time sequence when the individual thought that they saw or heard them.

Further, studies such as Sabom’s (1982) included a control group and other related data, which indicated that some observations clearly seemed to emanate from the vantage point and angle of the ceiling, precisely as claimed by the NDEr. In other words, many NDE patients had no doubt that they were looking down at themselves from the position of being up above their body, rather than relating normal comments that reflected an eye-level report, and such a ceiling-level viewpoint was frequently what their report revealed. When that elevated angle also turns out to reveal corroborated information, the higher position has to be favored, most of all when some of the reported items could *only* be seen from up above, such as the reported numbers.

Could resuscitation attempts “wake” the brain just long enough for the patient to observe certain elements in the room by natural means? Medical experts have asserted that re-achieving consciousness does not occur in just seconds, but rather does not really return until after the heart has actually been restarted again (Holden 2016). Furthermore, in several cases (Holden 2016), no resuscitation efforts at all had even been administered to the patient prior to the NDE observation, clearly eliminating this particular natural option. Moreover, neither would this natural retort even apply to most of the NDE reports, such as those over a distance, or to healthy persons sharing the NDEs, or to seeing deceased friends or loved ones who impart unknown information, to those occurring to blind persons, in other words, in an entire host of evidential scenarios.

Did any NDE reports occur beyond the point of “brain death”? This question must be addressed at more than one level. Increasingly, the most evidential NDE cases are usually thought to occur especially when the corroboration is produced during a state of cardiac arrest due to ventricular fibrillation. In such a condition, the heart stops and at least higher (cortical) brain activity usually ceases in a matter of just 11 to 15 seconds or so afterwards (especially De Vries *et al.* 1998; also Van Lommel 2006; Van Lommel *et al.* 2001; Greyson

2015; Parnia and Fenwick 2002). Then even lower brain activity likewise ceases very shortly afterwards (Van Lommel 2006; Holden 2016).

At any rate, the lower-level brain activity that is present just very slightly longer than upper level cortical consciousness would be insufficient by itself to explain high-level consciousness anyway, including the clearest, most realistic experiences that NDErs have ever reported in their entire lives, not to mention any reported confirmation (Sabom 2006, personal communication). In particular, even fully operating brains still could not even explain *the more than 100 corroborative cases* from a distance outside an individual's line of physical sight that were mentioned or cited in this chapter, or many of the over *300 total evidenced NDE observations* mentioned here.

But when addressing the question of there being any NDEs beyond brain death, there are still the cases mentioned above, namely, those where the NDEr sometimes reports encountering a deceased individual where the meeting is also evidential in nature. Here is an additional, crucial sort of data, in that while the NDEr returned to normal life, the purportedly deceased person in the encounter had been dead for some time, often for many years (Greyson 2010; cf. E. W. Kelly 2001). If this is indeed the best explanation for these events, then these NDEs definitely extend *far* beyond the irreversible, biological death of the deceased individual's brain! This bypasses the issue of cardiac arrest altogether! Such confirmed cases along with many others simply increase the likelihood that at least some NDE accounts apparently report actual data from conscious states that most likely extend beyond the death of the physical body. So quite clearly, the question of whether any NDEs extend beyond brain death may be answered affirmatively!

We have indicated many sorts of natural comebacks to NDEs, yet they do not have to be treated individually. That's because by the very nature of the topic, NDEs have a powerful advantage over virtually all other supernatural-natural issues that are debated by scholars. Due to the nature of this subject, there is a theoretical line in the sand that is automatically built into this particular discussion. Virtually every alternative hypothesis, including the most prominent ones, postulate conditions that are *internal* to the individual NDEr, such as oxygen deprivation, temporal lobe seizure, drugs, exaggerations, or hallucinations. Two more specific examples that we have already mentioned include Blackmore's dying brain or Marsh's waking brain hypotheses. That is, they rely chiefly upon *interior* physiological conditions and/or psychological states of the NDErs' mind. For the most part, then, NDE critiques depend upon *subjective* conditions mostly *inside* individuals, while often attempting simply to deny outside corroboration. But there's the chief rub.

The central problem for the overall natural position, then, is that in this chapter alone we have so far described or listed the sources for the more than 300 cases of *external* circumstances which argue for the veracity of these NDEs. Many reports are drawn from the immediate or surrounding vicinity of the NDEr, while others are derived from a distance away, beyond the eyes and ears of everyone present, and most of all, beyond the senses of the NDEr herself. Some describe individual events and others involve a sequence. Some of the ND experiences are reportedly shared by healthy onlookers, and still others occur simultaneously. A few of the NDErs are blind, and others claim to have met deceased individuals who impart evidential information that they and/or others did not know. Lastly, many NDEs are confirmed by more than one sort of evidence.

Therefore, if even just a handful of key examples among the 300 evidential cases accurately reported data from the "real world," it would seem that the subjective theses attempting to explain away these experiences would then fail by a large margin to account

for all of the key aspects of the reports. This is precisely what the NDErs themselves have always claimed all along on the grounds of their personal experience alone!

Here's the key: no matter what subjective, internal states the critics wish to discuss, by their very nature they cannot explain the existence of objective, externally corroborated NDEs. Something that exists objectively – *out there* – cannot be refuted or denied by internal human issues.

In short, here is the single, major problem with all natural theses against NDEs: they all utterly fail to explain the evidential NDE cases. In order for the natural suppositions to work, virtually *every one of these 300+ corroborative accounts cited here alone must be mistaken!* But how likely is that, particularly given the careful scientific efforts to determine the accuracy of many dozens of them? But there must be no remainder from the 300 cases if the natural challenge is to succeed, for by their nature, *the internal cases are trumped by the external, evidential ones!*

Since a plethora of well-evidenced NDEs cover so many different angles and circumstances, “shotgun” natural explanations are sometimes suggested, too. Perhaps some of these events are due to interviews that are too old, or drawn from false memories, or the evidence may have been exaggerated, or even consisting of lies. These multiple suggestions could potentially explain some of the NDE data. But it is highly dubious and untenable that the more than 300 evidenced cases referred to in this chapter can be accounted for even by the totality of such suggestions. Many testimonies were shared or collected immediately upon the spot or at least very soon afterwards, and many of the specific details were confirmed by a variety of witnesses, in addition to other checks and balances. Many truly exceptional NDE instances were singled out and checked even more meticulously. These evidential cases need to be explained.

It would seem, then, that Augustine's (2007b) own personal concession might come into play here, namely, that the presence of these sorts of confirmed NDE data would both “contradict” his naturalistic thesis, as well as indicate the likelihood of an afterlife. We noted above that Blackmore (1993) and Potts (2002) also made similar comments if the NDE evidence were to obtain. This serves as a warning to those who continue closing the naturalistic door that opposes NDE data seemingly no matter how great the quality or quantity of the corroborative reports that continue to accumulate. This is especially the case when there emerges such a wide variety of different types of NDE reports, all coming from different angles, particularly when multiple sorts of evidence affirm many of the accounts.

Like other debates of this sort, the natural view is not established because the alternate hypotheses occupy the best evidential ground, by a long shot in this case. Rather, NDE opponents *must* take the lesser stance because it is all that remains! The naturalist must choose the internal position while claiming that there are no true external evidences among the 300 NDE testimonies, which appears to be a highly unlikely view. This precisely is the line in the sand over which the natural positions have tripped!

## 14.4 Conclusion

Having come this far, perhaps a passing reflection on the chief issue of this volume, the mind-body debate, would be appropriate. The available evidence for NDEs seems clearly to be more than enough to establish their evidential reality, but does this contribute any impetus to mind-body dialogues? Is one view favored more than another on this NDE

evidence, with regard to the most likely concept of human nature? This question has been discussed in the NDE literature, for instance, by philosopher Mark Woodhouse (1981), but without being settled one way or another.

Yet it seems safe to say that the most common impression among scholars, is that aspects of this research such as the NDEr's perception of leaving her body and looking down at it from above seems quite strongly to favor a dualist perspective about the self and consciousness. This would result, of course, from the manifestation of material and immaterial aspects of the self. That the NDEr identifies herself with the location of her consciousness up above, often not even at least initially recognizing the body below, furthers this notion.

While this seems to make the most sense on the data we have, some researchers have pointed out that remote viewing of the data by the stationary NDEr could favor more of a unified notion of the body. While there are some differences, the NDE notions of the material and immaterial self do provide some helpful empirical hints, and dualists are leading the discussions.

Another major aspect seems clearly to emerge from discussions like this, as well. It often appears that the real, underlying issue in these matters is very frequently *not* about straightforward dialogues regarding where the best evidence lies, but is more about a momentous clash between worldviews. Much like the most heated political struggles, it usually makes far more difference which position the debater already favored prior to the beginning of the discussion. If this is accurate, then it seems that even strong evidential considerations are less likely to change minds. In fact, once our minds are made up, it is often simply amazing what ludicrous responses are often preferred just to keep from entertaining the possibility of real change!

For example, the naturalistic worldview has shown many signs in recent years of having major foundational fissures. Yet, it appears that many naturalists would say or do about anything to keep from conceding a major plank in their platform, like allowing for an afterlife. Naturalists seem to prefer opposing any and all evidence that usually favors even just the initial moments of an afterlife, which concerns the vast majority of the accounts that have been addressed here (see Habermas and Moreland 1998, chs. 8–9 for more details, including critiques of some “non-naturalistic” options, as well). But ignoring the quickly mounting numbers of highly evidenced and corroborated NDE incidents, or simply responding with guffaws in order to avoid such data fail to refute the NDE argument!

The many dozens of examples mentioned or listed in this essay include evidential corroboration beyond the cessation of heartbeat, at least cortical brain waves, with lower brain cessation coming just slightly afterwards. Usually all three types of termination occur roughly in tandem. Then there is the presence of meticulous and sequential evidenced reports from the resuscitation room, as well as observed details from a distance away, clearly beyond the range of one's physical senses. Adding to the mix, further, are reported evidential meetings with deceased friends and loved ones, often with additional unknown information, a few cases from those who were blind, as well as other NDEs that appear to have been shared, witnessed, or corroborated by healthy onlookers, and so on. Clearly, so many multiple natural explanations would have to obtain in order to explain each of these cases, that the naturalistic thesis rapidly reaches the breaking point here.

Alternative rejoinders no longer seem to have truly made as many gains in the most recent conversations, including Marsh's own hypothesis of a waking brain (although it must be noted again that Marsh himself is clearly not a naturalist). In fact, it could be argued that



the number of medical NDE articles centering on alternative theses have largely become less plentiful in recent years, with the majority of essays clearly favoring the possibility, if not the likelihood, of the NDE data.

Further, we pointed out earlier that Marsh gives comparatively little attention to the corroborative accounts anyway, and when he does look at a few, they seem often to be the older and less-established examples. This is very intriguing, especially when both sides in this debate, including Marsh himself, concede the crucial importance of explaining these evidenced claims.

After all, these corroborated cases may be precisely the keys to refuting the natural approaches to NDEs. The bottom line, then, appears to be this: neither Marsh's waking brain physiology nor any other natural theses have successfully explained the at least 300 evidential cases narrated or listed here, especially the weightier sorts like those that have been identified above. They appear to pack far more punch than do their natural counterparts.

This is the meaning of the clear line in the sand mentioned above. It certainly seems as though the natural case possesses far less explanatory power and its critiques are far more ad hoc. It often seems that the anti-NDE position only hangs by two very weak and virtually irrelevant desires: to preserve its natural worldview commitments at all costs, no matter what, and its strong dislike of any "spiritual" options. But one thing is clear: the naturalist position does not do well when attempting to refute the hundreds of NDE evidential cases. Hence, the naturalist view is by far the weaker explanation here; it is not even close.<sup>5</sup>

## Notes

1. Such as pages 15, 18–19, 22–27, 31–32, 49–50, 52, 63, 124–26. Of course, Marsh could protest that his chief volume in 2010 predated many of the best-evidenced NDE cases. But his 2016 article also cited here was published *after* the majority of the best cases and his thesis still did not change substantially. Further, though not having read his chapter in this present volume, he still apparently agreed to oppose the presence of objective NDEs.
2. Some noncrucial details have been changed.
3. Many more precise details are reported in Holden (2016).
4. Thus, along with several cases drawn from the author's own research collection that have been added here and there, these totals are only the count that is mentioned or cited in this chapter, not a total from all NDE literature.
5. The author wishes to thank long-time friends and researchers Bruce Greyson, Mike Sabom, and Jan Holden, plus Jonathan Kopel, for their exceptionally helpful discussion, information, and bibliographic support during this research.

## References

- Alexander, Eben. 2012. *Proof of Heaven: A Neurosurgeon's Journey into the Afterlife*. New York: Simon & Schuster.
- Anderson, J. Kerby. 1980. *Life, Death, and Beyond*. Grand Rapids, MI: Zondervan.
- Augustine, Keith. 2007a. "Does Paranormal Perception Occur in Near-Death Experiences?" *Journal of Near-Death Studies*, 25: 203–236.
- Augustine, Keith. 2007b. "'Does Paranormal Perception Occur in Near-Death Experiences?' Defended." *Journal of Near-Death Studies*, 25: 261–283.

- Augustine, Keith. 2007c. "Near-Death Experiences with Hallucinatory Features." *Journal of Near-Death Studies*, 26: 3–31.
- Augustine, Keith. 2007d. "Psychophysiological and Cultural Correlates Undermining a Survivalist Interpretation of Near-Death Experiences." *Journal of Near-Death Studies*, 26: 89–125.
- Beauregard, Mario, and Denyse O'Leary. 2007. *The Spiritual Brain: A Neuroscientist's Case for the Existence of the Soul*. New York: Harper Collins.
- Blackmore, Susan. 1993. *Dying to Live: Near-Death Experiences*. Buffalo, NY: Prometheus.
- Cook, Emily Williams, Bruce Greyson, and Ian Stevenson. 1998. "Do Any Near-Death Experiences Provide Evidence for the Survival of Human Personality after Death? Relevant Features and Illustrative Case Reports." *Journal of Scientific Exploration*, 12: 399–400.
- De Vries, Jaap W., P. F. Bakker, G. H. Visser, J. C. Diephuis, and A. C. van Huffelen. 1998. "Changes in Cerebral Oxygen Uptake and Cerebral Electrical Activity During Defibrillation Threshold Testing." *Anesthesia and Analgesia*, 87: 16–20.
- Fenwick, Peter, and Elizabeth Fenwick. 1995. *The Truth in the Light: Investigations of Over 300 Near-Death Experiences*. London: Headline Book Publications.
- Ford, Marvin, with Dave Balsiger and Don Tanner. 1978. *On the Other Side*. Plainfield, NJ: Logos International.
- Greyson, Bruce. 2007. "Comments on 'Does Paranormal Perception Occur in Near-Death Experiences?'" *Journal of Near-Death Studies*, 25: 237–239.
- Greyson, Bruce. 2010. "Seeing Dead People Not Known to Have Died: 'Peak in Darien' Experiences." *Anthropology and Humanism*, 35: 159–171.
- Greyson, Bruce. 2015. "Western Scientific Approaches to Near-Death Experiences." *Humanities*, 4: 775–796.
- Greyson, B., E. W. Kelly, and E. F. Kelly. 2009. "Explanatory Models for Near-Death Experiences." In *The Handbook of Near-Death Experiences: Thirty Years of Investigation*, edited by Janice Miner Holden, Bruce Greyson, and Debbie James, 213–234. Santa Barbara, CA: Praeger.
- Habermas, GARY R., and J. P. Moreland. 1998. *Beyond Death: Exploring the Evidence for Immortality*. Wheaton, IL: Crossway.
- Holden, Janice Miner. 2009. "Veridical Perception in Near-Death Experiences." In *The Handbook of Near-Death Experiences*, edited by Janice Holden, Bruce Greyson and Debbie James, 185–212. Santa Barbara, CA: Praeger.
- Holden, Janice Miner, ed. 2016. *The Self Does Not Die: Verified Paranormal Phenomena from Near-Death Experiences*. Durham, NC: International Association for Near-Death Studies (IANDS).
- Houlberg, L. 1992. "Coming Out of the Dark." *Nursing* (February).
- Kelly, Edward F. 2010. Review of Michael N. Marsh's *Out-of-Body and Near-Death Experiences: Brain-State Phenomena or Glimpses of Immortality?* *Journal of Scientific Exploration* 24: 729–737.
- Kelly, Emily Williams. 2001. "Near-Death Experiences with Reports of Meeting Deceased People." *Death Studies*, 25: 229–249.
- Kelly, E. W., B. Greyson, and I. Stevenson. 2000. "Can Experiences Near Death Furnish Evidence of Life After Death?" *Omega*, 40: 513–519.
- Kübler-Ross, Elisabeth. 1983. *On Children and Death*. New York: Macmillan.
- Long, Jeffrey, with Paul Perry. 2010. *Evidence of the Afterlife: The Science of Near-Death Experiences*. New York: Harper Collins.
- Marsh, Michael N. 2010. *Out-of-Body and Near-Death Experiences: Brain-State Phenomena or Glimpses of Immortality?* Oxford: Oxford University Press.
- Marsh, Michael N. 2016. "The Near-Death Experience: A Reality Check?" *Humanities*, 5: 1–25.
- Miller, J. Steve. 2012. *Near-Death Experiences: As Evidence for the Existence of God and Heaven*. Acworth, GA: Wisdom Creek.
- Moody, Raymond, with Paul Perry. 2010. *Glimpses of Eternity: Sharing a Loved One's Passage from this Life to the Next*. New York: Guideposts.

- Morse, Melvin, with Paul Perry. 1990. *Closer to the Light: Learning from Children's Near-Death Experiences*. New York: Random House.
- Morse, Melvin, with Paul Perry. 1992. *Transformed by the Light: The Powerful Effect of Near-Death Experiences on People's Lives*. New York: Random House.
- Morse, Melvin. 1994. "Near Death Experiences and Death-Related Visions in Children: Implications for the Clinician." *Current Problems in Pediatrics*, 24(February): 55–83.
- Myers, John, ed. 1971. *Voices from the Edge of Eternity*. Old Tappan, NJ: Fleming H. Revell.
- Parnia, Sam, and Peter Fenwick. 2002. "Near Death Experiences in Cardiac Arrest: Visions of a Dying Brain or Visions of a New Science of Consciousness." *Resuscitation*, 52: 5–11.
- Potts, Michael. 2002. "The Evidential Value of Near-Death Experiences for Belief in Life after Death." *Journal of Near-Death Studies*, 20: 233–258.
- Ring, Kenneth, and Sharon Cooper. 1997. "Near-Death and Out-of-Body Experiences in the Blind: A Study of Apparent Eyeless Sight." *Journal of Near-Death Studies*, 16: 101–147.
- Ring, Kenneth, and Madelaine Lawrence. 1993. "Further Evidence for Veridical Perception During Near-Death Experiences." *Journal of Near-Death Studies*, 11: 223–229.
- Ring, Kenneth, and Evelyn Elsaesser Valarino. 1998. *Lessons from the Light: What We Can Learn from the Near-Death Experiences*. New York: Insight Books.
- Sabom, Michael B. 1982. *Recollections of Death: A Medical Investigation*. New York: Harper & Row.
- Sartori, P., P. Badham, and P. Fenwick. 2006. "A Prospectively Studied Near-Death Experience with Corroborated Out-of-Body Perceptions and Unexplained Healing." *Journal of Near-Death Studies*, 25: 69–84.
- Sharp, Kimberly Clark. 2007. "The Other Shoe Drops: Commentary on 'Does Paranormal Perception Occur in Near-Death Experiences?'" *Journal of Near-Death Studies*, 25: 245–250.
- Van Lommel, Pim. 2006. "Near-Death Experience, Consciousness, and the Brain: A New Concept about the Continuity of our Consciousness Based on Recent Scientific Research on Near-Death Experience in Survivors of Cardiac Arrest." *World Futures*, 62: 134–151.
- Van Lommel, Pim. 2010. *Consciousness Beyond Life: The Science of the Near-Death Experience*. New York: Harper One.
- Van Lommel, Pim, R. van Wees, V. Meyers, and I. Efferich. 2001. "Near-Death Experience in Survivors of Cardiac Arrest: A Prospective Study in the Netherlands." *The Lancet*, 358: 2039–2045.
- Woodhouse, Mark. 1981. "Near-Death Experiences and the Mind-Body Problem." *Anabiosis: The Journal for Near-Death Studies*, 1: 57–65.

# The Phenomenology of Near-Death and Out-of-Body Experiences: No Heavenly Excursion for “Soul”

MICHAEL N. MARSH

The mind-body problem, of the relationship between brain with consciousness and mentality, of dualism versus monism, has occupied the thoughts of many through time. More recently, neurophysiology has probed the quest for material understandings not only of consciousness as the presumed basis of mind, but how, from a lump of stuff – the brain, functionally primed by neurochemical transmission and electrophysiological impulse – arises consciousness, or more properly, the state of being conscious.

Consciousness is susceptible to varied extrinsic and intrinsic influences. One is our capacity to render ourselves “un-conscious” to worldly activity in falling asleep thereby entering other realms of cerebral activity. During the six or eight nightly occurrences of REM-induced sleep, we dream (Hobson 2002). Dreaming is vivid and may be enlightening or extremely frightening, and is not associated with immediate bodily effects since our musculature is physiologically paralyzed. State-switching between wakeful alertness and sleep is brought about by rapid changes in neurochemical signaling. Dreams are never remembered, except those occurring at points of awakening – termed hypnopompic, and which have particular relevance for near-death and out-of-body experiences (ND/OBE). The latter often result from depressed levels of consciousness. It is advanced here that ND/OBE, likewise, are reawakening phenomena, based on the important premise that the experiences undergone are capable of recall, therefore demanding elaboration of new memory.

These events have gained enormously in popularity over recent years, although much of this folklore is anecdotal. The problem lies in the standardized stereotype implanted in the common mind by Moody (1976) who simplistically incorporated every element recalled to

him during interviews with ND/OBE subjects into an idealistic, imaginary account: darkness, a tunnel, light, divine figures, beautiful terrain, followed by abrupt returns to earth – at the point of regained conscious-wakefulness. Proponents interpret them to exemplify “consciousness,” “mind,” or “soul” traveling not only without a brain but external to the body, thereby gaining access to a “somewhere” supposedly beyond the physical world, offering “proof” of afterlife and the existence of quasi-religious beings.

I examine certain claims made for ND/OBE in this chapter, adding additional neurophysiological and theological insights. My overriding conclusion is that they are brain-generated phenomena arising during the period when subjects are regaining full conscious-awareness. Hence there are analogies with hypnopompic dream-awakenings. That is, ND/OBE are decidedly this-worldly events and have nothing to do with supposed journeys to spiritualized or nonphysical realms, nor amalgamations with so-called Cosmic Consciousness.

In exploring this topic (Marsh 2010), I have surveyed around 700–800 reports extracted from the basic canonical literature of five authors (Moody 1976; Ring 1980; Sabom 1982; Grey 1985; Fenwick and Fenwick 1998).

## 15.1 Evaluating the Claims Made for ND/OBE

### 15.1.1 *Hellish experiences*

Readers may be surprised by this initial departure but, in my view, these provide evidence of brain-based phenomenology. Curiously, such events are frequently veiled since they jar with preexisting concepts of attractively beautiful terrains: but they should not be.

Many accounts describe intense darkness, gloom, feelings of falling backward, of being propelled or sucked downwards, accompanied by screaming or other unintelligible noises, frightening creatures, red colors, overriding fear, panic, anguish, and desolation, followed then by either a pleasant NDE, or rapid return to conscious-awareness. Thus: “I was above my body and [below] a black pit with hands trying to grab me and pull me down into it. A lion came out and jumped on me” (Grey 63). The lion must, of course, be hallucinatory, originating from preexisting brain-based memories. Nevertheless, the perception of bodily manipulation by “hands” suggests an awareness of being handled by those aiding this woman during her faint.

Similar rough-handling was described by neurosurgeon Dr. Eben Alexander, hospitalized with *E.coli* meningitis (Alexander 2012), and US art critic Howard Storm, admitted to a Parisian hospital with peritonitis (Wilson 1997, 21). He was attacked by horrible little people who poked and threatened him verbally – in English. He held them off by praying, but this is no escape.

Certain generalizations arise from these terrifying recollections. First, why do some experiences involve terrible things and then proceed to happier occurrences – as with Alexander and Storm?

Second, darkness probably implies incipient returns to conscious-awareness, a view strengthened by my perception that certain subjects were either *infected* (Alexander) or in *toxic delirium* (Storm), indicating superficial levels of “unconsciousness.” Indeed, it is relevant (though rarely noticed) that they all felt being “plucked at,” reflecting knowledge of handling by carers during turning, cleaning, or clinical examination. This raises another important insight: that the precipitating causes of NDE may, by impacting different brain centers, both condition the level to which consciousness is depressed and the typologies undergone.

Third, what geo-cosmic locations are being sampled during these black episodes before the pleasanter NDE ensues? And do their “geo-coordinates” change in accommodating the latter? There is more to these happenings than the merely psycho-phenomenological “*failure to let go of one’s ego at death*” – which seems a rather limp explanation (Greyson and Bush 1992; Ring 1994).

Finally, neurosurgically based forms of brain pathology have been extensively detailed (Solms 2000). These influence REM dream-state modes, in addition to initiating other pathological or operative outcomes entailing horrific experiences. Drugs also influence cerebral function to produce fearful hallucinatory outcomes. Hellish NDE are therefore strongly consistent with a neurophysiological explanation, although there are no relevant scientific data currently available for reference.

### *15.1.2 History, world geography, and cultural influence*

ND reports are subject to historical influence, as recorded most spectacularly by the Venerable Bede (ca. 673–ca. 735) concerning the nobleman Drychthelm. He was delirious from pneumococcal pneumonia (see section above for relevant comments). In this subconscious state (termed in pre-antibiotic times a “vigil coma”), he was led by a white-clad man who showed him a medieval picture (ca. 750 CE) of “hell” where he was struck at by horrible vermin-like creatures and where “souls” were being alternately tossed between great fires and freezing cauldrons. Walking onwards they encountered “Heaven” inhabited by happy people, music and beautifully scented vistas (Bede 1968). Having emerged from his stupor, much to the horror of those keeping bedside vigil, he sold up his possessions, became a monk and did daily penances for fear of reentering hell. Karl Becker (1981) recalls similar scenes from ancient China.

In India, a man was hauled up before the Hindu god of death: before being seen, his legs had to be amputated. But then it was discovered the wrong man had been called, so having retrieved his legs from a cupboard (holding a cache of previously removed legs), he simply returned home. Alternatively, for modern Polynesians, “Heaven” comprises boats, cars, and other such commodities (Counts 1983). Those historic and geographic portrayals contrast sharply with the anodyne, soft loci of beautiful vistas, lovely flowers, and scented environments apparently sampled mainly by modern Americans and Britons. While such culturally determined reports clearly reflect brain-imprinted recall and experience, the critical question is whether these experiences could be deemed revelatory in some way. For example, Jesus used contemporary imagery to facilitate and punch home the point of his message. The corpus of NDE recollection, however, seems to lack a similarly intrinsic quality or value: we gain nothing particularly new, novel, or inspiring. That is, no message of ultimate relevance or importance ever seems to be communicated, as I show below.

### *15.1.3 Idiosyncracies, bizarreness, and banality*

Next, I consider the semantics of recalled ND/OBE experience. This is important simply because no authors hitherto have ever critically analyzed the verbal accounts offered: from such interrogations, what do we discover?

Following a heart attack, an ND/OBE subject meets his father who predeceased him by 15 years, although “dressed just like he used to be in grey trousers and a cardigan. He hadn’t changed a bit. We chatted quite naturally [about family members including his brother] and

he joked” (Grey 1985, 79). The obvious questions arising are firstly, during those 15 years, did his father ever undress and have a shower? And while showering, were his old gray trousers and cardigan laundered? Another woman, after a difficult Caesarian section, agonized: “I’m not staying here – [my husband] can’t cope with the new baby, he’s just got a new job and I’ve left a pile of shirts to be ironed and he doesn’t know how to do them” (Fenwick and Fenwick 1998, 80).

These are extraordinarily bizarre accounts yet offered as serious contributions to NDE literature. If my assertion is credible, then we cannot dismiss them, nor fail to notice their geophysical, anthropomorphic content offering dream-like themes of extreme banality, devoid of intelligible meaning. On these grounds, I would argue that NDE reporting is not corroborative of biblical testimony. The analysis I give offers further telling arguments against any serious import to NDE phenomenology, other than deriving from stored material in brains of subjects reporting.

#### 15.1.4 *Descriptions of Jesus and the geography of Heaven: true insights – or not?*

Similar criticisms apply, more specifically, to so-called viewings of Jesus. Do they give us the crucial fill-in detail lacking in the New Testament documents?

“I can see that form now – it had blonde-gold hair . . . and a beard – very light, and a moustache [and] a white garment on – there was a red spot here [points to Sacred Heart on gown] – and a chalice in his hand” (Ring 1980, 60). Or: “He was tall – had a white robe on – his face was beautiful – his skin was glowing and absolutely flawless” (Sabom 1982, 49). And “She [refers to NDE subject’s mother] was wearing a long sparkling silver gown – [as] did Christ – he had long hair – long beard” (Sabom 1982, 169). And: “I saw Christ. He was incredibly beautiful . . . his feet bare . . . hair down to his shoulders and a beard. There was light coming out of his head like a star” (Fenwick and Fenwick 1998, 86). Since Jesus was of Sephardic origins, he was probably of dark countenance with black hair.

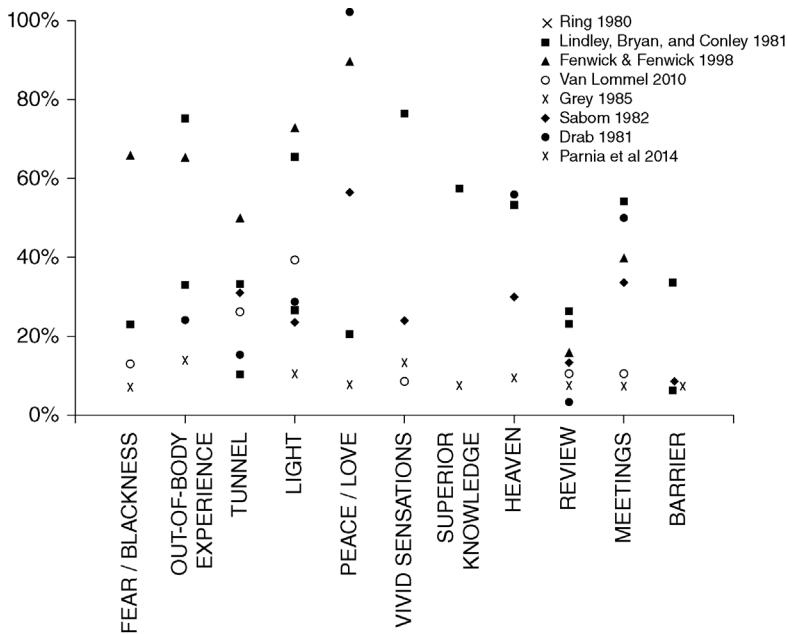
Next, consider the descriptive geography of the “otherworld.” For Sabom’s subjects (Sabom 1982, table XII) we get blue skies, clouds, parkland, mist, water, the Sea of Galilee, a road, a fence, a long corridor, a mountain top, golden gates, and geographic peoples undertaking cultural handicrafts. From Van Lommel’s (2010) and Long’s (2010) correspondents we encounter blue skies, rolling hills, flowers, colored terrains, music, distant city vistas, and so on. This content is culturally determined, failing to give useful insights into the supposed otherworldly realm. We derive absolutely nothing new or revelatory, but only a humdrum miscellany derivative of cinema, the media art galleries, Hollywood, and so on. Unfortunately, nothing retold is novel, or truly “transcendent”: the recollections, like dreams, lack sensible meaning or value in conveying serious insights about the afterlife.

In summary, we can arrive at some generalized conclusions. First, NDE accounts are nonidentical, nonuniform, personally idiosyncratic, evocative of dream-like bizarreness and banality: Ring’s “core” (1980) experience is totally inappropriate – there are alternatives (Table 15.1). Such widespread idiosyncrasies are reflected, for instance, in several canonical writings (see Marsh 2010) concerning the perceived “Light”: 16 percent (Ring 1980); 39 percent (Grey 1985; Lindley, Bryan, and Conley 1981); 28 percent (Sabom 1982); and 72 percent (Fenwick and Fenwick 1998). Those are vast differences. We should ask why this “other-world” is so *obstinately unequal* (Figure 15.1).

**Table 15.1** Near-death experiences: early and late. I have dismissed the so-called “core” trajectory of ND/OBE as not being truly representative of subjects’ recall, being more an artifact contrived by those reporting these events. That discrepancy is obviously revealed by their data displayed in Figure 15.1. A reclassification of ND/OBE in better reflecting what subjects actually recall is briefly summarized and classified here into early and late phase experiences. Seen in this more enlightened manner, the table honestly represents the key features of ND/OBE, thereby bypassing the “core” accounts which have become unfortunately ingrained in the public mind.

<i>Early</i>	<i>Late</i>
No pain	Increased pain perception
Ability to traverse physical objects	Hearing real voices
Absence of gravity – Floating/weightlessness	“Moral” imperative to assume earthly responsibilities
Sense of motion – Acceleration/gyration	Unable/unwilling to cross physical barriers
Seeing light and people	Bump return to body – weight/pain
No “moral qualms” when leaving family	

Source: Marsh 2010



**Figure 15.1** This scattergram plots the percentage (vertical axis) of ca. 800 subjects undergoing ND/OBE. This is the first occasion on which these results have been compositely assembled. As far as possible, data are extracted and forged into major experiential categories (horizontal axis). Contrary to expectation, the data reveal no obvious correlation, symmetry, or uniformity: the overall preponderance of experiences is < 55%. It is evident these cumulative data fail to underpin the smoothed accounts of ND/OBE phenomenology given in most publications. Indeed, any naive observer confronted with these data for the first time would neither conclude that they represented a reproducible “core” experience, nor that they provided a uniform account by each writer.



Viewed from earthbound perspectives, life is clearly unequal in the distribution of abilities, wealth or freedoms. Yet if NDE had anything to do with “spiritualized” outcomes, especially pertinent to the Kingdom of God, would we not reasonably expect some better form of equality? John Donne nicely expressed this ultimate desire (Sermon VIII), as “a house of God . . . with one equall Possession, one equall Communion and Identitie, [and] one equall Eternity” (Donne 2004).

Second, NDE accounts are historically, geographically, and culturally determined. Third, two or more experiences undergone by the same person are rarely identical. That is another grave problem surrounding experts’ assertions that NDE are truly otherworldly events. Fourth, the inconsistencies in reports about “Heaven” and “Jesus.” These are serious objections to the reality of ND/OBE, pointing instead to cerebrally induced illusion.

### 15.1.5 *Making sense of the phenomenal*

Despite the heterogeneous nature of individual reports, as briefly outlined above, there was one feature which stood out as uniform. This key insight came with my realization that NDE terminate as conscious-awareness is resumed: “[I came] rushing through a tunnel and snapped back into my body.” “When they put the shocks [electrical defibrillation] on me I fell back, down to my body, like a dead weight. And the next thing, I was in my body again” (Marsh 2010, 71–79). These remarkably abrupt endings into full conscious-awareness were succinctly put by Ring (1980, 101): “Once the individual has returned, painfully or otherwise to his body, the . . . experience is over . . . but . . . how exactly does one accomplish this return? – that is, re-unite with one’s body?”

However, the critical outcome is what *actually* needs to be reunited with the fleshly body, since it follows that if NDE are reawakening phenomena following periods of depressed consciousness, nothing whatsoever needs reuniting with the body. This additional speculative baggage merely confuses the issue. Readers should note that Ring fails to note whether consciousness, mind, or soul is reunited with its body: nor, in quasi-Platonic manner, how the former “recognizes” the latter for correct reunification.

I therefore firmly conclude that: (1) NDE terminate when each subject regains conscious-awareness; (2) with the use of word counts (as employed by dream researchers), it seems that most NDEs last only several seconds or minutes of real time, and, that the experiences undergone must therefore occur before, *and immediately up to*, that point at which conscious-awareness is fully regained: there is no escape from that conclusion; (3) this seems to be a sensible conclusion, because memories can only be set down when the brain is not in a moribund state, despite insistent contrary claims by so many other authors. OBE/NDE could best be envisaged as occurring, and importantly remembered, during those final moments when subjects are regaining conscious-awareness. It must, however, be acknowledged that claims of veridical perceptions by subjects supposedly “brain dead” exist, involving sightings of relatives in other parts of the hospital, or the use of specific clinical instruments. Such sightings are of critical importance in substantiating the view that NDE are valid pointers to otherworldly presence and activity. The underlying problem, and never tightly addressed by the NDE community, is whether these subjects were always in a continuous state of depressed consciousness – or not. Unfortunately, we have no neuro-physiological data on monitoring the levels of consciousness during NDE episodes, and thus on whether what is reported as veridical by a “dead” brain could, in fact, have actually been observed during a moment of transient conscious-awareness. But NDE are ephemeral

events, so that all the apparatus and dispassionate observers needed to make such a critical undertaking could hardly be mustered at the appropriate time. One could also conclude that the small handful of cases where veridicality of this sort has specifically been claimed becomes of scant significance relative to the many hundreds of thousands of people who never had such specific memories. Notwithstanding, my overall claim (4), as things stand at present, is that NDE/OBE are brain-generated, waking phenomena requiring memory for later recall, and are neither trips to Heaven, nor other cosmic spaces.

There are additional, supportive empirical data firmly substantiating my conclusions.

First, suicide survivors from the Golden Gate Bridge (Rosen 1975; Seiden 1968) experienced transcendent episodes during or immediately after their jumps. It takes about 4 seconds to fall 250 feet from the walkway, resulting in an impact speed into the water of about 70 mph (Snyder and Snow 1967): a few more moments are required for retrieval from the Bay. Some survivors could not remember hitting the water, so were probably temporarily rendered unconscious around the time of impact.

Next, consider examples from self-induced fainting attacks. During these laboratory-timed episodes ranging from 8 to 16 seconds, 60 percent of subjects (25 of 42) recalled visual and auditory adventures (Lempert, Bauer, and Schmidt 1994; Howard, Leathart, and Dornhorst 1951). The visual aspects ranged from darkness, grayness, colored variegated patches to fully formed visual situations with persons and things. Likewise, auditory experiences ranged from uninterpretable noises, swishing, rushing sounds, or screams, to human speech. Sixteen percent of those who fainted also had OBEs, while 8 percent went through tunnels. Tunnels, most clearly, are not the exclusive preserve of NDE: nor a common feature (about 10%) according to Drab (1981) – despite many contrary claims.

Finally, there are the reports from military pilots centrifuged to unconsciousness (Forster and Whinnery 1988). During a measured period not exceeding 40 seconds, these pilots underwent transcendental experiences (similar to recoveries from other preceding forms of unconsciousness) of beautiful, vividly colored surroundings, seeing deceased relatives, being deliriously happy, and undergoing such highly emotive feelings that they were reluctant to “return” to earth. For example, a pilot said: “I was floating in a blue ocean, on my back – the sun was up – and someone was trying to wake me. I did not want to wake up – I could see myself on the water and look at the sun.”

Authors involved with these pilots commented on “how such coherent visual illusions could be generated and remembered by the brain, within such short periods of time”: that is, as their subjects recovered from unconsciousness – just as happens during NDE. Wonderful what the human brain can do (Forster and Whinnery 1988).

## 15.2 Neuro-Physiological/Pathological Insights

### 15.2.1 *Somatic reference in ND/OBE*

The emphasis in the published accounts is that subjects are out of their bodies. If that were really true, then the irrelevance of corporeality should be strikingly obvious. But in many published accounts, dependency on flesh goes unnoticed. We saw one example above with the deceased father.

Accompanying this attempted suicide, a firm sense of self is evident: “I remember that I could see myself walking away. I was . . . 20 feet away . . . I could see me walking away. I was wearing this gray suit that I bought last year and I was walking away from myself

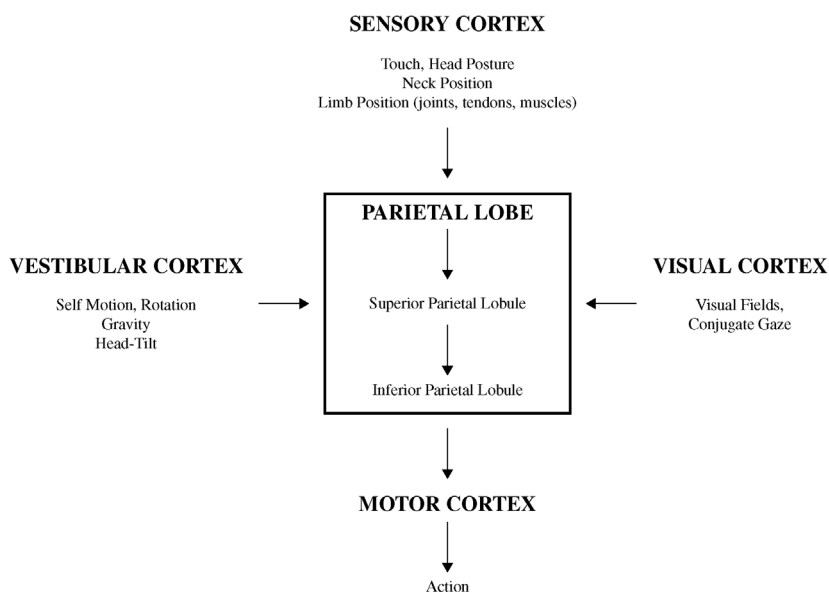
hanging there” (Ring 1980, 46). This suicide victim was *not* wearing a gray suit, revealing that what is reported is not necessarily a true account.

Another physically based example: “I had this piece of clothing on . . . very loose . . . and I remember having bare feet . . . it was very different . . . very thin, very delicate, very light. My face and hands were the same because I remember trying to touch my face to make sure everything was OK . . . [and] . . . I could feel it” (Ring 1980, 52). What is so interesting here is that the subject mentions her feet, which is most unusual. I have often wondered what subjects thought they had on their feet, and whether the otherworldly abode is boarded, paved, or carpeted. Such self-directed bodily reference points to the involvement of the temporo-parietal junction (TPJ) as the coordinating center for such outcomes (Ionta *et al.* 2011).

### 15.2.2 The vestibular apparatus and OBE

The vestibular apparatus is a large, brain-based system contributing to our sense of body image (egocentric) and its relationship to the immediate environment (peri-personal space), helped by additional incoming information from movements of the eyes, head, neck, skin, and joints (Figure 15.2).

All this information is received into a higher-level “multimodal” area within the temporo-parietal junction (TPJ) of the right cerebral hemisphere (Andersen 1997; Brandt and Dieterich 1999; Lackner 1988). Vestibular components comprise structures within the inner ear, including the semi-circular canals which orientate us spatially, and the saccule and utricle which sense gravity and alterations in bodily motion. The utricle and saccule are bony cavities containing large numbers of hairs bearing a crystal at their ends, the so-called “otoliths” (Mittelstaedt 1991), which bend in response to the strength and direction of



**Figure 15.2** This simple block diagram illustrates the main players contributing to the sense of egocentric and paracentric space. All incoming information from these three sources converges in the posterior parietal lobe where the “supramodal” perception of body image is created. For simplicity, the additional roles played by cerebellum, limbic system, hippocampus, and memory are excluded.

external environmental forces imposed upon the body. Their movements are traduced as signals into nerve endings which ultimately unite as the vestibular component of cranial nerve VIII. Their normal functioning, for example, allows us to continue walking upright in the dark while holding a cup of coffee without spillage. Their roles may also be perceived in the illusory continuation of the motion experienced after coming off a fairground round-about, or disembarking a ship which has just weathered heavy seas.

Much NDE phenomenology – hellish and heavenly – involves a great deal of hallucinatory motion: flying, ascending, floating, accelerating, spinning, falling, weightlessness, and bumping back into one’s body after episodes of weightlessness.

It is also noticeable, in particular, that OBE invariably occur when subjects are recumbent, due to reduced positional (horizontal) sensitivity of the utricle to gravitational force. The common illusion of floating beneath the ceiling during an OBE represents an anomalous projection of bodily (conscious) location, but only within para-centric space and never extending, it should be noticed, miles away to the horizon (Marsh 2013). This is the most extreme occasion when the illusion of self is being projected into space, in comparison with autoscapy and heautoscapy. The former is an inert recognition of one’s double (mirror image), while the latter is more intriguing, since consciousness then oscillates between the subject and the illusory double. Other aspects of duality of conscious experience are explored in the ensuing section.

### 15.2.3 *Duality of conscious-awareness during ND/OBE*

Next, and with relevance to this book, it is important to pay attention to many occurrences of a duality affecting our conscious-awareness. This is another very relevant factor determining how we view the phenomenology of ND/OBE. These depict occasions when *otherworldly* experiences are welded onto *this-worldly* events. How could a supposedly “dead” person, having undergone a medical crisis and thence inhabiting some kind of otherworld – simultaneously experience the commerce of this-worldly action? Since each is unlikely to exist within separate domains, the implication clearly reveals a bodily response to both aspects of the dual experience. These include hearing voices at the end of the bed, the placing of an oxygen mask over the face, being “poked” at by verminous creatures (see Section 15.1.1), feeling the searing burn after application of the defibrillating electrodes, or the painful stab of an injection – all interrupting the peaceful world of the NDE.

This form of dual conscious experiencing is not, of course, exclusive to NDE. It occurs with intrinsic brain abnormalities, or following external stimuli. These include lucid dreaming, temporal lobe auras preceding fits, licit and illicit drug-taking, or following direct electrical stimulation of the exposed brain (by Penfield, and by Blanke). Similar experiences of double threads of conscious experience occur in migraine sufferers. One lady could observe herself: “It was as if I was in another dimension – there was ‘I’ and there was ‘me’” (Lippman 1953). Another case by the same author described a woman who, during her pre-headache aura, felt that her phantom body was real, her actual body being regarded as the more illusory. In another case, the migraineur’s elevated body enjoyed a panoramic overview of her earthly activities, during which, “time was suspended.” These types of migrainous aura, associated in part to changes in cerebral blood perfusion, are truly reminiscent of OBE, reproducing several common phenomenological aspects. Even our daily lives are filled with illusory occurrences, but most of these are easily distinguished (Marsh 2015).

#### 15.2.4 *Neurological clarifications of ND/OBE phenomenology*

In drawing the various strands of this section together (somatic reference, disturbed vestibular locating of consciousness away from the physical body, and other forms of dual conscious-awareness), it is very tempting to suggest a close relationship with disturbed or aberrant neurological function, which would also include OBE (Blanke and Mohr 2005). In addition, computerized methodology (Ehrsson 2007) is allowing subjects to perceive themselves in near, para-centric space, so delivering further understandings about the many spatial illusions to which the human body, and its complex “vestibular system,” is subject (Marsh 2015). In addition, those prone to OBE appear to have a deficit in posterior hippocampal processing of the first person perspectival relationship between body and peri-personal space (Bergouignan, Nyberg, and Ehrsson 2014). However, it might still be objected that OBE subjects have recalled being aware of true events while undergoing an OBE (details of their resuscitations, or recognizing themselves being dealt with by clinical staff, and so on). The obstinate question is whether these subjects were either briefly conscious or, as noted in other recorded NDE events, were actually undergoing some form of duality of conscious experience (like epileptics or migraine sufferers). It must be remembered that many of these reports derive from cardiac arrest patients. The critical nature of these life-saving resuscitations means that clinical staff cannot sit back and make objective, scientifically valid observations on the conscious level of these patients throughout the resuscitation.

Some attempts have been made with computerized machines to analyze patients' EEGs during general anesthesia. But the results are not promising, and sometimes fail to provide good monitoring of the actual level to which consciousness has been depressed during operative surgery. Therefore, by the same token, it is not really possible to accurately measure varying levels of conscious-awareness in ND/OBE subjects, so that the validity of reports on these specific issues must remain unproven. Here, as another recent example, we should consider the long-awaited results of the so-called AWARE (*awareness during resuscitation*) project (Parnia *et al.* 2014). Although involving 2,060 cardiac arrests, only 9 (0.5%) *retrospectively* recollected experiences undergone during their resuscitations. These results are extremely disappointing, failing signally to produce the data originally anticipated. That is, they do not tell us whether the recollections actually occurred when subjects' brains were totally inactive. Nor do they show us, that if they were outside their bodies, patients would have reported on pictures placed beneath ward ceilings. Although “intrigued” by consciousness occurring during cardiac arrest, Parnia neither produced neurophysiological evidence for “brain death” at the crucial time nor addressed the most critical concern that flat-lined, “dead” brains are unable to effect memory. Here again, I reiterate the underlying issue that memory of these events demands either a conscious brain, or one in the throes of regaining full conscious-awareness. That is a problem which the NDE fraternity has yet to come to terms with and yet, in the absence of hard neurological data, continues to undermine their project. While we remain in a quandary about many of these issues, there are other details about subjects' brains which must be explored in context. One crucial issue is whether NDE subjects' brains are predisposed.

#### 15.2.5 *Concerning a cerebral predisposition toward ND/OBE*

Here, I refer briefly to another obvious question: that is, why the majority of subjects do *not* undergo ND/OBE during one or other lifetime crises. Even about 90 percent of those surviving cardiac arrest fail to report such experiences.

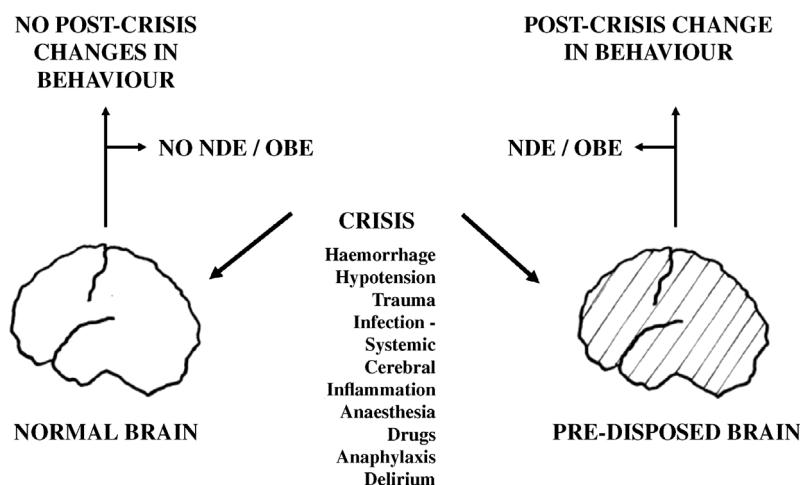
This manner of thinking inevitably suggests that the brains of those undergoing these phenomena are predisposed (Marsh 2016a). Analogous positions involve individuals with post-traumatic stress disorder (PTSD), for whom following the endurance of severe, life-threatening crises, there are marked later changes in phenotypic behavior, in addition to manifest epigenetic changes to the genome (Mehta *et al.* 2013). For many PTSD candidates, there is a clear history of childhood abuse – verbal, physical, even sexual. And with other forms of fearful, threatening assault, whether facing a gun or knife, or undergoing rape, the residual horrifying trauma of such encounters can be persistent and hence life-destroying. These early influences have profound, long-lasting influences on subjects' subsequent mental profiles – however, that profile might be realized in later years (Craig *et al.* 2009; Meyer-Lindenberg, Andreas and Buckholtz 2006; Teicher and Samson 2013).

Demographic factors (ethnicity, gender, age, profession, social class) are not precipitants of ND/OBE. Thus, other deeper factors need consideration, and appropriate methods employed to expose their likely contributions. Nor has the interdependence of gene and environment been well explored. For example, despite the fairly large populations reported, identical twins have not been reported to have had ND/OBE – or more critically, *not*.

Nevertheless, genetic factors are of great conceptual interest, because several polymorphic variations in key brain receptors (genetically determined alterations in the amino acid backbone of these molecules) for such important neurotransmitters as serotonin and dopamine, have crucial influences on environmental responses (Comings *et al.* 2000; Lorenzi *et al.* 2005). These include specific aspects of personality and temperament, even of “spirituality” and “transcendence.” We should note that after selective brain surgery (for variously graded malignant gliomas), a loss of “transcendence” (defined by these authors as undergoing spiritual epiphanies concerning beliefs in unseen realities and supernatural agents) was observed after interference specifically with the parietal lobes (Urgesi *et al.* 2010). Importantly, preoperative assessments contrasted with patients' postoperative behavioral profiles. There is an extensive literature pertaining to such influences, and similar methodologies need to be applied to large cohorts of ND/OBE subjects, compared with age/gender/ethnically matched control groups. How sure can we be that ND/OBE do not comprise a subtle group of individuals whose characteristics await identification and exploration?

For example, in addition to genomic traits, and gene-environment interactions with their resultant outcomes, Britton and Bootzin (2004) demonstrated that some ND/OBE subjects do have subclinical structural damage to their temporal lobes, a viewpoint still deserving wider, in-depth study. Their subjects revealed greater prevalences for inter-ictal epileptiform wave activity, albeit on the left-hand side thus showing little correlation with the specific features of each NDE undergone. But left-sided temporal lobe activity of this type is more likely to be associated with other possibilities within the ND/OBE experiential spectrum – sensing of invisible “presences”; the meeting of “spirits”; flashbacks; and life-reviews. Moreover, a neurological predisposition may be far subtler than other authors (Roberts *et al.* 1990; Persinger and Makarec 1993) have indicated. There are enough post-ND/OBE people around, so it is surprising (if not disappointing) that such intriguing observations have not been initiated.

From another perspective, there are NDE studies indicative of an abnormal perception of bodily *habitus*, dissatisfaction with private and public dispositions, and low regard for body shape and appearance (Murray and Fox 2005), pointing these perceptual differences in body shape to neurological dysfunction. My point here is that the propensity to undergo



**Figure 15.3** It is widely assumed that medical or psychological causes lead to ND/OBE, and out of which changes in post-event behavior are a necessary outcome. But an increasing body of neurological evidence suggests that varied behavioral aberrations occur because the brain is predisposed by earlier insults not routinely detected or clinically observable. It is therefore possible that changed behavioral phenotypes are the principal direct outcomes in individuals subject to one of these many causes (listed), the ND/OBE, itself, being merely epiphenomenal.

ND/OBE is tied closely to a cognitive misjudgment of body shape, and therefore is again reflective of a predisposed brain.

Despite a large body of neurophysiological evidence, far more prospective work needs to be done, involving detailed comparative investigation of brains of potential subjects – *before*, as well as *after*, they undergo ND/OBE. Several causes could be initiating triggers (Figure 15.3) which expose these subjects to an altered behavioral phenotype following exposure to acute psychological or physiological crisis. Moreover, much previous NDE reportage may be significantly biased by excessive use of male coronary care patients.

Finally, note that the parieto-temporal lobes lie at the vascular watershed between the forward carotid, and the rear vertebral-basilar, arterial systems to the brain. Thus acute reductions in pressure would make these regions highly susceptible to the effects of a reduced perfusion and oxygenation. Many studies are being conducted with coronary care patients with atherosclerotic blockage of the coronary arteries. But arterial atherosclerosis is a generalized disease, coincidentally affecting legs and brain, so that changes in cerebral perfusion could be all the more catastrophic. Moreover, as their brains recover, the reperfusion of blood could be uneven, thus allowing some centers to operate without the necessary physiological higher controlling centers being already in place.

The occurrence of NDE as conscious-awareness is being regained could thus be envisioned as a reactive epiphenomenon (Marsh 2013), dependent on the causative (crisis) factor coupled to underlying lesional pathologies (genetic, intra-uterine, or postnatal) in subjects' brains, or other longer-term incursions on cerebral function, as described above. This would not undermine the interesting propositions (Nelson *et al.* 2006; Nelson, Mattingly, and Schmitt 2007) related to genetic abnormalities in brain switching concerning REM intrusion modes, and allied phenomena.

### 15.2.6 REM intrusion and the emergent state of “otherworldly” possession

In their papers, Nelson and colleagues (2006, 2007) remind us that consciousness comprises wakefulness, and two sleep modes – deep sleep and REM sleep, during which the brain is neither nonfunctional nor fully operational. Possible disturbances could exist between these three states, but especially between wakefulness and REM sleep and its related dream-states. However, between these varying states of arousal, anomalous circumstances arise in which subjects cannot discern which state they inhabit, leading to feelings of detachment from themselves and the world, generating a spurious sense of belonging to a greater “union” or “composite association,” and usually defying both its comprehension and expression.

That state exemplifies the profound ineffability of the events undergone during these rare circumstances, and quite distinct from the superficial banality of blue skies and fluffy white clouds characterizing much ND/OBE reporting. Furthermore, this state of uncertainty is probably not much different from other “mystical states” expressing percepts as “Absolute Being,” union with God as expressed by Christian mystics, or other spiritualized encounters with presumptive divine figures (Nelson 2012, 224–246).

Nelson’s argument is partly developed from subjects with transiently abnormal forms of consciousness, such as sleep paralysis (Buzzi 2002), lucid dreaming (Kahan and LaBerge 1994; Voss *et al.* 2009), and narcolepsy. Narcolepsy, the disposition to fall asleep at any time, results from reductions in secretion of the protein hypocretin (orexin) which is genetically controlled (Lin *et al.* 1999; Thannickal *et al.* 2000; Kaur *et al.* 2009). There may also be strong connections with the condition known as *peduncular hallucinosis* (Marsh 2010, 183–186; Marsh 2016a). These disturbances in our varied states of arousal arise from abnormal upper brainstem neuropathology. State transitions occur as molecular switching between the relevant sets of mutually inhibitory neurons is operative, especially those secreting nor-adrenaline or acetylcholine, and other specialized neuronal groups and their specific neurotransmitters (Aston-Jones, Rajkowski and Cohen 1999; Lu *et al.* 2006).

Additional support for this view is provided by the observation that precipitants of ND/OBE, when occasioned by significant reductions in blood pressure (as frequently occur with cardiac arrests, heavy bleeding, or severe anaphylaxis) evoke distant effects on the brainstem, with the stimulation of REM sleep (Puizillout and Foutz 1976). The functional connection arises through cranial nerve X (Vagus) which transmits from the heart and great blood vessels to areas in the upper brainstem associated with switching mechanisms between conscious-awareness and REM sleep (the ventro-lateral peri-aqueductal gray matter, or vlPAG) (Nelson 2012; Vagg, Bandler, and Keay 2008).

These observations are of extreme importance in offering realistic explanations for the utter ineffability of those quintessentially rarer occurrences when subjects no longer know where they are. In these circumstances, subjects seemingly find themselves located in an amorphous existence which now suggests union with a greater outside being, or other nondefinable state of existence, and in which selfhood is deemed to have evaporated.

## 15.3 Theological Issues Arising

Here, I now continue in critically examining the notion of afterlife, as envisioned throughout the NDE literature, compared with Christian precepts of our creedal “Life Everlasting.” In thinking about death and dying, we are confronted by the questions: Why



are we here? What are we for? What is the meaning of life – as well as death? Nowadays, there are three options.

First, that there is no afterlife. Death means complete annihilation, being the collective view of many secularists and humanists.

Second, there is the view based on NDE reports, but added to by those professing specific memories of being born and emergence from the womb, together with the collective deposits based on paranormal and telepathic capacities for future prediction (Marsh 2010, 62–68, 244–252). The binding factor is based on energy-based wave-forms, derivative of the micro-world of classical quantum mechanics, not only extended into the macro-world of every day existence and experience, but firmly welded to all that is present beyond. This professed adherence to wave-forms constitutes the basic precept of (so-called) “Cosmic Consciousness.”

This envisages our puny little brains as mere receptors of a small bit of that universalized cosmic experience. However, when subjects find themselves divested of the physical restrictions of the brain – despite its billions of neurones, and astronomically larger collections of synapses – they taste reality as it actually *should* be tasted, with enhanced higher powers of thinking and knowing; of seeing things in a more lucid light; and with much greater understanding than when we are simply mucking around on earth. On Van Lommel’s account (2010), NDE fulfill these criteria, as well as constituting indelible “proof” of afterlife, even if subjects are only “clinically” dead. However, these subjects recover and in doing so, live to tell the tale. But the anecdotal material recalled, in my view, has little of substance to offer as firm proof for afterlife. Apart from anecdotes, the current problem now being put forward is that when “clinically” dead, a new world is undoubtedly sampled during that period of suspended animation. Unfortunately for Van Lommel, that does *not* mean irreversibly dead and awaiting the arrival of the undertaker’s hearse: that is the problem with his proposed theories.

I simply want people, first, to be aware that van Lommel, with Ring and Grey, have passed beyond any further questioning whether ND/OBE offer valid confirmation of afterlife, and second, to understand the implications of such speculative adventures arising from “clinical” deaths, of which Cosmic Consciousness seems to be the ultimate outcome – if not, in his view, the purpose of all things.

### 15.3.1 *The Kingdom of God*

Our third option, however, is to have belief and hope in a Creator God to whom we return when we have actually undergone death and passed on. But, what will that entail – since Heaven and Earth will also pass away? There is uncertainty about where we will ultimately find ourselves: a whiff of Hebrews (12: 26–29) here, perhaps? There is thus more to life than simply going to “Heaven” (especially of the type envisioned in the ND/OBE literature). Beginning his public ministry, Jesus declared (Isa 61:1–2): “The spirit of the Lord God is upon me – because the Lord has anointed me” – but for what purpose? Briefly: to bring good news to the poor; to release the imprisoned; to let the deaf hear; the blind see; the lame walk; and to comfort the widow, the fatherless, and the oppressed.

Through teaching these particular entities, Jesus inaugurated the Kingdom of God, a Kingdom representing His presence in coming toward us, among us, and within us: and, in engulfing us, offering promise of a society – even on earth – that ideally would comprise love, freedom, and justice.

The Kingdom continually breaks in upon the affairs of earthbound man, and against all that thwarts its advance, or destroys, denies, or usurps the spaces it would occupy. The Kingdom is God's dominion where the right for love and freedom is unfettered and unimpeded, a dominion not somewhere above the cloudless blue skies of the NDE, but focused intently on every impediment that would be a universal infringement of that right to enjoy the God-given realm of His Kingdom, right here – *on earth*. Of course, its realization will not come in this earthly life, and that is why the slaughtering of Jesus, in part, is redemptively intrinsic to salvation and the message which he brought.

From another angle, Marilyn McCord Adams reminds us (2000) that the metaphysical divide must be retained between Godly purity over human corruption and disease, a divide breached and reconciled through incarnational presence, in Man-in-man's descent to earth. There is a theological profundity concerning the in-breaking of the Kingdom here on earth – a foretelling of the perfect society. Yet that firmly expressed idea is considerably at odds with notions offered by a small number of NDE subjects who talk about being in a "heavenly realm":

First, with the near-death "heavenly" typologies (partly sampled above), yet based on very determinedly earthbound idiosyncratic, anthropomorphic, and geo-centered perceptual content.

Second, with so-called "Beings of Light," flattered with capital letters as if to give them some kind of stature and respect, but lacking a defined ontological status while epistemologically bereft of credible meaning contrary to their recognition either as Creator God of the Universe, or Messianic Savior. Moreover, Jesus is now the Christ, gloriously risen and ascended to the right hand of the Father. We just don't suddenly sidle up to these guys in a pool of white light: neither are they dressed in white garments of earthly construction, as the ND/OBE literature invariably portrays.

Third, with the way these sylvan pleasures of the near-death "heavenly vista" could possibly be related to the pain, deprivation, and depravity of living on earth, and the meanings and relationships of the two. What we do see are very simplistic, impoverished, and somewhat childish perceptions of a so-called "heavenly" or "otherworldly" domain "somewhere" up there.

And as we have seen, they reflect mere out-and-out geo-centered, earthbound properties – but we have also caught glimpses of the anguished side of subjects as their consciousness begins to reappear, indicative of frontal lobe activity assuming top-down ultimate control over the brain's cognitive functions – but thereby forcing a return to run-of-the-mill, earthly duties in the run-up to regaining full conscious-awareness. The Kingdom of Heaven is not geographic but within us – here on earth, and centered eucharistically among us, albeit in the Pauline sense of the "not yet" of final salvation. The Kingdom resides in you and me and in our behavioral attitudes to one another. To miss that point is, in reality, to miss the fundamental reality of Jesus's mission for which He, incarnate, came down on earth to establish. "Heaven" is certainly not some kind of hallucinatory tree-lined boulevard decorated with nice border flowers, and with blue skies and wispy white clouds above – up there – out there – somewhere: neither the place of God's abode, nor a place to which "souls" escape.

### *15.3.2 Soul, spirituality, and the NDE*

Classical spiritual encounters were discussed by William James (1902), and by William P. Alston (1993). We can compare classic examples of divine disclosure with those given by NDE subjects.

Consider quotations exemplifying the ineffability and noetic quality of classic disclosures: "One day at prayer . . . I saw . . . or was conscious of . . . Christ at my side . . . I saw nothing with my eyes" (St Teresa of Avila) (Alston 1993, 13). Or: "God surrounds me like the physical atmosphere. He is closer to me than my own breath" (James 1902, 71). And from S. Richard of Victor: "God sheds his sweetness . . . not his brightness . . . his beauty is not seen. He does not appear in the light . . . [but] in the fire . . . [which] warms rather than illumines" (Alston 1993, 52).

But compare those quiet, almost evasive, and remarkably introspective reflections of divine encounter with typical NDE reportage: "I could see my mother and Christ saying, 'Come home' and waving their hands at me. She [wore] a long sparkling silver gown, and so did Christ – long hair, long beard – they were both smiling" (Sabom 1982, 169). Or: "I heard his voice say 'Go back!' I said 'Why me, Lord?' and whoever spoke said my work on earth wasn't over yet . . . all I heard was his voice: it was loud, thundering, just like a clap of thunder coming out of nowhere" [*sic*] (Sabom 1982, 54). They recall Moody's comment, who so pertinently and perhaps unwittingly encapsulated the stark gulf that clearly exists between NDE experience and classical accounts of divine disclosure: "Again and again, my NDE subjects have described to me a panoramic, wrap-around, full-colour, three-dimensional vision of these events in their lives" (Moody 1976, 31).

Considering the "spiritual" properties of NDE reports, one might be somewhat reluctant to credit them with any serious divine import, or, like some of the other reports, of enlightening revelatory content. Furthermore, taken in context with other aspects of ND/OBE and the burgeoning corpus of explanatory neurophysiology associated with brain-based reawakening, it seems more probable that ND/OBE phenomenology neither offers any unique or newer insights into substance dualism, nor provides any compelling indication that these events are occasions when either mind, soul, or raw consciousness are capable of some form of extra-corporeal existence in some ill-defined otherworld.

## 15.4 Conclusion

In introducing this chapter, I declared myself a monist in indicating my belief that human beings are integrated "psychophysical units." Sustaining a stroke or vascular dementia informs us of the catastrophic cognitive and motor outcomes. The brain is the central coordinator of that psychophysical unit, while the "soul" could be envisaged as the emergent personality – encompassing body, developed mind, demeanor, vitality, extraversion or introversion, memory, planning for the future, conscientiousness, and so on. We, ourselves, are souls (Jeeves 1994, 134). Religious believers would wish to add a propensity for sin and guilt (Brown 1998), an apprehension of the divine and tendency for behaving spiritually. Cottingham (2005, 3–8) defines spirituality in terms of behaviors filling the creative and meditative space beyond material satisfactions, concerned thus with action not theory, ways of living rather than doctrinal allegiance, and praxis rather than belief. It is beyond truism that any part of the brain may be involved in religious experience (Saver and Rabin 1997). There is no cerebrally localized holy shrine, shrouded from the day-to-day vulgar commerce of secular neurophysiology.

Antithetically, true death is final and, eschatologically, must be firmly grasped. There can be no residual glowing embers, or "soul" like a beautiful butterfly emerging from the dried-up

shrunk chrysalis of a corpse (Fiddes 2000, 66). Fiddes envisages a cosmic incorporation into Christ's body where, ultimately, we shall finally come to see and know ourselves. My own view is that we shall be incorporated within the Godhead (Father, Son, and Spirit) through baptism. Baptism is here envisaged, within its cosmic and eschatological domains, as a "dying and rising with Christ": a "re-birth" – "not of blood, nor of the will of man, nor of the will of the flesh, but of God" – and thus from "above": an "adoption" anticipating ultimate union within the Godhead, in whom there is "perfect freedom." Therefore we "rejoice that our names are forever enrolled into the heavens." These and analogous quotations acquire meaning only in a metaphysical, rather than earthly, connotation (Marsh 2016b, 90–97, 243–249).

On these lines, it could never be convincingly argued that the recorded deposits of ND/OBE experience represent occasions on which, specifically, so-called "souls" leave the body on excursion to the "heavenly" realm, there to sample "divine presence." Nor that this deposited archive affords cogent revelatory perspectives, hitherto unrevealed, concerning the life eternal. That is far more subtle a prospect than any ND/OBE report comes near to understanding or elucidating. I have been at pains to demonstrate the neurological underpinnings of much of the phenomenology undergone, believing it to represent reawakening phenomena during which extremely vivid illusory/hallucinatory material (hypnopompic) is conjured up by recovering brains – whatever the anterior physiologic insult. Critically, memory is necessary to facilitate later recall: and that can only occur during the wakening period but never when the brain is temporarily inactive due to preceding metabolic breakdown.

## References

- Adams, Marilyn McCord. 2000. *Horrendous Evils and the Love of God*. London: Cornell University Press.
- Alexander, Eben. 2012. *Proof of Heaven*. London: Piatkus.
- Alston, William. 1993. *Perceiving God: The Epistemology of Religious Experience*. Ithaca, NY: Cornell University Press.
- Andersen, Richard. 1997. "Multimodal Integration for the Representation of Space in the Posterior Parietal Cortex." *Philosophical Transactions of the Royal Society London*, 352: 1421–1428.
- Aston-Jones, Gary, Janusz Rajkowski, and Jonathan Cohen. 1999. "Role of Locus Coeruleus in Attention and Behavioural Flexibility." *Biological Psychiatry*, 46: 1309–1320.
- Becker, Karl. 1981. "The Centrality of Near-Death Experiences in Chinese Pure Land Buddhism." *Journal of Near-Death Studies*, 1: 154–171.
- Bede. 1968. *A History of the English Church and People*, translated by Leo Sherley-Price, rev. edn. London: Penguin Classics.
- Bergouignan, L., L. Nyberg, and H. Ehrsson. 2014. "Out-of-Body-Induced Hippocampal Amnesia." *Proceedings of the National Academy of Sciences (USA)*, 111: 4421–4426.
- Blanke, Olaf, and Christine Mohr. 2005. "The Out-of-Body Experience: Disturbed Self-Processing at the Temporo-Parietal Junction." *Brain Research Reviews*, 50: 184–199.
- Brandt, Thomas, and Marianne Dieterich. 1999. "The Vestibular Cortex: Its Locations, Functions, and Disorders." *Annals of the New York Academy of Science*, 871: 293–312.
- Britton, Willoughby, and Richard Bootzin. 2004. "Near-Death Experiences and the Temporal Lobe." *Psychological Science*, 15: 254–258.
- Brown, Warren. 1998. "Cognitive Contributions to Soul." In *Whatever Happened to the Soul?* edited by Warren Brown, Nancy Murphy, and H. Newton Malony, 99–126. Minneapolis, MN: Fortress Press.

- Buzzi, Giorgio. 2002. "Near-Death Experiences." *Lancet*, 359: 2116–2117.
- Comings, David, Nancy Gonzales, Gerard Saucier, Patrick Johnson, and James MacMurray. 2000. "The DR4 Gene and the Spiritual Transcendence Scale of the Character Temperament Scale." *Psychiatric Genetics*, 10: 185–189.
- Cottingham, John. 2005. *The Spiritual Dimension*. Cambridge: Cambridge University Press.
- Counts, Dorothy. 1983. "Near-Death and Out-of-Body Experiences in a Melanesian Society." *Journal of Near-Death Studies*, 3: 115–135.
- Craig, M. C., M. Catani, Q. Deeley, R. Latham, E. Daly, R. Kanaan, et al. 2009. "Altered Connections on the Road to Psychopathy." *Molecular Psychiatry*, 14: 946–953.
- Donne, John. 2004. *One Equall Light: An Anthology of the Writings of John Donne*, edited by John Moses. Grand Rapids, MI: Eerdmans.
- Drab, Kevin. 1981. "The Tunnel Experience: Reality or Hallucination?" *Journal of Near-Death Studies*, 1: 126–152.
- Ehrsson, H. 2007. "The Experimental Induction of Out-of-Body Experience." *Science*, 317: 1048.
- Fenwick, Peter, and Elizabeth Fenwick. 1998. *The Truth in the Light*. New York: Berkeley Books.
- Fiddes, Paul. 2000. *The Promised End*. Oxford: Oxford University Press.
- Forster, Estelle, and James Whinnery. 1988. "Recovery from Forced Gz-induced Loss of Consciousness: Psychophysiologic Considerations." *Aviation, Space, and Environmental Medicine*, 59: 517–522.
- Grey, Margot. 1985. *Return from Death*. Boston, MA: Arkana.
- Greyson, Bruce, and Nancy Bush. 1992. "Distressing Near-Death Experiences." *Psychiatry*, 55: 95–110.
- Hobson, JA. 2002. *Dreaming: An Introduction to the Science of Sleep*. Oxford: Oxford University Press.
- Howard P., G. Leathart, and A. Dornhorst. 1951. "The 'Mess Trick' and the 'Fainting Lark.'" *British Medical Journal*, 3: 382–384.
- Ionta, Silvio, Lukas Heydrich, Bigna Lenggenhager, Michael Mouthon, Eleonora Fornari, Dominique Chapuis, et al. 2011. "Multisensory Mechanisms in Temporo-Parietal Cortex Support Self-Location and First-Person Perspective." *Neuron*, 70: 363–374.
- James, William. 1902. *The Varieties of Religious Experience*. London: Longmans, Green.
- Jeeves, Malcolm. 1994. *Mind Fields: Reflections on the Science of the Mind and Brain*. Leicester, UK: Apollos.
- Kahan, Tracey, and Stephen LaBerge. 1994. "Lucid Dreaming as Metacognition: Implications for Cognitive Science." *Consciousness and Cognition*, 3: 246–264.
- Kaur, Satvinder, Stephen Thankachan, Ssuraiya Begum, Meng Liu, Carlos Blanco-Centurion, and Priyattam Shiromani. 2009. "Hypocretin-2 Saporin Lesions of the Ventro-Lateral Periaqueductal Gray (vlPAG) Increase REM Sleep in Hypocretin Knockout Mice." *PLOS ONE*, 4(7): e6346.
- Lackner, James. 1988. "Some Proprioceptive Influences on the Perceptual Representation of Body Shape and Orientation." *Brain*, 111: 281–297.
- Lempert T., M. Bauer, and D. Schmidt. 1994. "Syncope and Near-Death Experience." *Lancet*, 344: 829–830.
- Lin, Ling, Juliette Faraco, Robin Li, Hiroshi Kadotani, William Rogers, Xiaoyan Lin, et al. 1999. "The Sleep Disorder Canine Narcolepsy is Caused by a Mutation in the Hypocretin (orexin) Receptor 2 Gene." *Cell*, 98: 365–376.
- Lindley, James, Sethyn Bryan, and Bob Conley. 1981. "Near-Death Experiences in a Pacific Northwest American Population." *Journal of Near-Death Studies*, 1: 104–124.
- Lippman, Caro. 1953. "Hallucinations of Physical Duality in Migraine." *American Journal of Psychiatry*, 117: 345–350.
- Long, Jeffrey. 2010. *Evidence of the Afterlife*. New York: Harper One.
- Lorenzi, Cristina, Serretti Alessandro, Mandelli Laura, Tubazio Viviana, Ploia Cristina, and Smeraldi Enrico. 2005. "5-HT 1A Polymorphisms and Self-Transcendence in Mood Disorders." *American Journal of Medical Genetics, Part B: Neuropsychiatric Genetics*, 137: 33–35.

- Lu, Jun, David Sherman, Marshall Devor, and Clifford B. Saper. 2006. "A Putative Flip-Flop Switch for Control of REM Sleep." *Nature*, 441: 589–594.
- Marsh, MICHAEL N. 2010. *Out-of-Body and Near-Death: Brain-State Phenomena or Glimpses of Immortality?* Oxford: Oxford University Press.
- Marsh, MICHAEL N. 2013. "Near-Death & Out-of-Body Experiences." In *Encyclopedia of Science and Religions*, edited by Anne Runehov, and Luis Oviedo, 1422–1433. Berlin: Springer.
- Marsh, MICHAEL N. 2015. "Hey! What's That Gorilla Doing Over There? On the Illusory-Hallucinatory Nature of Everyday Living." *European Review*, 23: 455–472.
- Marsh, MICHAEL N. 2016a. "The Near-Death Experience: Reality Check?" *Humanities* 5: 18–43. DOI: 10.3390/h5020018.
- Marsh, MICHAEL N. 2016b. *On Being Human: Distinctiveness, Dignity, Disability, Disposal*. New York: IFF Books.
- Mehta, Divya, Torsten Klengel, Karen Conneely, Alicia Smith, André Altmann, Thaddeus Pace, *et al.* 2013. "Childhood Maltreatment is Associated with Distinct Genomic and Epigenetic Profiles in Posttraumatic Stress Disorder." *Proceedings of the National Academy of Sciences (USA)*, 110: 8302–8307.
- Meyer-Lindenberg, Andreas, Joshua Buckholtz, Bhaskar Kolachana, Ahmad Hariri, Lukas Pezawas, Giuseppe Blasi, *et al.* 2006. "Neural Mechanisms of Genetic Risk for Impulsivity and Violence in Humans." *Proceedings of the National Academy of Sciences (USA)*, 103: 6269–6274.
- Mittelstaedt, Horst. 1991. "The Role of the Otoliths in the Perception of the Orientation of Self and World to the Vertical." *Zoologische Jahrbucher Physiology*, 95: 419–425.
- Moody, R. 1976. *Life after Life*. New York: Bantam Books.
- Murray, Craig, and Jezz Fox. 2005. "The Out-of-Body Experience and Body Image: Difference between Experiencers and Non-Experiencers." *The Journal of Mental and Nervous Disease*, 193: 70–72.
- Nelson, Kevin. 2012. *The Spiritual Doorway in the Brain: A Neurologist's Search for the Experience of God*. New York: Plume.
- Nelson, Kevin, Michelle Mattingly, Sherman Lee, and Frederick Schmitt. 2006. "Does the Arousal System Contribute to Near Death Experience?" *Neurology*, 66: 1003–1009.
- Nelson, Kevin, Michelle Mattingly, and Frederick Schmitt. 2007. "Out-of-Body Experience and Arousal." *Neurology*, 68: 794–795.
- Parnia, Sam, Ken Spearpoint, Gabriele de Vos, Peter Fenwick, Diana Goldberg, Jie Yang, *et al.* 2014. "AWARE – AWAREness during REsuscitation – a Prospective Study." *Resuscitation*, 85: 1799–1805.
- Persinger, Michael, and Katherine Makarec. 1993. "Complex Partial Epileptic Signs as a Continuum from Normals to Epileptics: Normative Data and CLINICAL Populations." *Journal Clinical Psychology*, 49: 33–45.
- Puizillout, J., and A. Foutz. 1976. "Vago-Aortic Nerves and REM Sleep: Evidence for a REM-Triggering and an REM-Maintenance Factor." *Brain Research*, 111: 181–184.
- Ring, Kenneth. 1980. *Life at Death*. New York: Coward, McCann & Geoghegan.
- Ring, Kenneth. 1994. "Solving the Riddle of Frightening Near-Death Experiences." *Journal of Near-Death Studies*, 13: 5–23.
- Roberts, Richard, Nils Varney, James Hulbert, Jane Paulsen, Emily Richardson, Jane Springer, *et al.* 1990. "The Neuropathology of Everyday Life: The Frequency of Partial Seizure Symptoms Among Normals." *Neuropsychology*, 4: 65–85.
- Rosen, David. 1975. "Suicide Survivors: A Follow-Up Study of Persons Who Survived Jumping from the Golden Gate and San Francisco-Oakland bridges." *The Western Journal of Medicine*, 122: 289–294.
- Sabom, M. 1982. *Recollections of Death*. New York: Harper & Row.
- Saver, Jeffrey, and John Rabin. 1997. "The Neural Substrate of Religious Experience." *Journal of Neuropsychiatry and Clinical Neurosciences*, 9(3): 498–510.

- Seiden 1968. "Where Are They Now? A Follow-Up Study of Suicide Attempters from the Golden Gate Bridge." *Suicide and Life-Threatening Behavior*, 8: 203–216.
- Snyder, Richard, and Clyde Snow. 1967. "Fatal Injuries Resulting from Extreme Water Impact." *Aerospace Medicine*, 38: 779–783.
- Solms, Mark. 2000. *The Neuropsychiatry of Dreams: A Clinico-Anatomical Study*. Mahwah, MA: Lawrence Erlbaum.
- Teicher, Martin, and Jacqueline Samson. 2013. "Childhood Maltreatment and Psychology: A Case for Ecophenotypic Variants as Clinically and Neurobiologically Distinct Subtypes." *American Journal of Psychiatry*, 170: 1114–1133.
- Thannickal, Thomas, Robert Moore, Robert Nienhuis, Lalini Ramanathan, Seema Gulyani, Michael Aldrich, *et al.* 2000. "Reduced Numbers of Hypocretin Neurons in Human Narcolepsy." *Neuron*, 27: 469–474.
- Urgesi, Cosimo, Salvatore M. Aglioti, Miran Skrap, and Franco Fabbro. 2010. "The Spiritual Brain: Selective Cortical Lesions Modulate Human Self-Transcendence." *Neuron*, 65: 309–319.
- Vagg, D., R. Bandler, and K. Keay. 2008. "Hypovolemic Shock: Critical Involvement of a Projection from the Vento-Lateral Peri-Aqueductal Gray to the Caudal Midline Medulla." *Neuroscience*, 152: 1099–1109.
- Van Lommel, Pim. 2010. *Consciousness Beyond Life: The Science of the Near-Death Experience*. New York: Harper One.
- Voss, Ursula, Romain Holzmann, Inka Tuin, and J. Allan Hobson. 2009. "Lucid Dreaming: A State of Consciousness with Features of Both Waking and Non-Lucid Dreaming." *Sleep*, 32: 1191–1200.
- Wilson, Ian. 1997. *Life after Death*. London: Pan Books.

# Why Reject Substance Dualism?

IAN RAVENSCROFT

In such a state of mind, the imagination swells with great but confused conceptions . . . and a full range is given to the fancy in the invisible regions, or world of Spirits.

David Hume, *Of Superstition and Enthusiasm* (1994 [1777])

As the chapters in this volume indicate, there has been a resurgence of enthusiasm for substance dualism (SD) in recent years. In what follows I will argue that such enthusiasm is unwarranted. My claim is not that SD is definitely false; rather my principal argument will be that its lack of explanatory power relative to physicalism rules it out of contention given the current evidence.

I begin in Section 16.1 with a few words about the nature of SD, and in Section 16.2 I briefly describe physicalism. Section 16.3 replies to some but not all of the arguments on offer for SD, and is entirely negative. (The restriction to *some* arguments is necessitated by considerations of space.) Section 16.4 briefly discusses substance dualism's relationship to three key elements of religious doctrine, and Section 16.5 discusses the idea of emergence. Section 16.6 is devoted to describing and defending what I take to be the most serious arguments for physicalism and consequently against substance dualism. While one well-known argument against substance dualism is found wanting, others are very strong. In Section 16.7 I draw an analogy between SD and one kind of creationism. In a brief conclusion I argue that SD is only a theory in a very attenuated sense of that term.

A word on terminology. Throughout I will avoid the phrase "soul," using instead the term "nonphysical mind" or, when the context makes it clear, "mind." The word "soul" should be retired in this context because of the religious connotations that often surround it. SD itself does not entail the continuation of mental life after bodily death and is uncommitted on the existence of God.



## 16.1 Varieties of SD

According to SD, mind and brain are radically different substances, the former nonphysical, the latter physical. The question of interaction between these two substances immediately arises. How, if at all, do the nonphysical mind and the physical brain interact? According to *interactionist* (or “Cartesian”) SD, the nonphysical mind causally impacts on the brain, and vice versa. According to *epiphenomenal* SD, the brain causally impacts on the mind, but not vice versa. *Parallelist* SD denies causal relations in both directions. On this view, God synchronizes the unfolding of mental and physical processes so they are always appropriately correlated. Finally, advocates of *occasionalist* SD claim that God appropriately correlates mental and physical events occasion by occasion.

There is another important way in which various types of SD can be distinguished. Physics recognizes a number of physical properties and relations such as charge, mass, and velocity. If nonphysical substance is radically distinct from physical substance, then presumably they have no properties in common and so nonphysical minds (or the components from which they are constituted) do not have properties like charge, mass, and velocity.

Understood in this way, both interactionist and epiphenomenal SD are extraordinarily radical doctrines. To see this, consider the physical properties of location in space and time (or the property of location in spacetime). If nonphysical minds really lack all physical properties then they are not located in space *or* time. However, substance dualists need not be committed to this extreme doctrine. They could argue that there is a set of nonphysical properties that all minds have essentially (say, being accessible to consciousness), but nonphysical minds can also contingently have a range of physical properties (e.g., location in spacetime). An interesting consequence of this proposal is that some properties cannot be neatly classified as either physical or nonphysical: location in spacetime, for example, may be neither exclusively physical nor exclusively nonphysical.

Finally, it should be noted that substance dualism is distinct from *property* dualism. The first asserts that the mind is a nonphysical substance; the other asserts that mental properties (e.g., believing that trout are fish) are nonphysical. This distinction is important in what follows.

## 16.2 Physicalism

Assessing SD requires understanding what physicalism amounts to. In the present context physicalism is the view that the mind is a physical object. Physicalists about the mind usually claim that the mind is the brain, but care is required here. While physicalists typically think that all mental states are brain states, they deny that all brain states are mental states. For example, they do not believe that the glial cells which play a support role in the brain are identical to mental states. In addition, some physicalists argue that the mind extends beyond the brain into the wider world (Clark and Chalmers 1998). On this view my notebook might literally be part of my memory. The extended mind hypothesis is physicalist because both the brain and the world into which the mind extends are physical. For ease of expression I will in what follows ignore the extended mind hypothesis and only talk about brain states.

An important approach to physicalism claims that mental properties (e.g., believing, fearing, and wanting) *supervene* on physical properties (e.g., the firing of neurons in a

certain part of the brain). Donald Davidson, an early proponent of supervenience in the philosophy of mind, described the supervenience of the mental on the physical like this: “an object cannot alter in some mental respect without altering in some physical respect” (Davidson 1980 [1970], 214). On this view, if physicalism is true I cannot acquire a new belief without my brain changing in some way. This account of supervenience will do for present purposes; more sophisticated versions can be found in, for example, David Lewis (1983) and Frank Jackson (1998).

According to *reductive* physicalism, types of mental states are identical to types of brain states. That is, each type of mental state – for example, believing that rabbits are cute – is identical to a type of brain state – for example, activity in part X of the cerebral cortex. (This example is drastically oversimplified, but it will do for present purposes.) If reductive physicalism is true then the mental supervenes on the physical, but the converse is not true because there are *nonreductive* versions of physicalism that are compatible with supervenience. According to this view, while some instances of believing that rabbits are cute might be identical to activity in part X of the cerebral cortex, others are identical to activity in part Y.

It’s important to get clear on what physicalism is *not* committed to. First, physicalism is not committed to *scientism* – the view that science is the only method for gathering genuine insights into the world. It follows that physicalism need not have some deep opposition to the first-person perspective. Second, physicalism is not committed to *eliminativism* – the view that there are no beliefs or desires. (We will return to the issue of eliminativism shortly.) Moreover, physicalists rarely advocate the claim that dualism is necessarily false. Finally, contrary to what some substance dualists seem to suggest (e.g., Baker and Goetz 2011; Baker 2011), the strongest arguments for physicalism do not rely on Ockham’s Razor – the claim that, other things being equal, simple hypotheses are to be preferred to complex ones.

## 16.3 Some Mistaken Arguments for SD

### 16.3.1 *Free will, folk psychology, and the “lived life”*

Substance dualists sometimes regard the existence of free will as deeply problematic for physicalism and therefore as supportive of SD (Baker and Goetz 2011). The arguments on offer seem to have the following structure (see, e.g., Taliaferro 2011; Robinson 2011).

Premise 1. SD is the only metaphysical view compatible with humans having free will.

Premise 2. At least some humans have free will.

Conclusion. SD is true.

This argument faces three objections. Notice first that Premise 1 would be brought into question if one or more plausible physicalist theories of free will were available. And such theories *are* available. To give just one example, Harry Frankfurt (1971) has argued that free will requires second-order desires (i.e., desires about desires) which control first-order desires. The smoker’s first-order desire for another cigarette isn’t free if she has the second-order desire to give up smoking. Frankfurt’s theory is entirely compatible with physicalism, so substance dualists who advocate arguments of the form given above need to offer substantial arguments against Frankfurt’s and other physicalists’ theories of free will.

Second, Premise 2 of the argument requires defense. It cannot simply be asserted that humans have free will. Granted, the idea that we have free will is deeply intuitive, but

that doesn't show it's true. Many other deeply intuitive claims have proven to be false. The problem of free will is a *problem* because there are good reasons to think that our thoughts and actions are entirely determined by prior states. This possibility cannot be casually ignored.

Finally, notice that substance dualists owe us an account of how the nonphysical mind has free will. What would a substance dualist theory of free will be like? A substance dualist might insist that nonphysical minds *just are* free, but that looks entirely ad hoc.

Some substance dualists (e.g., Taliaferro 2011; Robinson 2011) have misunderstood the philosophical debates about folk psychology and eliminativism, and seem to believe that objecting to eliminativism provides the basis for an argument against physicalism. The expression "folk psychology" is used in a variety of ways (Ravenscroft 2016), but for present purposes we can take folk psychology to be our everyday (or "folk") assumptions about human psychology which include the following: People have many beliefs, including false beliefs; they have desires or wants, some of which get satisfied; they think and have sensations and emotions; they reason from one belief or set of beliefs to other beliefs; and they are motivated to act in certain ways. At its most extreme, eliminativism is the idea that there are no such states as beliefs, desires, emotions, reasons, and so on, but more modest versions are available (see below). Eliminativists often argue that folk psychology is a *theory of human psychology* that is, moreover, false. They then go on to argue that the states and processes proposed by false theories don't exist and conclude that there are no beliefs, desires, and so forth (Ravenscroft 2016). Notice that eliminativism comes in degrees; it is possible, for example, to be eliminativist about beliefs and desires but not about sensations or emotions. Two of the most discussed works in the eliminativist literature (Churchland 1981 and Stich 1983) endorse a limited form of eliminativism, only denying the existence of states like beliefs and desires. Notice that the title of Churchland 1981 is "Eliminative Materialism and the Propositional Attitudes." The propositional attitudes are, very roughly, beliefs and desires, so Churchland is only advocating the limited view that there are no beliefs and desires. Again, the title of Stich (1983) is *From Folk Psychology to Cognitive Science: The Case Against Belief*, so Stich is only an eliminativist about beliefs and other propositional attitudes.

Eliminativists like Paul Churchland (1981) argue that folk psychology should be abandoned in favor of neuroscience in much the same way that Ptolemy's astronomy was replaced by Copernicus's. But this move isn't compulsory: eliminativists could argue that folk psychology should be replaced by a more respectable psychology which supervenes on, but is not identical to, neuroscience (Horgan and Woodward 1985).

It is hard to discern exactly how rejecting eliminativism is supposed to support SD. After all, a great many physicalists reject eliminativism (see, e.g., Jackson and Pettit 1990; Horgan and Woodward 1985). Some substance dualists (Baker and Goetz 2011; Taliaferro 2011) seem to suggest that introspection yields beliefs about our mental lives that are vastly more compelling than the arguments of the eliminativists (note that this argument does *not* appeal to the incorrigibility of introspection). But for three reasons this is hard to take seriously. First, the argument relies on being able to measure the reliability of introspection and compare that with the strength of arguments for eliminativism; no indication of how this would be achieved is given. Second, the claim that introspection is especially reliable is contradicted by a vast array of empirical research (for a brief introduction see Wilson 2002). To begin with, there are many important mental states and processes that are not available to introspection *at all*. For example, many of the states and processes involved in language processing cannot be introspected. In addition, humans frequently *make up* (or

“confabulate”) motivations to explain their behavior – motivations that can be ruled out by careful experimentation. The participants aren’t lying; rather, they are mistaken about their own motivations. In cases like this it is clear that introspection does not reveal what’s really going on. To insist in the face of the empirical data that introspection is highly reliable seems perverse. Finally, notice that eliminativist SD is not incoherent: maybe the workings of the nonphysical mind don’t conform to folk psychology.

Robinson (2011) uses the notion of “lived life” when defending SD, and seems to regard it as the basis of an anti-physicalist argument. Robinson says very little about what “lived life” is supposed to be, but I take it that this phrase refers to our experiences of ourselves, each other, and our world, including our aesthetic and moral experiences. For example, most of us think of ourselves as being motivated by beliefs, desires, and emotions, and that much of the time we act of our own free will. If we regard our “lived life” as epistemically prior – if we think of it as a datum to be taken very seriously – then arguments against free will and for eliminativism are unlikely to succeed.

There are three problems here. First, we can ask which elements of our human life fall under the concept of a “lived life”? Are certain kinds of aesthetic experiences an essential component? There are related questions about how universal the “lived life” is. Do the Yanomamo share my Western, middle class “lived life”? And there are developmental questions about “lived life.” Do three-year-olds have the same “lived life” as adults? These questions need to be answered before we can assess Robinson’s claim. Second, why should the “lived life” have this kind of epistemic supremacy? We need an argument for that highly controversial claim. Finally, it’s not clear how appealing to the “lived life” poses a problem for physicalism. We need an argument to show that the various aspects of “lived life,” whatever they are supposed to be, aren’t compatible with physicalism.

At one point Robinson (2011) argues that any neuroscientific investigation which identifies a folk psychological state with a brain state requires that the participants can accurately identify their folk psychological states. The primacy of folk psychology over neuroscience is thus established because unless the experimenter takes folk psychology seriously the experiment is pointless. But this argument also fails because it assumes that physicalism is committed to scientism and eagerly awaits the elimination of folk psychology. We have already seen that this conception of physicalism is very wide of the mark. Notice also that if eliminativism is correct there are no folk psychological states to identify with brain states. No primacy there. Finally, it is plausible that folk psychology and neuroscience will engage in a process of mutual adjustment, with results from folk psychological-level scientific investigations constraining those based on neuroscience, and vice versa. Again, the claim of primacy fails.

The issue of the primacy of folk psychology is closely related to the claim that the third-person perspective exemplified by science cannot yield an understanding of the first-person perspective. Drawing attention to bats’ use of sonar to navigate in the dark, Nagel (1974) asks how studying bats from the third-person perspective of science could yield an understanding of what bats’ sonar experiences are like. On this view there will always be something that science cannot examine. This argument is unconvincing for two reasons. First, as already emphasized, physicalism is not committed to scientism; that is, it is not committed to the claim that science is the only reliable way to understand the world. It follows that physicalists can accept that the first-person perspective is both important and not well understood from the third-person perspective. In addition, Nagel’s argument and its conclusion are not universally accepted, with some physicalists arguing that the third-person perspective *can* yield the first-person perspective. Arguments of this kind can be

found in, for example, Churchland (1985) and Jackson (2007). Nagel's argument leads naturally to the next argument for SD that I will consider.

### 16.3.2 *Only SD can account for phenomenal consciousness*

By "phenomenal consciousness" I mean the subjective qualities of some of our mental states. The idea is often captured by Nagel's phrase "what it is like" (Nagel 1974). There is something that it is like to see red, and what it is like to see red is different from what it is like to see green. That's a difference of phenomenal consciousness. Again, there is something that it is like to be hungry, and what it is like to be hungry is different from what it is like to be thirsty. That too is a difference of phenomenal consciousness.

Various arguments have been advanced in support of the claim that physical objects like the brain cannot be phenomenally conscious and therefore physicalism is false (see, e.g., Chalmers 1996). If phenomenally conscious experiences aren't physical, they must be nonphysical and so some version of dualism – perhaps SD – must be true.

This line of argument is not without problems. To begin with, physicalists have both developed counterarguments to the claim that phenomenal consciousness entails some kind of dualism and have put forward physicalist theories of phenomenal consciousness (among many other examples, see Dennett 1991; Papineau 2002). Second, notice that the proponents of anti-physicalism about phenomenal conscious are not universally proponents of substance dualism. For example, Frank Jackson (1982, 1986), in his celebrated knowledge argument against physicalism about phenomenal consciousness, endorsed not *substance* dualism but a restricted version of *epiphenomenal property* dualism. According to this doctrine, all substances are physical substances and most physical substance have only physical properties. However, a relatively small number of physical objects including the human brain have in addition nonphysical properties; in particular, they have phenomenal properties. So even if the knowledge argument is plausible, its success would not straightforwardly support SD. It's worth noting that since the 1990s Jackson has vigorously argued *against* the knowledge argument (Braddon-Mitchell and Jackson 1996; Jackson 2007; Jackson 2009). Little comfort for SD there.

Finally, notice that substance dualists have no account of how the nonphysical mind is phenomenally conscious, so even if it is accepted that physicalism struggles to account for phenomenal consciousness, there is no easy victory for SD. It might be asserted that the nonphysical mind *just is* conscious, but now the substance dualist must make that claim plausible. In addition, if it's OK to assert without argument that the nonphysical mind *just is* phenomenally conscious, why can't the physicalist help herself to the same move? The brain *just is* phenomenally conscious and no explanation of that fact is required.

## 16.4 First Interlude: SD and Religion

My guess is that among substance dualists there are more theists than atheists. I admit that I have no sociological data to establish this claim, but I think it's plausible. In this section I will argue that some religious doctrines sit more comfortably with physicalism than is often thought.

To begin with, theism does not entail substance dualism. In many religions, including Judaism, Christianity, and Islam, God is taken to be all-powerful, and presumably an

all-powerful being could create a physical mind. Deists could endorse this position, arguing that God created the physical world knowing that it would unfold to bring about beings with physical minds. Notice also that physicalism appears to be compatible with certain kinds of theism: some theists have defended the view that God is physical (see, e.g., Bishop 2009).

Religions typically – perhaps always – make normative claims about both thoughts and actions: one must not covet one’s neighbor’s wife; one must be charitable; and so on. Some substance dualists appear to believe that SD is preferable to physicalism because only the former can account for the existence of morality. But there is little reason to think that morality is incompatible with physicalism; after all, most metaethical theories today take physicalism for granted (see e.g., Smith 1994). Perhaps the reluctance of some substance dualists to accept that morality is compatible with physicalism is the belief that morality requires free will, and free will is in turn incompatible with physicalism. But as we have already seen (Section 16.3.1), physicalists have offered accounts of free will.

Very many religions – perhaps all – endorse the claim that we survive our bodily death, and some theists believe that after bodily death we live on as nonphysical substances. SD sits comfortably with the idea of nonphysical survival after death: if my mind is a nonphysical substance then it may survive the death of my body. But this move is not compulsory because it’s possible to give a physicalist account of reincarnation. Functionalists think that mental states are characterized by the functional roles they play. Very crudely, pain is the state that is caused by bodily damage, sometimes causes fear, and usually leads to grunting, swearing and avoidance behavior. Overwhelmingly, functionalists are physicalists, endorsing what I have called elsewhere (Ravenscroft 2005) the “transitivity argument”:

Premise 1. Mental state M = the occupant of functional role R.

Premise 2. The occupant of functional role R = brain state B.

Conclusion. Mental state M = brain state B.

(For ease of expression I have set aside the distinction between type and token physicalism.)

In principle the occupant of functional role R could be a state of a nonphysical substance and therefore substance dualists could be functionalists. However, I know of no substance dualist who has explored or endorsed this option.

Returning to the issue of life after bodily death, it *may* be the case that the functional roles occupied in my brain immediately before my death could be reproduced in another brain in a healthier body, in which case I would have been reincarnated. I emphasized “may” in the last paragraph because some philosophers would question the idea that replicating functional roles is enough to preserve a person’s identity over time.

I conclude that the supposed antagonism between physicalism and religious doctrine is not as strong as it is sometimes thought to be.

## 16.5 Second Interlude: Emergence

Some dualists are attracted to *emergence*, although it is unclear that it is a form of substance dualism; indeed, it is not clear that it is a form of dualism at all, and if it is it would seem to be a form of property dualism. It should be emphasized that the arguments about emergence have become very sophisticated (see, e.g., Kim 1999 and McLaughlin 1997) and that what follows barely scratches the surface.

Objects can have properties distinct from the properties of the objects from which they are composed. Balls, for example, have the property of being round, but their constituent parts may not be round. Again, trees have the property of being alive, but the atoms from which they are composed do not. While this much is uncontroversial, emergentists make the much more radical claim that higher-level properties are of a different metaphysical kind to lower-level properties. For example, it has been claimed that the property of being alive is distinct from, and cannot be reduced to, the properties of the matter from which living things are composed (Mill 1843; Broad 1925). In addition, emergentists claim that the emergent properties are governed by different sets of laws to those which govern the properties from which they emerged.

In the philosophy of mind emergentists say that mental properties emerge from neurological properties, and that the former are distinct from, and cannot be reduced to, the latter. Nor, it should be added, are mental properties governed by the same laws as those that govern neurological properties (see, e.g., Hasker 2011). Described this way, emergentism sounds like a form of property dualism because it claims that mental properties are distinct from, and cannot be reduced to, physical properties, and consequently it is not a form of substance dualism.

Notice that emergentism is compatible with the supervenience of the mental on the physical: emergentists do *not* claim that mental properties can change without changes occurring at the neurological level. Mental properties emerge from and are consequently dependent upon, neurobiological properties.

To assess emergence it is useful to consider the way physicalists respond to the observation that objects can have properties distinct from those of their constituent parts. According to physicalism, mental properties are physical properties if and only if they supervene on physical properties; that is, if and only if there can be no change in mental properties without a change in, for example, neurological properties. This is not to deny that psychological laws are distinct from neurobiological laws, nor is it to deny that psychological properties have causal powers distinct from those of neurological properties. After all, this is a familiar feature of the physical sciences (Fodor 1974). For example, the laws of geology are distinct from those of atomic science, but it does not follow that geological structures aren't physical. On the standard account of physicalism, then, the so-called "emergent" properties are just physical properties. Systems like the brain can be described at different levels, beginning with the atomic level and ending at the psychological level, but that is consistent with physicalism.

At the heart of the emergence thesis is the claim that the mental does not reduce to the physical. We saw in Section 16.2 that, according to reductive physicalism, each type of mental property is identical to a type of physical property. J. J. C. Smart (1959) insisted that the identities are not known a priori but are discovered by science. By analogy, the type water is identical to the type  $H_2O$ , and water is said to have been reduced to  $H_2O$ . If it can be shown that one type of mental property is identical to a type of neurological property, then emergence about the mental is false. And the current evidence is consistent with (a restricted kind of) mind-brain identity theory.

## 16.6 Some Arguments for Physicalism

In this section, I discuss what I take to be the most important arguments offered in favor of physicalism and therefore against SD. One well-known argument I won't discuss at any

length is the argument from the conservation of energy. The principle of conservation of energy says that the energy in any isolated physical system is fixed – it neither increases nor decreases. Both substance and epiphenomenal dualism seem to violate the principle because if a nonphysical mind were to influence a physical system like the brain then the energy of that system would increase (see, e.g., Flanagan 1991; Fodor 1994). Collins (2011) has argued that quantum physics does not support the principle of conservation of energy, and if he is right the conservation of energy argument against SD fails. However, I do not know enough about quantum physics to pass judgment on Collins's claims and so I will not pursue this objection to SD.

### 16.6.1 *Princess Elisabeth's objection*

Princess Elisabeth of Bohemia made the following astute objection to René Descartes's interactionist SD (Elisabeth and Descartes 2007). According to interactionist SD there are causal relations from mind to brain and vice versa. But also, according to SD, mind and brain are two quite different kinds of substance: one is physical, the other nonphysical. How can such disparate substances have causal relations? In effect, Elisabeth asked Descartes to square his interactionism with his substance dualism.

If it works, this kind of argument would also sound the death knell of epiphenomenal SD since, while epiphenomenal SD denies that mental states can influence brain states, it insists that brain states can influence mental states. Both parallelist and occasionalist SD (neither of which admit causal relations between mind and body) escape Elisabeth's objection, but they do so at considerable cost because they rely on God to correlate mental states with brain states. But the existence of God is, to put it mildly, deeply controversial among present-day philosophers, and anyway it is at least as puzzling as mind-brain interaction.

I think Princess Elisabeth's objection is problematic. To begin with, we can ask whether we need be so suspicious of causal relations between the physical brain and the nonphysical mind. Is there really a problem here? Notice that we do not find causal relations between quite different kinds of *physical* properties especially problematic. For example, sunlight – a form of electromagnetic radiation – causes iron – an array of atoms – to heat up. Sunlight and iron are very different, but no one raises objections to the idea that they causally interact.

One response to the sunlight-iron objection is that while we have well-established theories of how sunlight and iron interact, substance dualists have nothing to say about how the nonphysical mind interacts with the physical brain. I think that's an important observation; however, it does not solve our current problem because people accepted that sunlight causes iron to heat up long before they had the remotest idea of how that might occur. Presumably soldiers of the Roman Empire noticed that sunlight caused their iron equipment to heat up but they lacked sophisticated theories of that fact. Indeed, they may not have even noticed that the phenomenon is in need of explanation. The physicalist could insist that the physical and nonphysical are much more divergent than are sunlight and iron, but mere insistence isn't much to rest an argument on.

Elisabeth's concern may have been driven by the idea, prevalent at the time, that causation is a kind of pushing or pulling that requires an interface or "nexus" between cause and effect. However, the idea that causation requires a nexus between cause and effect was rejected by Hume (1975 [1777]) and, according to some writers, plays no role in modern physics (Russell 1912). If this is right, at least some of the counterarguments to dualism deriving from Elisabeth's objection have been cleared away.



### 16.6.2 Causal closure

What I will call the *strong* principle of causal closure says that all physical events have only physical causes. The physical world is “closed” to nonphysical influence. It is clear that the strong principle of causal closure entails that interactive SD is false because SD requires that some physical events are caused by nonphysical events. However, I do not want to rest my case against SD on the strong principle of closure. Rather, I will rely on the much weaker principle that at the neural level and above there are no nonphysical causal influences.

Say that Harriet sees her friend and catches his attention by raising her arm. Her arm raising was brought about by the contraction of muscles in her arm and shoulder. Muscle contractions are physical events whose immediate causes are, as far as we can ascertain, prior physical events; in particular, the immediate cause is activity in the relevant motor neurons. That neural activity was caused by, as far as we can tell, activity in the motor cortex which in turn was caused by, as far as we can tell, further neural activity in other parts of the brain. Finally, this long chain of neural events that are, as far as we can tell, all physical, terminates in the activation of Harriet’s retina by light reflected off her friend.

Neuroscientists investigating the chain of events starting at Harriet’s retina and ending with her arm raising do not find physical events whose occurrence can’t be explained by other physical events. All the available evidence points to the chain of physical events being causally closed. Of course, neuroscientists may one day discover a class of neural events that, no matter how thoroughly they carry out their investigations, appear to have no physical causes. The weak principle of physical closure is an empirical hypothesis that is open to disconfirmation; nevertheless, it remains the case that all the available evidence points to its being true.

### 16.6.3 Explanatory weakness

In this subsection, I will advance what I take to be the most serious objection to SD – the observation that physicalist theories of the mind have vastly more explanatory power than SD.

Human minds have many features some of which can usefully be cataloged as follows.

- 1 *Perception*: We perceive the world in a variety of modes (visual, tactile, olfactory, etc.). Perception involves causal relationships between the world and mind. For example, light from objects in our visual field can cause changes in our beliefs. In addition, we are sometime (but not always) phenomenally conscious of our perceptions.
- 2 *Reasoning*: We reason in at least two modes. *Theoretical* reasoning carries us from one belief or set of beliefs to other beliefs; importantly, these processes are subject to a number of biases. *Practical* reasoning carries us from beliefs and desires to decisions to act thus and so. Sometimes we are conscious of both our reasoning and of the outputs of our reasoning, but very often we are not.
- 3 *Action*: We sometimes act on our decisions.
- 4 *Emotions*: We experience a wide range of emotions, and emotions are powerful motivators. Many – perhaps all – emotions are phenomenally conscious, although the claim that all emotions are phenomenally conscious is controversial.
- 5 *Coordination*: Even simple bodily movements require very sophisticated coordination that involves (among other processes) vision, reflexes, balance, and proprioception. In

addition, coordination very often involves feedback loops from perception to movement and back to perception.

- 6 *Memory*: We remember in a number of different modes. There are many different memory processes, including short and long term, and procedural memory. An important distinction is made between implicit and explicit memory; only the latter is available to consciousness.
- 7 *Language*: Language involves, among many other processes, a representation of the grammar of the language in question. Strikingly, these processes are not available to introspection. They are in Stephen Stich's useful phrase "subdoxastic" (Stich 1978).
- 8 *Brain damage*: It has long been known that brain damage can cause mental deficits. In many cases damage to specific parts of the brain brings about specific mental deficits. For example, damage to Broca's area leads to deficits in language processing.
- 9 *Mental disorder*: We are subject to a wide range of mental disorders including depression, schizophrenia, and autism.

For each of these features scientists have extraordinarily detailed theories, evidence, and explanations – explanations that make no appeal to nonphysical substances or properties. It would take millions of words to describe in detail the extraordinary work done by scientists on each of these features, so of necessity I will restrict myself to a few sentences about three features.

To begin with, we know a great deal about the psychology and neurobiology of perception in its various modes. For example, we have a good understanding of the parts of the brain involved in vision. Intriguingly, we are beginning to understand perceptual imagination in the various perceptual modalities, and have explored the significant overlaps between perceptual deficits and deficits of the imagination. For example, patients with hemispherical neglect, that is, they only attend to one side of their visual field, have a parallel deficit of visual imagination (the classic reference is Bisiach and Luzzatti 1978).

Consider next the emotions. We now know an enormous amount about the anatomy, physiology, and function of emotions. We know, for example, that there is a set of *basic* emotions that include happiness, sadness, and anger (Ekman 1992). Darwin identified a similar set of emotions in his *The Expression of the Emotions in Man and Animals* (Darwin 1962 [1872]) and proposed that they function as signals. The basic emotions have characteristic facial features, plus a suite of standard hormonal and cardiac features, and their functioning is at least partly independent of the person's beliefs: even if you believe the snake is harmless you may still be afraid of it.

Cognitive science has also told us a great deal about reasoning including the brain areas principally involved in the various kinds of reasoning; its development in children and decline in dementia; and its biases. For example, we tend to continue to hold onto our beliefs in the face of counterevidence ("belief-perseverance"). See, for example, Ross, Lepper, and Hubbard (1975).

Finally, consider our understanding of the psychology, pharmacology, and neuroscience of mental disorders. Our understanding of the brain structures and neurotransmitters involved in, for example, both depression and schizophrenia, is now considerable and some pharmacological relief is now available for people suffering these terrible disorders. This is not to deny that some forms of therapy are pitched at the level of beliefs, desires, and emotions. For example, in cognitive behavioral therapy (CBT) the patient is encouraged to examine her beliefs and modify or eliminate those that lead to problematic emotional

experiences or behaviors. But CBT's appeal to such states does not unseat the modest version of physicalism endorsed in this chapter; it only unseats eliminativist versions of physicalism (Ravenscroft 2009).

Now ask yourself what SD has contributed to our understanding of perception, reasoning, action, emotions, coordination, memory, language, brain damage, and mental disorder. *The answer is nothing.* There are no detailed substance dualist theories of any of these important features of our mental life in this sense: the hypothesis that mental states are nonphysical has done no work at all. Substance dualists have advanced no testable hypotheses, let alone confirmed them. They have made no important discoveries, let alone discoveries that might ameliorate human suffering. Substance dualism is a point in logical space where in principle a psychological theory could be constructed, and nothing more. Notice that this objection to substance dualism depends on neither scientism nor eliminativism. The claim is not that the only level at which perception, reasoning, emotions, and so on, can be understood is the neuroscientific level; rather, the claim is that the hypothesis that mental states are nonphysical has yet to contribute to our understanding of these phenomena.

In desperation a substance dualist might reach for the claim that SD is compatible with all the physical sciences of the mind that I have been lauding. The nonphysical mind is a shadowy puppeteer, pulling the strings of the brain. Reason, for example, with all its biases, is an action of the nonphysical mind that is expressed through the brain. But this move doesn't meet the objection. The substance dualist still needs to explain how the nonphysical mind could achieve all the outcomes that are (supposedly) expressed by the brain. The explanatory gap still yawns.

The astonishing chasm between the explanatory power of physicalism and that of SD is in my view the core objection to SD. SD has nothing to say about all the crucial questions, and is therefore of little value.

Before moving on, it is worth noting that the argument from explanatory power does not in any way rely on Ockham's Razor. At no point does the argument appeal to the greater simplicity of physicalism compared with SD and consequently any qualms one does or should have about Ockham's Razor do not impact the argument sketched here.

## 16.7 Analogy with the Design Argument

In the previous subsection, I argued for physicalism and against substance dualism on the grounds that the former has much greater explanatory power than the latter. The design argument is instructive because, on one interpretation, it too is an argument from the best explanation that purports to establish the existence of a nonphysical entity – God (Dawkins 1996; Sober 2000).

Interpreting William Paley's (1963 [1802]) design argument is difficult. It can be read as a simple argument from analogy, although as an argument from analogy it is, as Hume (2007 [1779]) argued, extraordinarily weak. An alternative reading of Paley is that he is presenting an argument from the best explanation, and I will pursue that idea here.

Paley begins by drawing attention to the way living organisms are well-adapted to their environments. He offers two hypotheses for adaptation:

Hypothesis 1. Adaptation in the natural world occurs by accident.

Hypothesis 2. Adaptation in the natural world is the work of God.

Paley argues convincingly that Hypothesis 2 is considerably more plausible than Hypothesis 1. Since Hypothesis 2 is the best explanation available, Paley concluded that God exists.

But Paley had the great misfortune of living before Darwin published *On the Origin of Species* (2008 [1859]) because Darwin offered a third hypothesis about adaptation:

Hypothesis 3. Adaptation in the natural world is the outcome of natural selection.

So assessing the design argument comes down to determining which hypothesis has the greater explanatory power: Hypothesis 2 or Hypothesis 3? (There is in fact a hidden premise in the argument that follows: it assumes that God is, in John Bishop's (2009) useful phrase, "omnigod"; that is, God is all-knowing, all-powerful, and all-good. If this conception of God – or something close to it – is rejected, the argument collapses.)

Richard Dawkins (1996) and Elliott Sober (2000) have argued that Hypothesis 3 has much greater explanatory power than Hypothesis 2. Their argument largely rests on observations of the natural world. To give just one example, the human eye has a "design" fault. (I have put "design" in quotation marks to avoid begging the question against Darwin. If the theory of natural selection is right, there is no design in the natural world.) The retina, which consists of light-sensitive cells, is at the back of the eye opposite the pupil. Information gathered by the retina is passed to the visual cortex by the optic nerve. One would expect the optic nerve would be "wired" into the back of the retina but, surprisingly, it is "wired" into the front. Consequently, the optic nerve must pass through the retina and so there are no light-sensitive retinal cells at that point, resulting in a blind spot. Clearly this arrangement is suboptimal, and that creates a problem for Hypothesis 2. Why would God structure our eyes so we have a blind spot, with the problems it entails?

So far we do not have much of an objection against the design argument; the real action begins when we ask to what extent Hypothesis 3 can explain the blind spot. Can it offer a better explanation? The answer is *yes*. Natural selection works by making tiny changes to existing structures. Changes that increase the ability of the organism to survive and reproduce tend to be preserved; changes that don't tend to be eliminated. If this process is repeated many times, organisms will tend to become better adapted to their environment. However, the process of selecting small changes is imperfect. Natural selection has to build on existing structures and that limits its capacity to develop optimal structures. It cannot recognize accumulated mistakes and go back and start again, which is why Dawkins called his book *The Blind Watchmaker* (1996). Strikingly, a series of organisms exist whose eyes occupy many of the numerous stages between a very primitive eye that consists only of a small number of light-sensitive cells through to much more sophisticated eyes like ours. (In passing, notice that the argument just sketched does not rely on Dawkins's controversial claim that people's actions are almost always motivated in ways that promote their own interest or that of their close kin; Dawkins 1976.) It is worth emphasizing that this is only one example of the explanatory advantages of Hypothesis 3 over Hypothesis 2. Many, many more examples can be found in the evolutionary biology literature.

Notice how the argument we have been considering in this section goes. It evaluates the explanatory power of three hypotheses about the origin of the human eye, and argues that Darwin's hypothesis (Hypothesis 3) has the most explanatory power and therefore should be accepted. Notice also that both Paley's and Darwin's arguments rely on empirical data. These are scientific investigations, not armchair conjectures. The arguments in the previous

subsection evaluated the explanatory power of SD and physicalism, and concluded that physicalism is the better theory. There too the decisive factor was empirical data. The design argument and substance dualism fail for the same reason: neither can explain what's going on.

## 16.8 Concluding Remarks

What's striking about SD is that it is a theory in only the most attenuated sense of the word. Substance dualists have nothing to say about the components of the nonphysical mind, or about how those components are structured into a working whole. SD has no account of reasoning, of perception, or of the coordination of action. It is silent on mental illness and on memory, emotions, dreams, and imagination. SD offers nothing on psychological development in children or cognitive decline in the elderly. It makes no serious, testable predictions on any aspect of human cognition. Worse still, it's not clear how any of these failings could be addressed. What kind of investigations would yield information about the components and structure of the nonphysical mind? How would a research program in substance dualist psychology even begin?

It would be absurd to say that SD is certainly false, but it would also be absurd to regard it, on the basis of current evidence, as a serious option in the study of mind-brain relations.

## References

- Baker, M. 2011. "Brains and Souls; Grammar and Speaking." In *The Soul Hypothesis: Investigations into the Existence of the Soul*, edited by M. Baker and S. Goetz, 73–93. New York: Continuum.
- Baker, M., and S. Goetz. 2011. "Introduction." In *The Soul Hypothesis: Investigations into the Existence of the Soul*, edited by M. Baker and S. Goetz, 1–20. New York: Continuum.
- Bishop, J. 2009. "Towards a Religiously Adequate Alternative to Omnigod Theism." *Ratio*, 48: 419–433.
- Bisiach, E., and C. Luzzatti. 1978. "Unilateral Neglect of Representational Space." *Cortex*, 14: 129–133.
- Braddon-Mitchell, D., and F. Jackson. 1996. *Philosophy of Mind and Cognition*. Malden, MA: Blackwell.
- Broad, C. 1925. *The Mind and its Place in Nature*. London: Routledge and Kegan Paul.
- Chalmers, D. 1996. *The Conscious Mind: In Search of a Fundamental Theory*. Oxford: Oxford University Press.
- Churchland, P. 1981. "Eliminative Materialism and the Propositional Attitudes." *Journal of Philosophy*, 78: 67–90.
- Churchland, P. 1985. "Reduction, Qualia and the Direct Introspection of Brain States." *Journal of Philosophy*, 82: 8–28.
- Clark, A., and D. Chalmers. 1998. "The Extended Mind." *Analysis*, 58: 10–23.
- Collins, R. 2011. "The Energy of the Soul." In *The Soul Hypothesis: Investigations into the Existence of the Soul*, edited by M. Baker and S. Goetz, 123–133. New York: Continuum.
- Darwin, C. 2008 [1859]. *On the Origin of Species*. Oxford: Oxford University Press.
- Darwin, C. 1962 [1872]. *The Expression of the Emotions in Man and Animals*. Chicago: Chicago University Press.
- Davidson, D. 1980 [1970]. "Mental Events." In *Essays on Actions and Events*, 207–224. Oxford: University Press.
- Dawkins, R. 1976. *The Selfish Gene*. Oxford: Oxford University Press.
- Dawkins, R. 1996. *The Blind Watchmaker*. New York: W. W. Norton.

- Dennett, D. 1991. *Consciousness Explained*. New York: Little, Brown.
- Ekman, P. 1992. "An Argument for Basic Emotions." *Psychological Review*, 99: 550–553.
- Elisabeth, Princess of Bohemia, and R. Descartes. 2007. *The Correspondence between Princess Elisabeth of Bohemia and René Descartes*, edited and translated by L. Shapiro. Chicago: Chicago University Press.
- Flanagan, O. 1991. *The Science of the Mind*. Cambridge, MA: MIT Press.
- Fodor, J. 1974. "Special Sciences." *Synthese*, 28: 97–115.
- Fodor, J. 1994. "The Mind-Body Problem." In *The Mind-Body Problem*, edited by R. Warner and T. Szubka. Oxford: Blackwell.
- Frankfurt, H. 1971. "Freedom of the Will and the Concept of a Person." *Journal of Philosophy*, 68: 5–20.
- Hasker, W. 2011. "Souls Bestly and Human." In *The Soul Hypothesis: Investigations into the Existence of the Soul*, edited by M. Baker and S. Goetz, 202–217. New York: Continuum.
- Horgan, T., and J. Woodward. 1985. "Folk Psychology is Here to Stay." *Philosophical Review*, 94: 97–226.
- Hume, D. 1975 [1777]. *An Enquiry Concerning Human Understanding*. Oxford: Clarendon Press.
- Hume, D. 1994 [1777]. "Of Superstition and Enthusiasm." In *Essays: Moral, Political, and Literary*, edited by W. Todd. Indianapolis, IN: Liberty Fund.
- Hume, D. 2007 [1779]. *Dialogues Concerning Natural Religion*. Cambridge: Cambridge University Press.
- Jackson, F. 1982. "Epiphenomenal Qualia." *Philosophical Quarterly*, 32: 127–136.
- Jackson, F. 1986. "What Mary Knew." *Journal of Philosophy*, 83: 291–295.
- Jackson, F. 1998. *From Metaphysics to Ethics: A Defense of Conceptual Analysis*. Oxford: Oxford University Press.
- Jackson, F. 2007. "The Knowledge Argument, Diaphonousness, Representationalism." In *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*, edited by T. Alter and S. Walter. Oxford: Oxford University Press.
- Jackson, F. 2009. "Replies." In *Minds, Worlds and Conditional: Essays on the Philosophy of Frank Jackson*, edited by I. Ravenscroft. Oxford: Oxford University Press.
- Jackson, F., and P. Pettit. 1990. "In Defense of Folk Psychology." *Philosophical Studies*, 59: 31–54.
- Kim, J. 1999. "Making Sense of Emergence." *Philosophical Studies*, 95: 3–36.
- Lewis, D. 1983. "New Work for a Theory of Universals." *Australian Journal of Philosophy*, 61: 343–377.
- McLaughlin, B. 1997. "Emergence and Supervenience." *Intellectia*, 25: 25–43.
- Mill, J. 1843. *System of Logic*. London: Longmans, Green, Reader and Dyer.
- Nagel, T. 1974. "What Is It Like to Be a Bat?" *Philosophical Review*, 83: 435–450.
- Paley, W. 1963 [1802]. *Natural Theology*. Indianapolis, IN: Bobbs-Merrill.
- Papineau, D. 2002. *Thinking about Consciousness*. Oxford: Oxford University Press.
- Ravenscroft, I. 2005. *Philosophy of Mind: A Beginner's Guide*. Oxford: Oxford University Press.
- Ravenscroft, I. 2009. "Is Folk Psychology a Theory?" In *The Routledge Companion to Philosophy of Psychology*, edited by John Symons and Paco Calvo, 131–147. Abingdon, UK: Routledge.
- Ravenscroft, I. 2016. "Folk Psychology as a Theory." *Stanford Encyclopedia of Philosophy*, edited by E. Zalta. Accessed June 29, 2017. <http://plato.stanford.edu/entries/folkpsych-theory/>.
- Robinson, D. 2011. "Minds, Brains, and Brains in Vats." In *The Soul Hypothesis: Investigations into the Existence of the Soul*, edited by M. Baker and S. Goetz, 46–67. New York: Continuum.
- Ross, L., M. Lepper, and M. Hubbard. 1975. "Perseverance in Self-Perception and Social Perception: Biased Attributional Processes in the Debriefing Paradigm." *Journal of Personality and Social Psychology*, 32: 880–892.
- Russell, B. 1912. "On the Notion of Cause." *Proceedings of the Aristotelian Society*, 13: 1–26.
- Smart, J. 1959. "Sensations and Brain Processes." *Philosophical Review*, 68: 141–156.
- Smith, M. 1994. *The Moral Problem*. Malden, MA: Wiley-Blackwell.

- Sober, E. 2000. *Philosophy of Biology*. Oxford: Oxford University Press.
- Stich, S. 1978. "Beliefs and Subdoxastic States." *Philosophy of Science*, 45: 499–518.
- Stich, S. 1983. *From Folk Psychology to Cognitive Science: The Case Against Belief*. Cambridge, MA: MIT Press.
- Taliaferro, C. 2011. "The Soul of the Matter." In *The Soul Hypothesis: Investigations into the Existence of the Soul*, edited by M. Baker and S. Goetz, 26–40. New York: Continuum.
- Wilson, T. 2002. *Strangers to Ourselves*. Cambridge, MA: Harvard University Press.

### Further Reading

- Descartes, R. 1996 [1641]. *Meditations on First Philosophy*, translated by J. Cottingham. Cambridge: Cambridge University Press.
- Lewis, D. 1973. "Causation." *Journal of Philosophy*, 70: 556–567.

## PART II

# Alternatives to Substance Dualism





# Why Should a Christian Embrace Materialism (about Human Persons)?

KEVIN CORCORAN

Why should a Christian embrace materialism about human persons? There are both short and long answers to this question. One short answer is that I think Christians, like everyone, should hold views that are *true*, and it seems to me that materialism about human persons is the *truth* about our nature. And that's why Christians should embrace materialism about human persons.<sup>1</sup> A longer answer is a bit more complicated, of course, as it involves providing *reasons* for believing that materialism about human persons is true. In the pages that follow, I aim to deliver a longer answer, to say why I, as a Christian, hold a materialist view of human persons, and why I think you should, too.

There are three terms in the title to this chapter that cry out for definition. Those terms are *materialism*, *human*, and *person*. Let me begin by first saying what I understand by each of these terms and then go on to say, as simply as I can, what I mean by *a materialist view of human persons*. That out of the way, I will in subsequent sections say just why I hold such a view. In the final section, I consider what many believe to be a crippling problem for a materialist view of human nature, and I show why actually it isn't a problem at all for the view.

## 17.1 Defining our Terms

Let's begin with "materialism." While often used interchangeably, I want to distinguish between *materialism* and *physicalism*. Let us understand "materialism" to be a claim about the metaphysical *composition* of human persons. And let us understand "physicalism" to be a claim about the nature of *consciousness*, *subjectivity*, and *the mental*. According to such a distinction, *materialism* about human persons is a view according to which human persons are *wholly physical* entities, *exhaustively composed* of physical particles (the sorts of entities

it is the job of physicists to investigate) and, as such, human persons are neither identical to *immaterial* souls nor do they have *immaterial* souls as parts.

Physicalism, by contrast, is a view according to which one's *compositional* nature *exhausts her entire nature*. In other words, according to physicalism (as understood here) there are no human person facts over and above compositional and related structural facts. Once a human person's compositional nature is fixed, every other fact about her (including all mental facts) is entirely fixed, without remainder. Physicalism – so understood as a claim about the mind and the mental – is *reductive*. It is the view that phenomenal consciousness – all the “*what it's like*” features of consciousness and subjectivity in general – is *entirely* physical; that is, *all* facts about phenomenal consciousness *are* physical facts. On such a view, phenomenal consciousness and first-person subjectivity are *reducible, without remainder*, to “facts about structure and function.” When Francis Crick says, “your joys and your sorrows, your memories and your ambitions . . . are in fact no more than the behavior of a vast assembly of nerve cells and their associated molecules” (Crick 1994, 3), Crick expresses a physicalist view of the mind and the mental.

I reject physicalism (about consciousness and the mental) but embrace materialism (about the composition of human persons). That is to say, I believe that human persons are *entirely* composed of physical stuff, but deny the claim that there are no human person facts over and above compositional and related structural facts.

Now, what about the terms *human* and *person*? Here again, these terms are often used interchangeably. And here again, I want to distinguish between the two. By “human,” or more precisely *human being*, I simply mean an animal of the species *Homo sapiens* (or *Homo sapiens sapiens*). If something is a human being, then it has a certain animal ancestry. Now by *person*, I mean an entity with a *first-person perspective* or the capacity for a first-person perspective, that is, a capacity to think of oneself *as oneself*, without the need for a name or description. When I think of myself as “Kevin Corcoran” or “the only Irishman in the Calvin College philosophy department,” or “the biological father of Shannon and Rowan Corcoran,” I am *not* thinking of myself in the first-person way. After all, I might, tragically, *forget* my name is Kevin Corcoran and *forget* that I teach at Calvin or even *forget* that I have children. When I think such thoughts as these, however, “I wonder if I will live long enough to see my grandchildren grow up and marry?” or “I wonder whether I will ever finish writing this chapter?,” I think of myself in the first-person way. For while I could be mistaken about the fact that I'm Kevin Corcoran or teach at Calvin or have children and grandchildren, I can't be mistaken that it's *me* that I'm wondering about when I wonder whether *I* will live long enough to see *my* grandchildren grow up and marry.

Putting these two ideas together, something is a *person* if, and only if, it possesses a capacity for a first-person perspective. And in order for something to be a *human person*, it must have a biological body, that is, an organism of the species *Homo sapiens*, that it, and it alone, can refer to in a first-person way, such as when one thinks, “I wonder if this headache of mine will ever go away.”

With these definitions stipulated, we can say the following: while everything that's a human *person* is human, by virtue of having a biological body of the species of *Homo sapiens* that it, and it alone, can refer to in the first-person way, it's not the case that everything that's a human *being* is a human *person*. A human being (i.e., an animal of the species *Homo sapiens*) lacking *all capacity* for a first-person perspective would, on this understanding, be a human, *nonperson*. Note that in Corcoran (2003) I argue that there are no untoward ethical

implications for a view of human persons according to which some humans are nonpersons.

What would be a concrete example of something that satisfies the identity criteria for *human*, but not for *person*, and so would be, on this understanding of things, a human, *nonperson*? Perhaps someone like Terri Schiavo at the end of her life, whose cerebral cortex had completely atrophied, as graphically illustrated in her CT scan. We know enough about consciousness in human animals to know that absent a cerebral cortex the individual in question will lack the neural complexity required to underwrite consciousness, let alone a first-person perspective. For we know enough about consciousness in humans to know that without a cerebral cortex there *is* no consciousness; and without consciousness, as I say, there is no *first-person consciousness*.

So far, everything I've said about human personhood is consistent with *animalism* about human persons (the view held by such philosophers as Eric Olson and Peter van Inwagen) and a *constitution account* of human persons (the view held by Lynne Baker and myself). If one is an animalist about human persons, then one will say that the thing that is a person is identical to the thing that is a human animal, and personhood is a *contingent* property of that thing. In other words, the thing that is a human animal can exist without being a person; it may be a person during some intervals of its existence (the interval between its 10th and 30th year, for example) and not during others (when it's a 10 millimeter large embryo or a 150 pound adult *sans* a cerebral cortex, for example). And if one is a constitution theorist about human persons, then one will say that human persons are wholly *constituted by* without being *identical to* human animals. But on both an animalist view and a constitution view, human *nonpersons* are not only *conceptually* possible but, perhaps sadly, sometimes actual.

By "a materialist view of human persons," therefore, I mean any view according to which a human person's *compositional nature* is exhausted by physical entities, regardless of whether the view takes human persons and human animals to stand in an identity relation (à la animalism) or a constitution relation (à la constitution view).

## 17.2 Reasons for Believing a Materialist View of Human Persons

### 17.2.1 *An aesthetic reason*

Now, why do I think materialism is the truth about our nature? I suppose the first reason for thinking that materialism is the truth about our nature is that to suppose otherwise, to believe (say) in some form or other of *substance* dualism, is to introduce an unnecessary and inelegant cleavage into the natural world. That is not something I would expect *within* creation. In other words, that human persons are not of a piece with the rest of nature, which would be the case if human persons of all the animals are the only creatures endowed by God with immaterial souls, strikes me as odd. Granted, this is a kind of aesthetic argument, but, I would expect the natural world, as a product of divine creation, to be elegant and seamless with respect to its inhabitants' compositional nature.

### 17.2.2 *A biological/developmental reason*

Materialism about human persons fits what we know about the slow, gradual emergence and development of increasingly complex living things, as well as what we know about the

gradual development of consciousness itself within biological organisms. We know, for example, that new conscious capacities emerge as brain structure develops in complexity. The slow, gradual development of increasingly complex conscious capacities fits the picture we get from nature. If what dualists say about the soul is true – that it's a *nonphysical, partless* entity and that *it* is the bearer of psychological properties and states of human consciousness – then, it seems difficult, if not impossible, to account for the *gradual* development of conscious capacities, as it's hard to see how a *partless* thing can develop at all. In other words, it would seem that the soul that is *you* must have had the capacity to engage in the solving of differential equations from its earliest days as the soul of the embryo that became your body. Yet we know that the neural structure of an embryo lacks the kind of complexity necessary to underwrite the sort of cognitive sophistication one needs to think about let alone solve differential equations. At the very least, the gradual development of conscious capacities, in tandem with the gradual development of complexity in neural structures in biological brains, seems to provide a *prima facie* challenge to dualism and, at the same time, a *prima facie* reason to think that materialism about human persons is true.

### 17.2.3 *Evidence from the neurosciences*

Materialism is a better fit with evidence from the neurosciences than dualism and this makes materialism, in my view, a more plausible theory of human persons than dualism. More specifically, there's a tension between dualism and the neural *dependence* of our consciousness and mental life.

Suppose that the dualist is right and the subject of conscious experience (in both its cognitive and phenomenal aspects) is a *nonphysical* substance devoid of any mereological structure. Then while we would expect a fairly high degree of some kinds of dependence (e.g., dependence of the soul on the brain for sensory input), we would not expect to find the sort of radical, causal dependence of our conscious lives on the physical that we do in fact find. The point is best appreciated by considering the *fine-grained* dependence that examples from the neurosciences illustrate. While conscious experience may seem to be seamless and unified, the examples we are about to consider reveal a fairly high degree of *structural* complexity in which distinct aspects of experience are implemented in anatomically distinct neural structures. For example, a condition known as blindsight (considered in section 17.2.3.1) reveals that visual processing involves *two* pathways, only one of which is conscious, that control different aspects of visual perception. The fact that damage to a *specific* part of the brain, in this case the primary visual cortex, results in a highly selective visual impairment, rather than the *total* loss of all capacity for visual processing, suggests that different parts of the visual system depend on distinct neural structures. This, in conjunction with other cases, suggests an overall picture in which consciousness and cognition rest on a structurally complex architecture that maps onto a neural structure that is at least equally complex, in which each element of our mental architecture is mapped onto a different neural structure on which it depends.

The upshot is that we would not expect to find this sort of systematic, fine-grained dependence of consciousness and experience on patterns of neural activity if dualism were true. Why? Well, if the dualist were correct, then not only would all conscious experience take place within a nonphysical subject, but additionally the kind of neural activity mapped out by the neurosciences should be explanatorily *irrelevant* to its occurrence and character. This supposed explanatory irrelevance of the physical for the features of consciousness is

the driving force of a typical anti-materialist argument. Here's Leibniz's famous statement of that argument:

We must admit that perception, and whatever depends on it, cannot be explained on mechanical principles, i.e. by shapes and movements. If we pretend that there is a machine whose structure makes it think, sense and have perceptions, then we can conceive of it enlarged, but keeping to the same proportions, so that we might go inside it as into a mill. Suppose that we do: then if we inspect the interior we shall find there nothing but parts which push one another, and never anything which could explain a perception. (Leibniz 1989 [1714], 213–224; see also Plantinga 2006, 2007)

Thus, according to Leibniz and anti-materialism, perception must be sought in *simple* substance, not in what is composite. But if complex patterns of neural firings were explanatorily irrelevant in a way that demanded a nonphysical subject of conscious experience, then – in the absence of strong reasons to the contrary – we'd expect conscious experience to be nomologically independent of such neural activities. But conscious experience is precisely *not* nomologically independent of neural activity. So in the absence of strong reasons to expect otherwise, the fine-grained, neurobiological dependence of consciousness is strongly suggestive of a materialist view of human persons.

#### 17.2.3.1 Blindsight

Blindsight is a fascinating condition that renders patients who suffer from it able to “see” without *consciously* seeing (Ramachandran and Blakeslee 1998; Ramachandran 2004). A patient suffering from blindsight, under experimental conditions, can be shown (say) a spot of light in their “blind” field of “vision” and asked “what is it you see?” They will answer “nothing.” Which is exactly what we would expect. But if asked to point to the spot of light, they can do it. If the object is moving up or down, left or right, *in the blind field* they can tell you with near perfect accuracy which direction the “unseen” object is moving. Similarly, patients suffering from blindsight can reach for objects, state whether a stick is oriented vertically or horizontally, and even catch a ball thrown toward them – all in their blind field of vision. Most astonishing, they do this with remarkable accuracy even while insisting that they are merely guessing.<sup>2</sup> But how can that be? How can someone “see” without seeing?

The answer lies in the neuroanatomy of vision. It's long been known that damage to the primary visual cortex (a part of the occipital lobe devoted to conscious visual processing) results in blindness. For example, damage to the right primary visual cortex results in blindness in the left field of vision. This damage inhibits the functionality of the visual pathway running from the retina, through the thalamus to the primary visual cortex and from the primary visual cortex to a cluster of thirty or so highly specialized visual processing areas, for example, V4 is devoted to color perception and MT to perception of motion. As this pathway specializes in processing aspects of conscious visual perception, you need it in order to see something *consciously*. But there is a second, evolutionarily older, visual pathway that subserves a different aspect of vision and runs from the retina to the superior colliculus in the brain stem and from there to the parietal lobe (a higher cortical area concerned with spatial representation and spatial navigation). This older pathway is involved in reflexive behavior and orienting toward something important in your visual field (perhaps something that threatens your continued

existence). This second pathway is found in birds and reptiles (nonhuman animals which some believe do not enjoy consciousness). It is the higher centers of bird, reptile, and human brains that substantially differ and that accounts for why the one (humans) and (likely) not the other (reptiles and birds) enjoys consciousness.

In blindsight part of the newer pathway, the primary visual cortex, is damaged. Therefore blindsight patients do not *consciously* see anything. But the older pathway, the one that subserves the “fight or flight” reflex and that’s responsible for object location and orientation in a visual field, remains intact. An object in the blind field of vision activates this older pathway allowing the patient to locate and respond to it. Since the older pathway is not conscious, it can process visual stimuli without the blindsighted individual being consciously aware of those stimuli. It is in virtue of this evolutionarily older pathway that blindsight patients are able nonconsciously to “see” the direction of objects they do not “see” consciously.

The kind of dependence of consciousness on brain activity illustrated in the phenomenon of blindsight motivates another reason to believe that a materialist view of human persons is the truth about our nature, and not some version of dualism. For the case of blindsight illustrates how certain remarkable features of consciousness are explainable in terms of neuroscience and so explaining facets and features of consciousness is achieved without any need to invoke a nonphysical soul. Many other features of consciousness too, from the ability to discriminate and react to environmental stimulæ and the ability to access our own internal conscious states, to the focus of attention and the difference between wakefulness and sleep, to the very existence of human emotions, *all* are in fact *explainable* in terms of brain structures and neural activity. In other words, science explains these features of consciousness without invoking a nonphysical soul. So there is, at least with respect to all of these aspects of conscious experience, no explanatory work for a nonphysical soul to do. The steady march of progress and explanatory successes already achieved in the neurosciences and its cognate disciplines, coupled with the *continued* progress and successes we can expect in the future, leaves less and less room for a nonphysical soul. The ongoing success of the neurosciences is squeezing the soul out of the explanatory picture and providing more and more evidence for a materialist view of human persons.

### 17.2.3.2 Phantom limb

In the past ten to fifteen years there has been a lot of talk about neural plasticity. During most of the twentieth century the consensus was that brain structure was pretty much set and immutable after some specified critical period of childhood development. More recent research, however, shows that while the brain is not endlessly plastic, it is malleable within limits or boundaries. Within certain circumscribed boundaries, our brains can, indeed, be “re-wired.” In fact, experience can change not just brain structure or anatomy but also the functional organization or physiology of the brain. Connections or neural pathways within the brain are removed or preserved almost entirely based on use or nonuse. Not only that, but neighboring neurons that fire together or simultaneously, often fuse to make a new pathway within the network. Likewise, neighboring neurons that do not fire together form independent neural routes.

A vivid example of brain plasticity is found in phantom limb patients. These are patients who have lost a limb but continue to feel the presence of their missing limb and experience

pains, itches, and other sorts of sensory stimulation in it. For example, one patient who was missing a right arm continued to feel it itch whenever he shaved the left side of his face. Other patients feel cold sensations in their phantom limb when an ice cube touches their face and still others feel water trickle down their missing arm when a drop of water trickles down their face (Ramachandran and Blakeslee 1998). But how is this possible? How can someone feel an itch in their right arm when they don't have a right arm? How can someone feel water trickle down their right arm when, again, they don't have a right arm? How can touching a person's cheek result in a sensation in a limb that's been missing for a decade?

In a series of experiments performed during brain surgeries using only local anesthesia, Wilder Penfield discovered that stimulating the post-central gyrus, a narrow vertical region of cortex in the parietal lobe, produces sensations in different parts of the body (Ramachandran and Blakeslee 1998). Penfield discovered that this strip of cortex contains an entire neural map, called the *Penfield homunculus*, of our whole body. As it turns out, the location of our right arm on this map is adjacent to that of the face. When the patient loses a limb, the cortical region corresponding to the right arm region of the map no longer receives stimulation as the limb is missing and is no longer sending it signals. The neurons in this understimulated cortical region are so "hungry" for sensory input that they begin over-reaching their borders and extending their way into the adjacent region, that of the face, where they receive signals and stimulation. These signals are then interpreted as being sent from the hand as well as the face. Thus, sensory stimulation originating in the face results in the stimulation of both the facial and hand regions of the Penfield map and this is why when the patient shaves (or stimulates that area) he also feels it as an itch on his phantom arm.<sup>3</sup> Neural plasticity makes limited neural remapping possible and this remapping explains the felt location of sensations as if they are occurring in a missing limb.

Phantom arm and hand phenomena are explained in terms of a facial tactile stimulation sending a sensory message to adjacent areas in the Penfield map. The felt location of a sensation is explained by stimulation of the part of the sensory cortex corresponding to that location in the Penfield map. Similar explanations are given for other features of conscious experience involving phantom arms or hands, for example, the type of sensation experienced (i.e., whether it's an itch, a vibration, or the feeling of a trickle of water) and its intensity – all without positing a nonphysical soul. The lesson to be drawn, I believe, is that a nonphysical soul plays no role in explaining these features of consciousness. And the fact that more and more aspects of consciousness are being explained in terms of neurophysiology provides ever more optimism that experience itself will one day yield to explanation; explanation wholly in terms of principles and laws that are consistent with everything else we know about the natural world and the workings of the brain. Neurology, neurobiology, and the neurosciences, therefore, threaten to screen off any contribution for the soul to make in explaining consciousness.

Based on the above and similar cases in the neuroscience literature, I think we can safely say that dualism is a less plausible thesis than materialism when it comes to an explanation of consciousness. In fact, I think the neuroscientific evidence is sufficient to make it doubtful that an immaterial soul has any role to play in an explanation of consciousness. The more we learn about the neural correlates of consciousness, the more explanatorily irrelevant a nonphysical soul becomes and, correspondingly, the more reason we have to believe a materialist view of human persons.



### 17.3 The Hard Problem of Consciousness and a Materialist View of Human Persons

I can imagine someone raising at this point what at first sight may seem a formidable problem for a materialist view of human persons, and that is the so-called *hard problem* of consciousness.

The hard problem of consciousness, unlike the preceding “easy” problems, is just the sheer fact of phenomenal consciousness itself. How is it that the complicated interaction of one hundred billion nerve cells and their several hundred trillion synaptic connections should be accompanied by phenomenal experience? How is it that something inherently *subjective* like a first-person perspective can be explained in terms of *objective*, physical mechanisms? There is, after all, absolutely nothing about the biochemical properties of human neural structures that even begin to suggest the accompanying presence of the qualitative character and subjectivity of phenomenal experiences. Nor is there anything about the biochemical properties of human neural structures to ground a reductive explanation of consciousness in terms of neural mechanisms and their biochemical properties, nothing that makes it intelligible how features of the latter sort are responsible for features of the former sort. Any such attempt is bound to leave out what is distinctive of phenomenal consciousness. Doesn’t this fact lead to the conclusion that it is *dualism*, not materialism, about human persons that is the truth about our nature?

In a word, no. And here’s why. What, exactly, do we expect an account of, an explanation of, consciousness to do? Ought it to say *how* phenomenal consciousness arises out of complex neural circuitry? Ought it to say *why* *those* features of external, physical objects get unified and tagged with *those* phenomenal markers? What is it to *explain* some phenomenon anyway? What is it to give an *account of* some phenomenon?

According to one understanding of what’s involved in an explanation, there are two components of peculiarly *scientific* explanations, an *explanandum* (that which needs explaining) and an *explanans* (that which does the explaining, or accounts for the *explanandum*). On one historically influential account, the *explanans* involves specifying a set of deterministic laws and initial conditions from which the *explanandum* is logically deduced or derived. The derivation of the *explanandum* from the *explanans* contributes to our understanding of why the *explanandum* occurred and makes its occurrence intelligible.

Another model of explanation proceeds by identifying the lower-level mechanism that explains the presence of a higher-level feature. When the lower-level mechanism successfully explains the presence of the higher-level feature without remainder, we can consider the higher-level feature reductively explained. Consider *brittleness*. Brittleness, Michael Tye points out, is caused by “the irregular alignment of crystals” and, as a result of this type of alignment, the forces that hold the crystals together are weak. That is why brittle pieces of glass shatter easily (Tye 2007). The problem for physicalism, as we defined it at the start, is that it is difficult to see how any *objective* mechanism (by itself) could explain the *subjective* character of phenomenal states: even if we understood *all* of the fine-grained structures and chemical changes associated with the mechanisms of our brains, there is something that still requires explaining, namely, *why* and *how* this collection of neural and/or chemical changes produces that subjective feeling, or any subjective feeling at all. And this is *the hard problem* of consciousness.

The underlying assumption here is that if there were in fact a *complete* explanation of consciousness in terms of the proposed reducing mechanism, there'd be no room for questions like, "but how does *this* neural structure produce *that* subjective feeling?" For in the case of brittleness – a paradigm case of mechanistic explanation – the reducing mechanism closes off exactly this sort of question. Once we fully understand the nature of the lower-level reducing mechanism, for example, irregular alignment of crystals and the resulting weak bonding relations, we have an intelligible link between the reducing mechanism and the reduced property and it is inconceivable that the higher-level property fails to be instantiated (given the reducing mechanism). This a priori constraint on successful mechanistic explanations is most clearly stated by Joseph Levine (2004), who focuses on a mechanistic explanation of why the boiling point of water is 212°F at sea level: "Given a sufficiently rich elaboration of the [chemical composition of water and the behavior of H<sub>2</sub>O molecules when their average kinetic energy increases] it is inconceivable that H<sub>2</sub>O should not boil at 212°F at sea level" (Levine 2004, 79).

Turning again to phenomenal consciousness, since any neural mechanism *can* be conceived of *in the absence* of first-person phenomenal reality, it will *never* be the case that a complete understanding of a proposed reducing mechanism will render it inconceivable that a creature with the relevant neural structure fail to possess a first-person perspective and a capacity for phenomenal consciousness. As Levine put it:

No matter how rich the neurophysiological story gets, it still seems quite coherent to imagine that all that should be going on without there being anything it's like to undergo the states in question. (Levine 2004, 79)

Thus, if a reductive explanation of consciousness requires closing the explanatory gap and closing that gap requires positing a mechanism that makes it inconceivable that the higher-level property be instantiated (given the proposed mechanism), then when it comes to consciousness, the explanatory gap is unclosable *in principle*. And assuming that physicalism requires a *reductive* explanation of exactly the sort excluded by the explanatory gap, it follows that physicalism is a failure.

Did I not just admit that subjectivity and phenomenal consciousness *cannot* be explained? Whether subjectivity can be explained depends on what we mean by explain. I said at the beginning of this chapter that I reject physicalism about the mind and the mental, while embracing materialism about human persons. So it should come as no surprise that I deny that the sheer fact of subjectivity can be explained in ways analogous to the brittleness of glass and the boiling point of water at sea level. In other words, I deny that subjectivity is susceptible to a reductive, mechanistic explanation that renders the instantiation of higher-level properties intelligible in terms of lower-level mechanisms.

It is important to note, however, that this does not imply that consciousness can't be explained. What it implies is that it can't be explained *reductively*. I think consciousness *can* be explained. But, contrary to physicalism, I take the first-person perspective and phenomenal consciousness as *fundamental*. That there is a first-person perspective and that phenomenal consciousness exists is undeniable. How objectively existing third person observable phenomena such as the wetware and functioning of the human brain should produce such a phenomenon as a singular, subjective point of view not only has not yet been reductively explained, but its explanation may lay forever beyond our grasp. For it is plausible to believe that we are constitutionally incapable of ever providing an explanation

of phenomenal consciousness in terms of neural structures and mechanisms, of saying just how those objectively existing elements – neurons and synapses, and networks of them – give rise to phenomenal consciousness and a singular subjective experiencer. Colin McGinn (1989) believes exactly that – that how consciousness emerges from the wetware of the human brain may lie forever beyond our grasp. If that is right, then reductive explanations of phenomenal consciousness in terms of neural circuitry and the like are bound to fail, *in principle*.

While clearly there have been enormous gains and impressive strides over the past few decades when it comes to “explaining consciousness,” we must be clear about what, exactly, we’re explaining when we claim we’re explaining consciousness. What we’re doing is uncovering the neural mechanisms and structures implicated in specific facets and features of conscious experience. We can put it, if you like, in terms of uncovering the physical grounds of first-person, subjective experiences, but we’re never saying *how* these physical grounds (i.e., mechanisms or structures) are paired with those particular phenomenal features of a conscious experience. That is to say, while the cognitive and neurosciences are making great gains in discovering the physical mechanisms that underlie our conscious mental lives, they are not providing experimental data that will close the “explanatory gap,” despite what some optimistic physicalists and neuroscientists may claim.

## 17.4 Conclusion

To sum up: the hard problem of consciousness is a problem for a materialist view of human persons *only if* a materialist view of human persons entails physicalism about the mind and the mental. But it doesn’t. Phenomenal consciousness is an indisputable feature of the natural world and there is an ineliminable and irreducible first-person perspective to phenomenal consciousness. This much is granted. While perhaps it’s incumbent upon materialism to provide *some kind* of explanation of consciousness, I’ve been suggesting that the kind of explanation required has no *reductive* implications and so the failure of *physicalism* to explain the sheer fact of phenomenal consciousness is no failure at all for a *materialist* view of human persons. The failure of physicalism concedes nothing to dualism. It is, in fact, exactly what we’d expect if we embrace *materialism* about human persons, but reject the reductionistic implications of physicalism.

## Notes

1. When I first thought about writing this chapter, and my charge in writing it – “we want you to write a chapter in which you say why Christians should be materialists about human persons,” – my first thought was, “well, I don’t know that Christians *should be* materialists, but I think they *can be*.” But my friend Paul Manata reminded me that I think materialism about human persons is *true* and surely I want Christians to believe what’s true. Hence, the short answer to the question.
2. See Weiskrantz (2009) for a detailed discussion of the empirical issues and summary of research. Accessible discussions can also be found in Ramachandran and Blakeslee (1998) and Ramachandran (2004, 2012). Varying assessments have been made of the impact of blindsight on a philosophical theory of consciousness (Block 1995; Dennett 1991; Tye 1995).
3. Ramachandran discovered a map of a patient’s entire hand on his face as well as a second map of the hand on the patient’s arm a few inches above the amputation line. The second map is also a result of neural remapping, as the hand region of the Penfield map is adjacent to both the facial and upper arm regions (Ramachandran and Blakeslee 1998, 29).

## References

- Block, Ned. 1995. "On A Confusion About a Function of Consciousness." *Behavioral and Brain Sciences*, 18: 227–247.
- Corcoran, Kevin. 2003. "Material Persons, Immaterial Souls and an Ethic of Life." *Faith and Philosophy*, 20(2): 218–228.
- Crick, Francis. 1994. *The Astonishing Hypothesis*. New York: Charles Scribner's Sons.
- Dennett, Daniel. 1991. *Consciousness Explained*. New York: Back Bay Books.
- Leibniz, Gottfried Wilhelm. 1989 [1714]. "Monadology,." In *Philosophical Essays*, edited by Roger Ariew and Daniel Garber. Indianapolis, IN: Hackett.
- Levine, Joseph. 2004. *Purple Haze: The Puzzle of Consciousness*. Oxford: Oxford University Press.
- McGinn, Colin. 1989. "Can We Solve the Mind-Body Problem?" *Mind*, 98(391): 349–366.
- Plantinga, Alvin. 2006. "Against Materialism." *Faith and Philosophy*, 23: 3–32.
- Plantinga, Alvin. 2007. "Materialism and Christian Belief." In *Persons: Human and Divine*, edited by Peter van Inwagen and Dean Zimmerman, 99–141. Oxford: Oxford University Press.
- Ramachandran, V. S. 2004. *A Brief Tour of Human Consciousness: From Imposter Poodles to Purple Numbers*. New York: Pi Press.
- Ramachandran, V. S. 2012. *The Tell-Tale Brain: A Neuroscientist's Quest for What Makes Us Human*. New York: W. W. Norton.
- Ramachandran, V. S., and Sandra Blakeslee. 1998. *Phantoms in the Brain: Probing the Mysteries of the Human Mind*. New York: Morrow.
- Tye, Michael. 1995. *Ten Problems of Consciousness*. Cambridge, MA: Bradford Books.
- Tye, Michael. 2007. "Philosophical Problems of Consciousness." In *The Blackwell Companion to Consciousness*, edited by M. Velmans and S. Schneider, 23–35. Oxford: Blackwell.
- Weiskrantz, Lawrence. 2009. *Blindsight: A Case Study Spanning 35 Years and New Developments*. Oxford: Oxford University Press.

# Debating Animalism

# For Animalism

ERIC T. OLSON

## 18.1 What Animalism Does and Doesn't Say

What are we? The question has many answers. Some are evident and undisputed: we are people; we are subjects of consciousness; we are human beings (and not Martian foundlings). Others are subject to debate: we are creatures made in God's image; we are products of evolution by natural selection; we are material things, made up entirely of chemical atoms. One such disputed answer is that we are biological organisms. We are material things of a specific sort: animals of the primate species *Homo sapiens*. This is the view known as animalism. Before discussing why it is disputed and whether it's true, I want to distinguish it from some similar-sounding claims.

Animalism says that we human people (or, as the lawyers say, persons) are organisms. It does not say that this is true by definition. Specifically, it does not say that a person is by definition a sort of organism. Most definitions of personhood say that to be a person is to have certain special mental capacities – to be rational, perhaps, capable of consciousness, and able to think about oneself in the first person (as in, “I wish I weren't such a coward”). A god or an angel, were there such a thing, would be a person in this sense, though not an organism. This is perfectly compatible with the claim that we *human* people are organisms. Animalism does not propose any definition of personhood – any account of what it is to be a person, as opposed to a nonperson. Nor do animalists claim to know a priori, just by rational reflection, that we are organisms. It's the senses, and not reason alone, that tell us that we are animals.

Animalism does not say that we are animals *essentially*. Something is an animal essentially if it is an animal and could not possibly exist without being an animal. By contrast, something is an animal accidentally or contingently if it is an animal but could exist without being one (in the way that students can exist without being students). If human animals are animals essentially, then our being animals implies that *we* are animals essentially. If they are animals accidentally, then animalism implies that we are animals only accidentally. Which of these is the case is independent of whether we are animals.

More generally, animalism does not imply or presuppose any account of the metaphysical nature of animals. Aristotelians say that an animal (or any other material object) is a compound of matter and form. And what makes something an animal, or more specifically a human animal or a spider or a worm, is its form rather than its matter. Scholars disagree about what this means, but whether it is true is, again, independent of whether you and I are animals or nonanimals. Both Aristotle and Aquinas thought we were animals, and were thus animalists *avant la lettre*.<sup>1</sup>

Finally, animalism does not say that our entire nature consists in or follows from our being animals. It does not imply that our behavior is biologically determined, or that we have only biological properties (whatever these claims might mean) – any more than our being physical objects implies that our only properties are those studied in physics. There is nothing “reductionistic,” in any meaningful sense of the term, about animalism. Animalists accept the indisputable fact that we have important features not shared by any nonhuman organisms that we know of: the capacity for sophisticated rational thought, for instance, and to speak a language with a complex grammatical structure. That we are animals is only the beginning of an account of human nature.<sup>2</sup>

## 18.2 What We Appear to Be

Why suppose that we are animals? Well, that’s how it appears. We seem to be material things made of the same stuff that makes up sticks and stones. (This thought is expressed in the book of Genesis: “Dust you are, and to dust you will return.”) Which material things do we seem to be? If you examine yourself in the mirror, you see an animal. The animal appears to be the same size as you – no bigger and no smaller. Like animals, we seem to extend just as far as the surface of our skin. Each of us seems to have the physical and biological properties of an animal: its mass, temperature, chemistry, anatomy, and so on. Nor is there any difference in behavior between a human animal and a human person. The appearance is that we *are* the animals in the mirror. Of course, appearances can be deceptive. We might really be something else. But animals are what we appear to be.

Consider the alternatives. One is that we are material things other than animals: smaller parts of animals, for instance, such as brains. This would mean that we each weigh less than three pounds. A human person is composed of soft, yellowish-pink tissue and located entirely within her skull. Most likely you have never literally seen yourself or anyone else, and may not want to.

Another alternative is that each of us might be physically identical to the animal in the mirror and made of precisely the same matter, yet not an animal but rather a nonanimal “constituted by” it. How could something be physically just like an animal without being an animal itself? Well, we might have different modal properties from animals. Perhaps we are *essentially* able to think: it is absolutely impossible for us to exist without having that ability. Yet no animal is essentially able to think: each human animal begins its existence as an embryo without any mental capacities, and could end its life in an irreversible coma. If you are essentially able to think but no animal is, then you could not be an animal. (In fact you must have come into being several months after the animal did: you could not have existed before the animal was able to think.)

Or we might not be material things at all. We might be nonphysical, invisible, intangible, immaterial things. Or each of us might be composed of two things, one immaterial and one material.<sup>3</sup>

There are other alternatives to our being animals. Many of them are respectable views with important advocates. But these alternatives are surprising. They are not how things appear. We don't *seem* to be smaller parts of animals, or nonanimals made of the same matter as animals, or wholly immaterial things. We seem to be animals. We ought to deny that we are animals only if we have arguments strong enough to overturn this appearance. But we don't need any argument in order to believe that we are animals, because that's how things appear before the arguments are given.

That we are animals is like the view that time is real. That too is how it appears: things seem to happen one after another. Some philosophers believe this appearance to be an illusion, but only on the basis of arguments they take to be powerful enough to overturn the appearance. We are not entitled to deny the reality of time without any argument. On the contrary: it's reasonable to believe that time is real without having any argument for it, because that's how things appear before the arguments are given. Of course, the reality of time is more widely held than animalism. That's because the objections to the reality of time are weaker than the objections to our being animals (I'll come to these presently). My point is simply that both look true on the face of it.

Someone might say that what appears to be the case before we consider the arguments is not that we *are* animals, but only that we have animal bodies. *Our bodies* appear to be animals. We don't.

In order to assess this claim, we need to know what it is for a thing to be someone's body. What does it *mean* to say that *x* is *y*'s body? It's no easy question. The most common answer is that someone's body is an object that she can move and feel in an especially direct way: you can move and feel your body without moving or feeling anything else (except parts of it). Of course, you can move and feel your left hand without moving or feeling anything else, yet your left hand is not your body. Perhaps your body is the *largest* object that has this feature (see Olson 1997, 142–153). If this is correct, then for our bodies to be animals is for us to be able to move and feel the animal we see in the mirror, and no larger object, in that direct way. Since we do appear to have that ability, our bodies seem to be animals.

But this does nothing to diminish the appearance that we ourselves are animals. You can move and feel your hands without moving or feeling anything else. And your hands appear to be parts of you, as well as parts of your body. They are not merely instruments that you use to tie your shoes and turn the pages of books. To move or feel your hands is to move or feel yourself. That's how it seems, anyway. You appear to be a material thing of which your hands are parts, not an immaterial thing in two-way communication with your hands. In fact, you appear to be the largest object that you can move and feel directly: you appear to *be* your body.

It may be that the phrase "*x*'s body" means nothing like "the largest object *x* can move and feel directly." In that case we cannot evaluate the claim that only our bodies appear to be animals until we have an alternative account of what that phrase means. But whatever exactly it comes to, the claim that our bodies appear to be animals is unlikely to imply that we ourselves do not.



### 18.3 The View from Within

Someone might say that although we appear *to the senses* to be animals, we appear *in introspection* to be something very different: immaterial entities not composed of parts.

Close your eyes, plug up your ears, and ignore all bodily sensations. What do you appear, from this perspective, to be? You don't appear to be an animal. Without sensory information, you can't even tell whether there *are* any animals. Do you appear *not* to be an animal, then?

Descartes seems to have thought so.<sup>4</sup> He noted in the Sixth Meditation that we cannot distinguish any parts of ourselves by introspection. You can tell by introspection that you are thinking, and more specifically that you have certain beliefs, desires, emotions, and the like. But you can't tell whether you are composed of parts. (Your beliefs and desires are not parts of you, any more than your movements are.) Descartes inferred from this that we are *not* composed of parts. Yet if we were animals we should have many parts: organs, cells, atoms, and so on. It would follow that we are not animals. Or if the argument does not establish conclusively that we have no parts, it may show that we at least appear to have no parts when we look within, contrary to the way we appear outwardly to the senses. That would make appearances equivocal.

But is the inability to detect parts of ourselves by introspection any reason to suppose that there are no such parts to be detected? Does it show that we appear, to introspection at least, to have no parts? The answer depends on how it would appear to introspection if we *did* have parts. Suppose we were animals, composed of organs, cells, and atoms in vast numbers. Would that give us a different inward appearance from what we actually observe? Descartes gives no reason to think so. For all he says, our appearance to introspection would be exactly the same whether we were animals composed of parts or simple, immaterial entities. If so, introspection provides no evidence against our being animals.

Compare the fact that we cannot detect any parts of ourselves by introspection with the fact that we cannot detect any penguins by introspection. This does not mean that there appear in introspection to be no penguins, contrary to the appearance given by the senses. Although introspection does not give the appearance that there are penguins, it does not give the appearance that there are no penguins either. It is simply silent on the existence of penguins. It says nothing either for or against. Introspection is equally silent on whether we are made up of parts. It does not give the appearance that we have parts, but neither does it give the appearance that we have no parts. It is not evidence of absence, but mere absence of evidence. We appear unequivocally, before the arguments are given, to be animals.<sup>5</sup>

### 18.4 People and Their Bodies

I have said that we appear to be animals and we ought to believe otherwise only on the basis of evidence strong enough to outweigh this appearance. And many philosophers *have* believed otherwise. Nearly all the major figures in the history of Western philosophy (with the notable exception of Aristotle and his followers) denied that we are animals. The most common reason for this was the conviction that no material thing could think (a topic discussed elsewhere in this volume). Since *we* clearly think, we must be immaterial, and therefore not animals.

Few contemporary philosophers take us to be immaterial. Yet many deny that we are animals.<sup>6</sup> Why do they reject the appearance?

One reason is based on the thought that to be an animal is to be a mere body. And it sounds wrong to say that people are the same thing as their bodies. People are one thing, the thought goes; their bodies are something else. Since animals are bodies, it follows that people are not animals.

Whether our being animals really does imply that we are our bodies depends on what it is for something to be someone's body. The answer to this question that we considered in Section 18.2 suggests that it does have this implication. I don't set much store by that answer, but I will concede the point for the sake of argument. The important question, then, is why we should deny that people are their bodies. Presumably it's because it sounds wrong to say things like this:

- 1 Descartes's body gave two proofs for the existence of God.
- 2 Descartes's body read *The Guardian*.
- 3 Descartes's body asked the shop manager for a refund.

The reason why these statements sound wrong, the objection claims, is that our bodies don't give philosophical arguments or read newspapers or ask shop managers for refunds. Those are simply not things that our bodies can do. They lack the mental and behavioral properties that we have.

But it's doubtful whether this really is why 1–3 sound wrong. Consider these:

- 4 Descartes's body was born in 1596.
- 5 Descartes's body had dark hair.
- 6 Descartes's body was seen entering the shop at 6:24 p.m.
- 7 Descartes's body died of pneumonia.

They sound wrong too. And this would appear to be for the same reason that 1–3 sound wrong. (They have the same sort of "wrong" feel.) Yet if Descartes's body was an animal, as the objection claims, they cannot be wrong because his body was not born in 1596, did not have any particular hair color, was not seen entering a shop, and did not die of pneumonia. These are things that clearly *can* be true of animals. Why such statements sound wrong – why we can say that Descartes, but not his body, died of pneumonia, for instance – is not easy to say. But the answer cannot be simply that human bodies are incapable of doing the things described in the statements. (Not, anyway, if they are animals.) That undermines the claim that 1–3 sound wrong because human bodies cannot think or act as we do. The way in which we can and cannot use such phrases as 'Descartes's body' provides no reason to deny that human animals have the mental and behavioral properties that we have. It is therefore no objection to our being animals.

## 18.5 Life after Death

Another thought has to do with life after death. To have life after death is to exist, in a conscious state, after one has died and one's remains have decayed to the point

where they can no longer function. This does not seem possible for a biological organism. Once an organism has been consumed by worms, it no longer exists, and so cannot be conscious. Animals, even human animals, cannot have life after death. If *we* could have life after death, we cannot be animals. Or at least not animals essentially. If we were animals accidentally, we might be able to have life after death by first becoming nonanimals.

The argument does not require that we actually have life after death, but only that we *could* have it: that it is possible in the broadest sense. If there were a god, he could bring it about. This is not because a god would be able to do just anything. Not even an omnipotent being could make a liquid giraffe, or bring it about that he himself had never existed. Nor could he give an organism life after death. But he could give *us* life after death.

Is life after death really possible for us? It's not obviously *impossible*. We seem able to imagine it, whereas we can't even begin to imagine a liquid giraffe. But that we are animals is not obviously impossible either. If we can imagine having life after death, we can certainly imagine being animals. In fact we don't have to imagine it: that's exactly how it appears. So if it's possible for us to have life after death, it ought to be equally possible for us to be animals.

But these things cannot both be possible. If we are animals, then we are necessarily animals. (So the objection to animalism presupposes, anyway.) Whether we are animals is not a contingent matter: it could not be that we are in fact animals but we could have been immaterial entities instead, or vice versa. We are animals either in all possible worlds in which we exist, or in none. It follows that if it's possible for us to be animals – logically or metaphysically possible – then we are necessarily animals. And if we are necessarily animals, then it is not possible for us to have life after death – assuming, anyway, that no animal could have life after death. So it cannot be possible for us to have life after death and also possible for us to be animals. At least one of the two must be impossible. But it's hard to know which it is.

The reason it's hard to know is that things can be impossible without being obviously so. Consider these two statements:

- 8 Every even number is the sum of two prime numbers.
- 9 Not every even number is the sum of two prime numbers.

Since they are contradictory, only one of them can be true. Indeed, only one is possible. The one that's false, like all mathematical falsehoods, is necessarily false – that is, impossible. But no one knows which it is. No one has ever found a contradiction in the thought that every even number is the sum of two primes. Nor has anyone found a contradiction in the idea of an even number that is *not* the sum of two primes. It is, as the mathematicians say, an open problem. Both statements may seem possible, yet one of them is impossible for an unknown reason. And the statement that we have life after death could be the same: it might be impossible for an unknown reason.

But even if it really were possible for us to have life after death, this would rule out our being animals only if it's impossible for an animal to have life after death. And this is disputed. Eminent philosophers – Christian philosophers, no less – have argued that we could have life after death even if we are animals, and animals essentially (see Chapter 32, this volume). If they are right, the objection to animalism collapses.

## 18.6 Brain Transplants

The most common reason for rejecting animalism is that it has unattractive consequences about what it takes for us to persist through time. These consequences do not occur in real life, but we can imagine cases where they do.

Suppose your brain were transplanted into my head, and my own brain destroyed. Because the brain is the organ primarily responsible for your psychological features, it seems that the resulting person – the one with your brain – would have your plans, preferences, personality, and cognitive skills, for the most part at least (even if there would also be messy side effects). He would remember your life and not mine. My own memories, plans, preferences, personality, and cognitive skills would be destroyed along with my brain. The resulting person would think he was you and not me. It's tempting to think that he really would be you. The operation would pare you down to the size of a brain, move you across the room, and then give you a new set of parts to replace the ones you lost. It would not give *me* a new brain, with new memories, plans, and so on. It would give *you* a new body.

But the operation would not give any animal a new body. It would simply move an organ from one animal to another, just as a liver transplant would. The animal previously associated with you would lose an organ, and with it the capacity for thought and consciousness. That organ would then become a part of the animal previously associated with me. If you and I are animals, the operation *would* give me a new brain, together with memories of things I never did and false autobiographical beliefs. And it's easy to believe that this is wrong. If it is, then we must be something other than animals.

In other words: a brain transplant would move a person from one animal to another. But it would not move an animal from one animal to another. It follows that a person is not an animal. Even if you never have a brain transplant, you have a property that no animal has, namely being such that you would go with your brain if it were ever transplanted. And if you have a property that no animal has, you are not an animal.

How strong is this argument? The crucial premise is that the person would go with her transplanted brain: the one who ended up with that organ would be the donor and not the recipient. Attractive though this may sound, it looks less compelling if we tell the story in a different way (Snowdon 2014, 234). Suppose you had an illness that would kill you unless your brain were replaced with a healthy donated organ. This, of course, would have grave side effects: it would destroy your memories, plans, preferences, and other mental properties. It may not be clear whether you could survive such a thing, even if the operation were completely successful. But it's not obvious that you *couldn't* survive it either. (Whether your survival in these circumstances would be any benefit to you is another matter.) Maybe the operation could save your life, though at great cost. We can't confidently rule this out. Nor could we confidently rule it out if the new brain gave you memories, plans, and preferences from the donor. But if it's not obvious that the brain recipient would not be you, then it's not obvious that it would be the donor. Maybe the donor would simply lose an organ, and the organ would become a part of you, saving your life. The claim is not that this is obviously true, but that it's not obviously false. And in that case it's not obvious that a person must go with her transplanted brain, contrary to the transplant objection. Though forceful, the objection does not settle the matter.

## 18.7 An Argument for Animalism

I have argued that we appear to be animals and we don't appear to be nonanimals (Section 18.2). I have tried to counter objections both to this (Section 18.3) and to the claim that we actually are animals (Section 18.4, Section 18.5, and Section 18.6). I will conclude with an argument for animalism.

The main premise of the argument is that it's possible for a biological organism to have mental properties. Dogs, for instance, can be conscious. They can feel pain, and be happy or miserable. They like some things and dislike others. That's why we have animal-welfare laws. The mental powers of dogs may be feeble compared to ours, but they *have* mental powers.

This makes it hard to deny that we are animals. If any organism can have mental properties, a human animal can. If dogs can be happy or miserable, human animals can too. And in that case, normal, adult human animals actually have mental properties. They are conscious; they are happy or miserable; they like some things and dislike others. But they don't have different moods or likes or dislikes from ours. If the animal sitting here is happy, I'm happy. If it prefers red wine to ditchwater, I do too. And presumably the converse also holds: if I feel or like or dislike something, the animal does as well. There appears to be no psychological difference between the animal and me. That's hardly surprising, seeing as we share the same history, the same surroundings, the same sense organs, and of course the same brain and nervous system.

Suppose this is right: each of us is psychologically indistinguishable from a certain human animal. It would follow that a normal, mature human animal is a person – that is, a rational, conscious being that can think about itself in the first person. And if you and I were not those animals, it would follow that you were something other than the animal person thinking your thoughts. Being a person yourself, you would be one of *two* people within your skin thinking, in exactly the same way, about philosophy. There would be twice as many people as we thought there were.

This would threaten to make the view that we are not animals self-undermining: even if it were true, we could never know that it was. How could you know that you were the nonanimal person thinking your thoughts and not the animal person? If you think you are the nonanimal, the animal will think, using the same reasoning, that it is too. For all you could ever know, *you* might be the one making this mistake.

Your epistemic situation would be like that of someone who had just been duplicated (Olson 2015a, sect. 6). Suppose we had a duplicating machine. When you step into the “in” box, the machine reads off your complete physical (and mental) condition and uses this information to assemble a perfect duplicate of you in the “out” box. The process renders you briefly unconscious but is otherwise harmless. One person wakes up in each box. The boxes are indistinguishable. Because each person will have the same memories and perceive identical surroundings, each will think, for the same reasons, that he or she is you. But only one will be right. If this happened to you, you would have no reason to suppose, afterwards, that you were the original person who stepped into the machine rather than the freshly made duplicate. (Suppose the technicians who work the machine are sworn to secrecy and immune to bribes.) You would think, “Who am I? Did I do the things I seem to remember doing, or did I come into being only a moment ago, complete with false memories of someone else's life?” And you would have no way of answering these questions.

So even if you were not the animal, you could never know it. Nor could you know whether you would go with your brain if it were transplanted or stay behind with an empty head, undermining the transplant objection. I take this to be an absurd outcome. The obvious solution to the problem is to suppose that you are not a second person in addition to the animal. You are the animal person.

This conclusion follows from four premises (Olson 2003):

- 1 There is an animal where you are.
- 2 The animal thinks.
- 3 You think.
- 4 You are not one of two thinkers where you are.

If you are not an animal, at least one of these claims must be false. The one opponents of animalism are most likely to reject is Premise 2. If the human animal where you are is *not* thinking, then it's not a person, and not a second thinker of your thoughts. It does not mistakenly take itself to be you, and you can know that you are the person and not the animal.

But this has startling consequences. If the animal is not thinking, then no human animal can ever think. And that can only be because it is metaphysically impossible for any biological organism to have any mental property. It follows that dogs cannot feel pain or prefer some things to others. They are no more sentient or intelligent than stones. At most a dog might relate to a conscious canine nonorganism in the way that a human organism relates to a human person, whatever way that might be. What appears to be a conscious animal is really two things: a conscious nonanimal and an unconscious animal.

And if biological organisms cannot have mental properties, there has to be a reason why not. Why should it be absolutely impossible for an animal to be conscious? What prevents human animals from using their brains – *our* brains – to think? This is a hard question (see Olson 2007, 31–35; Shoemaker 2011). Opponents of animalism desperately need an answer to it.<sup>7</sup>

## Notes

1. Toner (2014) is an interesting discussion of Thomistic animalism. My own thoughts on Thomism, for what they're worth, are in Olson (2007, 171–176).
2. For more on the topic of this section, see Olson (2015b).
3. Parfit (2012) argues that we are brains. On the constitution view, see Baker (2000) and Shoemaker (2011). Many famous historical figures have taken us to be immaterial: Plato, Augustine, Descartes, Leibniz, Berkeley, Hume, and Kant, for instance. For more on these and other alternatives to animalism, see Olson (2007).
4. AT VII: 86. The argument presented here is an oversimplification. Bennett's interpretation (2001, 67–71) is more plausible, though equally ineffective as an argument for the claim that we are or appear to be immaterial.
5. This is not to deny that there are metaphysical arguments for our being simple (Lowe 2001; Olson 2007, 153–164, 176–179), or to imply that such arguments are worthless. The point is simply about how things appear before we consider such arguments.
6. A recent survey of more than 900 professional philosophers (Bourget and Chalmers 2009) found that only 17 percent favored a "biological view" of personal identity, which is more or less equivalent to animalism. The survey did not include a question about whether we are immaterial, presumably because the authors thought there was little current debate over it.
7. I thank Stewart Goetz, Jonathan Loose, Angus Menuge, and J. P. Moreland for comments on earlier versions.

## References

- Baker, L. R. 2000. *Persons and Bodies: A Constitution View*. Cambridge: Cambridge University Press.
- Bennett, J. 2001. *Learning from Six Philosophers*, vol. 1. Oxford: Oxford University Press.
- Bourget, D., and D. Chalmers. 2009. "The PhilPapers Surveys." Accessed March 21, 2016. <http://philpapers.org/surveys/>.
- Lowe, E. J. 2001. "Identity, Composition, and the Simplicity of the Self." In *Soul, Body, and Survival: Essays on the Metaphysics of Human Persons*, edited by Kevin Corcoran, 139–158. Ithaca, NY: Cornell University Press.
- Olson, E. 1997. *The Human Animal: Personal Identity Without Psychology*. Oxford: Oxford University Press.
- Olson, E. 2003. "An Argument for Animalism." In *Personal Identity*, edited by R. Martin and J. Barresi, 318–334. Malden, MA: Blackwell.
- Olson, E. 2007. *What Are We? A Study in Personal Ontology*. Oxford: Oxford University Press.
- Olson, E. 2015a. "Personal Identity." In *The Stanford Encyclopedia of Philosophy*, edited by E. Zalta. Accessed June 20, 2017. <http://plato.stanford.edu/archives/fall2015/entries/identity-personal/>.
- Olson, E. 2015b. "What It Means to Say that We Are Animals." *Journal of Consciousness Studies*, 22(11–12): 84–107.
- Parfit, D. 2012. "We Are Not Human Beings." *Philosophy*, 87: 5–28.
- Shoemaker, S. 2011. "On What We Are." In *The Oxford Handbook of the Self*, edited by S. Gallagher, 352–371. Oxford: Oxford University Press.
- Snowdon, P. 2014. *Persons, Animals, Ourselves*. Oxford: Oxford University Press.
- Toner, P. 2014. "Hylemorphism, Remnant Persons and Personhood." *Canadian Journal of Philosophy*, 44: 76–96.

# Against Animalism

STEWART GOETZ

Are we souls or animals? In this chapter, I explain why I think a belief in the soul is reasonable and a belief in animalism is not. In Section I, I point out how ordinary people for ages have believed in the soul and explain why I think this is the case. After that, I spend some time discussing the nature of the soul with a focus on those properties that will be important for my consideration of animalism, which is found in Section II. Therein, I first question Eric Olson's claim that we seem to be animals and then set forth his characterization of animalism for the purpose of raising concerns about his explanation of an organism's identity through time in terms of a life. Finally, in Section III I briefly consider the nature of the dialectic between persons like Olson and me.

## I

According to Eric Olson,

it would be an understatement to say that immaterialism [the view that we are immaterial substances or souls] is out of favor nowadays. Most philosophers of mind treat it as little more than a historical curiosity . . . In the current intellectual climate, the interesting question about immaterialism is not whether it might be true, but how the likes of Plato, Descartes, and Leibniz could ever have believed it. (Olson 2007, 151)

I believe the answer to this question, which at least some contemporary philosophers either forget or find difficult to accept, is fairly obvious: these philosophical luminaries could believe that we are souls because the view is so intuitively plausible. While Olson is right to point out that the *contemporary academic establishment* (he mentions philosophers) is firmly opposed to the existence of the soul, it is nevertheless the case that belief in the soul is daily bread and butter for *ordinary people*. As far as the ordinary person is concerned, not only are we souls but we also have bodies: human beings are souls with material bodies, which is the view commonly referred to as "soul-body dualism," "substance dualism," or for my purposes, simply "dualism." The experimental cognitive scientist Jesse Bering, while he



denies the truth of dualism, acknowledges that ordinary “folk” as such certainly affirm it (Bering 2006). And according to the developmental psychologist Paul Bloom, who also denies dualism, all of us naturally develop a belief in it (Bloom 2004). And he cites other academics (e.g., anthropologist Alfred Gell and neuropsychologist Paul Broks) who make the same point (while also denying dualism) (Bloom 2004, 195). Our ordinary belief in dualism is corroborated by reports of near-death experiences. The actual truth or falsity of these reports is of no concern here. What is relevant is the fact that both those making the reports and those hearing of them almost never question the metaphysical possibility of what the reports describe. What they wonder about is the integrity of the evidence for and against what is reported. And literature also makes effective use of our ordinary belief in dualism. J. K. Rowling, in her hugely successful *Harry Potter* stories, depicts the worst death one can die as having one’s soul sucked out of one’s body by the kiss of a dementor.

What *explains* the natural human proclivity to affirm the existence of the soul? If one believes people like Daniel Dennett, there is ultimately only one legitimate explanation (a universal explanatory acid) for the continuing existence of anything: random (caused but purposeless) changes resulting in arrangements of matter that proved to be adaptive (Dennett 1995). Given this explanatory paradigm, belief in the existence of the soul arose in something like the following way: People initially postulated the existence of agents in the form of invisible souls to explain the rustlings in the brush and strange noises downstairs. This ascription proved to be advantageous for survival (it is better to believe that there is someone downstairs and take the appropriate action than to roll over and go back to sleep; even if the former alternative is often wrong – a false positive, it is better to be wrong and live than to do nothing and be killed) and as a result was passed on to descendants (see Schloss and Murray 2009 for an interesting treatment of these issues).

This explanation of soul belief (it is adaptive, but erroneous) is an instance of what philosophers regard as a third-person approach to an issue: it brackets experiences and beliefs about the subject itself and considers its attitude toward other objects. A different kind of approach is first-person in nature and takes seriously a subject’s beliefs about itself and what explains them. What can we learn from the first-person perspective about why people are natural-born believers in souls? Well, we learn that people believe that they themselves are souls and that they do not believe this because they are out to explain something and postulate the soul’s existence for explanatory purposes. They believe they are souls because they are directly aware of themselves as such, or at least fail to be aware of themselves as material entities. There are two alternatives here.

First, one might think that one is directly aware of oneself as a substantively simple entity in the sense that one is aware of oneself as a thing that is not made up of other things. One is aware of oneself as not having separable parts. For example, René Descartes wrote the following:

Mind and body . . . are in fact substances which are really distinct one from the other . . . This conclusion is confirmed . . . by the fact that we cannot understand a body except as being divisible, while by contrast we cannot understand a mind except as being indivisible. For we cannot conceive of half of a mind, while we can always conceive of half of a body, however small; and this leads us to recognize that the natures of mind and body are not only different, but in some way opposite. (Descartes 1984, 9–10)

When I consider the mind [soul] or myself in so far as I am merely a thinking thing, I am unable to distinguish any parts within myself; I understand myself to be something quite single and

complete. Although the whole mind seems to be united to the whole body, I recognize that if a foot, or an arm, or any other part of the body is cut off, nothing has thereby been taken away from the mind. As for the faculties [powers and capacities] of willing, of understanding, of sensory perception, and so on, these cannot be termed parts of the mind, since it is one and the same mind that wills, and understands and has sensory perceptions. (Descartes 1984, 59)

Second, one might fail to be aware that one has any separable parts. Roderick Chisholm nicely explains this idea:

I may perceive myself to be thinking and know that I am doing so and yet be unable to know whether I am perceiving any proper part of anything that I am perceiving. It may be, for all anyone knows, that whenever I perceive myself to be thinking, I *do* perceive some part of myself. This would be the case, for example, if I could not perceive myself to be thinking without perceiving some part of my body, and if, moreover, I were identical with my body or with that part of my body. But it is not true that, whenever I perceive myself to be thinking, I thereby perceive what I can *know* to be a part of myself. (Chisholm 1976, 100)

The position represented by Descartes is epistemologically stronger than that suggested by Chisholm in the sense that it entails, as Chisholm's does not, that the self (soul) has no separable parts. On Chisholm's view, one could fail to be aware that one has such parts and on this basis alone unjustifiably infer that one is substantively simple because one mistakenly assumes that one would be aware of those parts were one to have them. But regardless of whether this unjustified inference is being made, it is easy to understand how one would believe that one is a soul on either the Cartesian or Chisholmian perspective.

In focusing on the soul's substantive simplicity, I am not pointing out anything new. Olson correctly writes that "most philosophers who believe in souls take them to be mereologically simple – that is to lack proper parts" (Olson 2007, 150). But what is sometimes not mentioned when discussing the idea that the soul is substantively simple is the important point that a soul's lacking separable parts is compatible with its having a multiplicity of properties, because properties are not separable parts. For example, both Descartes and Chisholm would affirm that the soul has properties like the power to choose, the capacity to experience pain, the capacity to experience pleasure, the capacity to believe, and so on. Each of these properties is irreducibly distinct from the others so there is a genuine plurality of properties. Yet complexity at the level of properties is consistent with simplicity at the level of substance.

If one is a substantively simple soul with a multiplicity of properties, where is one located? Descartes seems to have concluded that the soul is not located in space at all. In brief, his reasoning at this point apparently went as follows: A body is that which is extended, has shape, is divisible into separable parts, and can be moved, though it cannot move itself. A soul is substantially simple and, therefore, not extended, has no shape, is not divisible into separable parts, and is immovable, whether by something else or itself (Goetz and Taliaferro 2011).

While contemporary philosophers often identify dualism with Descartes's dualism (Cartesian dualism) and, thereby, assume that the soul, were it to exist, would not be in space, dualism is not identical with its Cartesian form. Descartes's dualism is a species of a genus. Thus, there are non-Cartesian dualisms. Most importantly for present purposes, there is a kind of dualism that locates the soul in space. Indeed, this was the orthodox species of dualism up until the time of Descartes (the beginning of the seventeenth

century). The standard view until then had the soul located in its entirety at every point in space that is occupied by its physical body. The idea is that souls and bodies both occupy space, but they do so in different ways. Bodies occupy space in virtue of their substantive parts occupying subregions of the space occupied by the complex whole. Because souls have no separable parts, they must be present in their entirety at every point in the space that they occupy. The upshot is that souls, while having no shape themselves, occupy shaped regions of space. And because souls and bodies occupy space in different ways, they can occupy the same space.<sup>1</sup>

Undoubtedly, this is all somewhat mysterious. So why affirm the mystery? Why affirm the soul occupies the space of its physical body? The answer is not mysterious in the least: it is plausible to maintain that the soul occupies the space occupied by its physical body because that is the space it seems to occupy. For example, when your foot is stepped on, you feel pain in *your foot*. And when your head is hit, you feel pain in *your head*. *You, in your entirety*, feel the pains in these places *simultaneously*, so that you must be present in your entirety and at the same time in your foot and in your head. Furthermore, even without feeling these pains when and where you feel them, you feel like you occupy the space of your body. You feel right now as if you are simultaneously in your arms, legs, head, and so on. The term for the capacity for you to feel in this way is “proprioception.” Descartes was well aware of how we feel. In his *Meditations*, he writes: “But as to the nature of this soul, either I did not think about this or else I imagined it to be something tenuous, like a wind or fire or ether, which permeated my more solid parts” (Descartes 1984, 17). Ultimately, Descartes believed he had theoretical considerations that justified his concluding that his feeling of occupying the space filled by his body is illusory. But he fully recognized that initially it seemed he occupied the entirety of his body.

Souls seem to exist in space and they feel pain and pleasure, believe, think, choose, and so on. Souls also seem to persist or have identity through or over time. I, my soul, exist now and I, the numerically same entity, also existed yesterday, last week, last year, ten years ago, and for many years prior to that.

## II

According to dualism, I am essentially a soul. Am I also an animal? Yes, because I am a soul that has a body. But I am an animal only accidentally, not essentially. In other words, while I am presently an animal in the sense that a particular animal body is mine – I have it, that animal body is not me in whole or in part. Because it is not, I might continue to exist without it, either by having a different body or no body at all.

Eric Olson is not a dualist. He is what he terms an “animalist,” where an animalist is someone who affirms that we are animals. Animals are material, biological organisms, where an organism differs from other material things by having a life (Olson 2007, 28). A life is

a self-organizing biological event that maintains the organism’s complex internal structure . . . An organism must constantly take in new particles, reconfigure and assimilate them into its living fabric, and expel those that are no longer useful to it. An organism’s life enables it to persist . . . despite constant material turnover. (Olson 2007, 28)

Olson maintains that “an organism persists if and only if its life continues” (Olson 2007, 29).

Why believe animalism? According to Olson, “we seem to be animals” (Olson 2007, 23). Here, we need to be careful. A dualist affirms that we seem to be animals *accidentally*. That is, a dualist agrees that we seem to have animal bodies. However, a dualist denies that we seem to be animals *essentially*. As I pointed out at the beginning of Section I, a dualist affirms what ordinary folk believe, which is that we are essentially souls. Therefore, when Olson writes that we seem to be animals and means by this that we seem to be animals *essentially*, a dualist must respectfully disagree. The referent of “we” in Olson’s assertion is certainly not ordinary people.<sup>2</sup>

This disagreement between Olson and a dualist is significant for discussions about what we are. For example, Olson considers what he terms the “divisibility argument” for the existence of the soul (Olson 2007, 152–153). It goes as follows:

- 1 Any material thing that is a candidate for being a thinker is divisible into parts.
- 2 No thinking thing is divisible into parts.

Therefore,

- 3 Thinkers, ourselves included, must be simple (without substantive parts).

Olson asks how we could know Premise 2 is true. How could we know that no thinking thing is divisible into parts? The only way to know this is by already knowing that no thinking thing could be material. But how could we know that? Well, we could know that we think and seem to be simple. Maybe it is a bit presumptuous to move from the fact that we seem to be simple to the conclusion that only simple things could be thinkers.<sup>3</sup> But I suspect that what Olson really objects to is not this move but the claim that we seem to be simple. How things seem to us bears a good bit of argumentative weight.

Acknowledging the importance of the “seemingness of being souls” goes a long way toward explaining a point made by Olson in response to what he calls “the argument from disembodied survival” (Olson 2007, 152). Olson rightly points out that some argue for the existence of the soul from our ability to survive in a disembodied state. No material thing can survive disembodied because in that case it would cease to be a material thing (which is not possible because a material thing is essentially material). Therefore, anything that could survive disembodied must be immaterial before its disembodiment. Olson rightly answers that the basis for our knowing that we could survive in a disembodied state must be our knowledge that we are not material things. But the argument from disembodied survival was designed to conclude that we have this knowledge. It was not supposed to require that we have it from the outset. If we were to say that for all we know we are material (think here of Chisholm’s argument in the previous section: while we are not aware of ourselves being material, for all we know we might be), then even if we are material we can understand how one might think we have grounds for thinking that we are not. However, Olson would rightly respond at this point with the following: “How can we be sure that we’re not mistaken in this way? Only by ruling out the possibility of our being material ourselves. But that is what we were trying to establish in the first place” (Olson 2007, 153).

But at this juncture dualists could push back in either of two ways. First, they might affirm the stronger position suggested by Descartes’s comments and maintain they are directly aware of being souls. Second, they might affirm the weaker position represented by Chisholm’s remarks and ask Olson for some reason why we should think that we are material things. As we have already seen, Olson’s claim that we *seem* to be animals is suspect. At best, we seem accidentally to be animals. We certainly do not seem to be animals

essentially. So a dualist might request from Olson a reason why we should believe animalism is true. It seems that Olson must argue that the ordinary person's belief in the soul is mistaken, while a dualist need not completely reject the idea that we seem to be animals, but reject only the view that we seem to be animals essentially.

Is there a way to reconcile the view that we seem to be souls with Olson's claim that we seem to be animals? As I have suggested elsewhere (Goetz 2005, 55–56) there is a natural way to harmonize the two seemings. Because one seems to occupy the space that is occupied by one's material, animal body one might mistakenly slide from an awareness of this fact to asserting that one seems to be a material animal. However, all that is warranted is an assertion that one seems to occupy the space that is also occupied by one's animal body. That one is an animal body is a view that requires an argument.

As Olson recognizes, we all believe that we persist across time. It is helpful at this juncture to see how differently a dualist and an animalist account for this persistence. I begin with a dualist. According to a dualist, we are able to persist across time because the numerically same simple soul, which lacks separable parts, persists across that time. Obviously, an animalist cannot provide such an account of persistence because an animal is a complex material entity whose parts are constantly changing over time. This change is orderly or structured, but it is nevertheless real change. How, then, is it possible for an animal to persist through time? Here, Olson invokes the concept of a *life*, where a certain particular material entity *M* is a part of a particular organism *O* if and only if *M* is caught up in the life of *O*.

By a life I mean . . . a self-organizing biological event that maintains the organism's complex internal structure. The materials that organisms are made up of are intrinsically unstable and must therefore be constantly repaired and renewed . . . An organism's life enables it to persist and retain its characteristic structure despite constant material turnover. (Olson 2007, 28)

Concerning persistence through time, Olson writes that he is "inclined to believe that an organism persists if and only if its life continues" (Olson 2007, 29). Presumably, "continues" means "persists." So an animal persists if and only if the biological life in which its parts are caught up persists, and animal *A*<sub>1</sub> at time *t*<sub>1</sub> is the same particular entity as animal *A*<sub>2</sub> at time *t*<sub>2</sub> if and only if *A*<sub>1</sub>'s life at *t*<sub>1</sub> is the same as *A*<sub>2</sub>'s life at *t*<sub>2</sub>, where the sameness of life from one moment to the next is the sameness of numerical identity.

So according to animalism, the best account of the conditions under which an animal persists involves the concept of a life and its persistence. What is it for a life to persist? Olson says that a life is an *event*. Is it, then, a simple event with no event parts? Or is it a complex event with event parts? Assertions by Olson clearly indicate that he thinks of a life as a complex event. For example, he endorses Peter van Inwagen's comments that a biological life is composed of the activities of the things that compose it (Olson 2007, 226).<sup>4</sup> And Olson himself writes that, "the objects that *compose* the organism are the ones whose activities constitute its life" (Olson 1997, 138). Activities are events, so an implication of Olson's view of a life is that it is a complex event composed of other events. Here, we need to remember that the parts of an animal organism are constantly changing over time. Hence, if a life is composed of the activities of these parts and these parts are constantly changing, it follows that the activities of a life are constantly changing. So a life is a complex event whose event parts are constantly changing.

But what, then, explains the persistence of a life across time?<sup>5</sup> It seems we end up with a complex event that remains numerically the same through time, even though its parts are

constantly changing. So an animal whose parts are constantly changing is the numerically same animal over time because those constantly changing parts are caught up in the numerically same life. But when we ask about persistence of that life over time we are informed that it is subject to the same kind of impermanence as the animal organism. If the persistence through time of a life has a compositional nature like that of the animal organism whose life it is, why invoke the concept of a life at all as an account of the conditions under which an animal persists? Why not just say that an animal persists through time, period?

In light of these concerns about the conceptual adequacy of animalism's account of persistence, how does that account compare with that which makes use of a soul? It seems to me that a dualist has a more straightforward account of persistence. Because a soul is a substantively simple entity with no separable parts, an account of its persistence is not encumbered by the kinds of issues that arise with the persistence of an organism and its life. So a soul seems eminently more qualified to be the subject of our persistence through time. We persist because we are souls whose substantive simplicity guarantees avoidance of problems about the persistence of wholes made up of parts.

### III

I close with some brief thoughts about a belief in the soul's existence as opposed to a belief in animalism. Is there a sound argument that supports belief in the soul's existence against animalism or for animalism against the existence of the soul? I doubt that there is. It seems to me that one simply starts with a basic belief in one or the other and considers objections that might be raised against that belief. If one is persuaded by one or more of the objections, one will be troubled and perhaps moved to give up the view.

Does claiming that a belief in the soul's existence or animalism is basic amount to saying that the belief is ungrounded or arbitrary? Not in the least. Because I am a dualist, I will address this question in terms of my belief in the soul. In terms of the discussion in Section I, I believe in the soul's existence because I am directly aware of being a soul (see also Moreland 2011). So my belief is grounded in my awareness. Do I expect that my claim that I am aware of being a soul will persuade Olson of the falsity of animalism? No. He might respond to me that I am confusing an awareness of being a soul with a failure to be aware of being a material entity that has substantive parts. At that point, we will have a disagreement for which quite probably there is no resolution. As a believer in the soul's existence, I have read over the years much criticism of the concept of the soul, where this criticism has mostly been presented in the form of criticisms of dualism. I have learned a great deal from reading these criticisms. But I have never been persuaded that my beliefs in the soul and dualism are false.<sup>6</sup>

My belief in the soul's existence is also, at least in part, the source of my views about identity and persistence. Given these views, I cannot help but be puzzled about animalism's account of persistence through time. Here, I follow in the tradition of people like Joseph Butler, Thomas Reid (Perry 1975, 99–118) and Roderick Chisholm (1994, 89–113, 145–158). But if Olson simply believes he is an animal, then I doubt he will find my concerns about the identity of an animal and its persistence across time all that problematic. My concerns will not get much, if any, traction with him. And I will leave the matter there.

## Notes

1. This account of how the soul is located in space helps with a response to a concern about soul-body interaction raised by Olson. He writes that “if the soul interacts with the body, it interacts with different parts of the body, specifically different parts of the brain, at the same time: for instance, it receives visual information from the occipital lobe at the same time as it receives auditory information from the temporal lobe . . . If this interaction requires spatial co-location, then the soul needs to be as big as the brain, threatening the claim that it is simple” (Olson 2007, 165). Descartes’s dualist predecessors would have responded to Olson that the soul does not need to be as big as the brain with whose parts it simultaneously interacts, because it is not spatially “spread out” across its brain. Rather, it is wholly present at each point of the brain where it causally interacts with it, assuming the brain is where the soul causally interacts with its body.
2. In private correspondence, Olson acknowledges that a dualist’s claim that “we are animals accidentally may be true when properly understood, but it is apt to mislead.” It is apt to mislead because “it suggests that Olson and Goetz agree that we are animals, and disagree only over a modal claim [the difference between being an animal accidentally and being one essentially]. It also makes it sound as if we agree that we are material things, since animals are material things. But we don’t really agree about these things at all.” Olson maintains that “the most perspicacious way to express [the dualist view] is to say that I am an immaterial soul that (accidentally) has an animal body.” I am completely at ease with expressing the dualist view as Olson suggests. Indeed, I affirmed at the beginning of Section I that we are souls. However, I am just as thoroughly at ease expressing the dualist view as “I am an animal, but only accidentally.” Is the latter expression really apt to mislead? Olson has no problem understanding what I mean. No philosopher whose work I have read and/or interacted with has had a problem understanding dualism. What about ordinary people? They have never accused me of misleading them when we discuss what we are. Similarly, while standing in a filling station, I have heard customers say things like “I am out on the bypass out of petrol/gas” to employees. What I have never heard is the response “Don’t mislead me! You are not down at the corner but standing right here in front of me.” The fact is that we all, whether philosophers or not, understand that it is perfectly natural to identify ourselves with objects with which we are closely associated. According to a dualist, to say that I am an animal is just one more instance of this, as are assertions like “I weigh X” and “I am Y feet tall.”
3. However, David Barnett (2008, 2010) argues that the ordinary person’s naive conception of conscious beings demands that they be simple.
4. Olson references van Inwagen (1995).
5. For what follows, I am indebted to an insightful unpublished paper entitled “Grounding Personal Identity: An Enduring Problem for Animalism” (Rickabaugh, n.d.).
6. For an admission by a materialist that the standard objections to dualism are not convincing, see Chapter 2.

## References

- Barnett, David. 2008. “The Simplicity Intuition and Its Hidden Influence on Philosophy of Mind.” *Noûs*, 42: 308–335.
- Barnett, David. 2010. “You Are Simple.” In *The Waning of Materialism*, edited by Robert C. Koons and George Bealer, 161–174. Oxford: Oxford University Press.
- Bering, Jesse. 2006. “The Folk Psychology of Souls.” *Behavioral and Brain Sciences*, 29: 453–498.
- Bloom, Paul. 2004. *Descartes’ Baby: How the Science of Child Development Explains What Makes Us Human*. New York: Basic Books.
- Chisholm, Roderick M. 1976. *Person and Object: A Metaphysical Study*. La Salle, IL: Open Court.
- Chisholm, Roderick M. 1994. “On the Observability of the Self.” In *Self-Knowledge*, edited by Quassim Cassam, 94–108. Oxford: Oxford University Press.
- Dennett, Daniel. 1995. *Darwin’s Dangerous Idea*. New York: Simon & Schuster.
- Descartes, René. 1984. *The Philosophical Writings of Descartes*, translated by John Cottingham, Robert Stoothoff, and Douglas Murdoch, vol. 2. Cambridge: Cambridge University Press.
- Goetz, Stewart. 2005. “Substance Dualism.” In *In Search of the Soul: Four Views of the Mind-Body Problem*, edited by Joel B. Green and Stuart L. Palmer, 33–60. Downers Grove, IL: InterVarsity Press.

- Goetz, Stewart, and Charles Taliaferro. 2011. *A Brief History of the Soul*. Oxford: Wiley-Blackwell.
- Lycan, William. 2009. "Giving Dualism Its Due." *Australian Journal of Philosophy*, 87(4): 551–563.
- Moreland, J. P. 2011. "Substance Dualism and the Argument from Self-Awareness." *Philosophia Christi*, 13(1): 21–34.
- Olson, Eric. 1997. *The Human Animal: Personal Identity without Psychology*. Oxford: Oxford University Press.
- Olson, Eric. 2007. *What Are We? A Study in Personal Ontology*. Oxford: Oxford University Press.
- Perry, John, ed. 1975. *Personal Identity*. Berkeley: University of California Press.
- Rickabaugh, Brandon L. 2017. n.d. "Grounding Personal Identity: An Enduring Problem for Animalism." Unpublished paper.
- Schloss, Jeffrey, and Michael J. Murray, eds. 2009. *The Believing Primate*. Oxford: Oxford University Press.
- van Inwagen, Peter. 1995. *Material Beings*. Ithaca, NY: Cornell University Press.



# Debating Nonreductive Physicalism

# For Nonreductive Physicalism

NANCEY CLAIRE MURPHY

## 20.1 Introduction

I begin with notes on terminology. I call my position “nonreductive physicalism” because it is the most common term in current philosophy of mind. It is meant to signal opposition to anthropological dualisms of body and either mind or soul, as well as to physicalist accounts that reduce humans to *nothing but* complex animals.

There are problems with “physicalism,” however. Some believe that the term automatically biases an account of human nature in a reductionist direction – humans are then thought to be entirely controlled by physical laws. Some prefer terms such as “emergent monism,” “dual-aspect monism,” or “biological naturalism” for roughly the same position that I hold.

Among philosophers, the dualism to be defended or defeated is most often body-mind dualism, while in religious contexts, it is body-soul dualism. I shall not make this distinction here, since historically the relation between the terms “mind” and “soul” has varied considerably.

I was influenced in planning my argument by examining this volume’s table of contents. The structure of the book appears to represent a concept of philosophy that has increasingly been called into question during the past half century. The major clue is that, although many of the authors in Parts 1 and 2 have biblical or theological interests, there is a *separate* Part 3 reserved for Christian theological and biblical chapters. Also, if Bible and theology require a section of their own, why not a section for science?<sup>1</sup>

I take this segregation of disciplines to be characteristic of the “analytic” method in modern philosophy: either analysis of concepts such as *mind* (note that the very concept of a concept has shifted during this period), or of linguistic terms such as “mind,” “brain,” “experience.” It was common in the 1970s to introduce students to philosophy by teaching them to differentiate philosophical questions from empirical ones (and also, perhaps from faith questions).

Many recent philosophers, however, have concluded that methods of this sort, expressly contrasting philosophy with empirical studies and with historical study of the origins and

changes in concepts through the centuries, were bound to fail. Many conclude that the notion of “conceptual analysis” needs to be replaced by a (much more empirical) notion of “conceptual archaeology,” that is, by specification of the who and when of a concept’s employment (Stout 1981). For an explanation of this change, see, for example, John Searle’s (1996) “Contemporary Philosophy in the United States” in *The Blackwell Companion to Philosophy*. In brief, it has become apparent that significant conceptual shifts have occurred during the development of Western thought, but have been disguised by continuities in terminology. A relevant example is shifts in the concept of *soul*, which has been obscured by continued use of “soul” and its various modern translations. These discontinuities often need to be discovered by noting shifting relations between philosophical terms and nonphilosophical bodies of knowledge. Hence, the attempt to seal philosophy off from other disciplines has been misguided.

The argument of this chapter, then, will intentionally draw from biblical studies and theology, and from (a bit of) cognitive neuroscience. I claim that both theological and scientific developments are “resonant” with nonreductive physicalism. I borrow the term “resonance” from Warren Brown (2004), who adapts a position on theological method from John Wesley. Wesley believed that the sources for doctrines were Scripture, church tradition, Christian experience, and reason. Brown adds a fifth source, science, and claims that nonreductive physicalism is the position in Christian anthropology that best harmonizes all five sources.

I, however, omit experience as a source because individual experience is so often shaped by the other factors that Brown includes. To illustrate, consider the contrast between Stewart Goetz’s claim that he just finds himself having the belief that he is a soul distinct from his body (Goetz 2005, 33), and Peter van Inwagen’s claim that “when I enter most deeply into that which I call *myself*, I seem to discover that I am a living animal. And therefore dualism seems to me to be an unnecessarily complicated theory about my nature” (van Inwagen 1995, quoted by Goetz 2005, 55).

My rendering of the role of reason is not to take it as a distinct body of knowledge, but as the *use* of reason to solve problems such as inconsistencies in our thinking. In particular, the problem with physicalism is that there is no widely accepted way to show that it need not be reductionistic. The main thrust of my chapter, then, will be to argue that “nonreductive physicalism” is not inconsistent; it is not the case that all we think and do is simply determined by biology. For theological purposes Christians need to pursue concepts of agency, moral accountability, relationality, and so forth.

While many nonreligious philosophers and scientists, from Thomas Hobbes in the seventeenth century to E. O. Wilson in the present, are happy to accept reductionism and determinism in accounts of human life, causal reductionism simply has to be false. If thought can be reduced entirely to the laws of neurobiology, then the very idea of *arguing* for the position makes no sense, since what may appear to be acceptance of a conclusion based on reason must instead be merely the result of meaningless causal processes.

However, as mentioned above, presently there seems to be no agreed-upon strategy for defeating reductionism. Warren Brown and I (2007) have argued, though, that an understanding of downward causation in complex systems allows for the defeat of neurobiological reductionism. So the task of this chapter is to show that *nonreductive* physicalism is philosophically defensible, compatible with mainstream cognitive neuroscience, and is also acceptable biblically and theologically.

## 20.2 Biblical Studies and Tradition

Chapter 27 and Chapter 28 illustrate the complexity of determining the biblical view on the metaphysics of human beings. If I am to argue as a Christian for nonreductive physicalism, it is necessary that I be able to explain how Christians for centuries could have been wrong in believing dualism to be biblical teaching.

Significant help comes from New Testament scholar James Dunn. Dunn (1998) distinguishes between “aspective” and “partitive” accounts of human nature. Roughly, he says, Greek philosophers tended to be interested in partitive accounts: what are the essential parts that make up a human being? In contrast, biblical authors were interested in aspective accounts. Here each “part” stands for the whole person thought of from a certain angle. What the biblical authors are concerned with, then, is human beings in relationship to the world, to one another, and especially to God. Paul’s distinction between spirit and flesh is not the later distinction between soul and body. Paul is concerned with two ways of living: one in conformity with the Spirit of God, and the other in conformity to the *eon* before Christ.

Another part of the answer involves translation. The Septuagint is a Greek translation of the Hebrew scriptures, dating from around 250 BCE. This text translated Hebrew anthropological terminology into Greek, and it then contained terms that, *in the minds of Christians influenced by Greek philosophy*, referred to constituent *parts* of humans. Later Christians have obligingly read and translated them in this way. One instance is the Hebrew word *nephesh*, which was translated as *psyche* in the Septuagint and later into English as “soul.” To see the difference translation can make, compare the King James Version to a recent version of the Bible. The number of times the word “soul” is used has shrunk dramatically. In most cases the Hebrew or Greek term is taken simply to be a way of referring to the whole living person.

However, various meanings can be attributed to *nephesh*. The original meaning is “throat” or “neck.” It can also refer to breath, since the breath flows through the throat. By further extension it signifies a living human being since it is the flowing of the breath that makes a person alive (Gillman 1997, 76). Many older references to the soul, in the Psalms, for instance, fit well with current understandings of body-soul dualism. But consider this example from Psalm 7, verses 1–2. In the King James Version it reads “O Lord my God in thee do I put my trust: save me from all them that persecute me: Lest he tear my *soul* like a lion, rending it in pieces.” In the New Revised Version, it reads “O Lord my God, I take refuge in you; save and deliver me from all who pursue me; or they will tear *me* like a lion.” Notice though, that given the propensities of lions, a more fitting worry would be that the persecutors will tear *my throat* like a lion.

I can only touch briefly on ways in which Dunn’s aspective–partitive distinction might change our reading of Christian tradition (and consequently our propensities to read either dualism or physicalism *back* into the Bible). My example is the reformation controversy regarding the “intermediate state”: the question of Christians’ cognitive state between their deaths and the general resurrection at the end of time. Calvin was a proponent of conscious awareness of God after death, while Luther and many of the radical reformers (those who practiced believers’ baptism and advocated the separation of church and state, also called Anabaptists) adopted some form of “soul sleep,” either an unconscious waiting period or, since sleep is often a euphemism for death, the actual death of the soul. I have argued that although “soul sleep” *seems* to entail substance dualism, it does not in the case of the majority of radical reformers. I came to this recognition serendipitously. In a review by

Terry Hiebert (2008) of my book *Bodies and Souls, or Spirited Bodies?* (2006), Hiebert claims that my nonreductive physicalism is incompatible with the original teachings of (my own) radical-reformation tradition.

Hiebert focuses on three early leaders of the movement, Michael Sattler, Menno Simons, and Balthasar Hubmaier, noting the variety of positions held by these early teachers, and what he calls the “admitted paradoxes surrounding the Anabaptist views of the soul and death” (Hiebert 2008, 187).

Hiebert recounts my appreciation for Dunn’s distinction between aspective and partitive accounts of human nature. I suggest, however, that these are only apparent paradoxes if indeed we distinguish these uses of language. He concludes his review by saying that “historic Anabaptist writers . . . were remarkably unified in using the biblical language of the soul . . . Apparently most Anabaptists . . . were either implicit or explicit dualists” (Hiebert 2008, 195). But then he says: “Anabaptists practiced a restrained use of this interior language of the soul in worship and ministry. They believed in the soul’s engagement with God and people through baptism, the Lord’s Supper, footwashing . . . [and so on]. This rich *metaphor* evokes the moral, emotional, *relational*, volitional, rational, and spiritual *dimensions* of human experience” (Hiebert 2008, 196; emphasis added).

Notice how he has shifted his use of the word “soul” in these two quotations: In the first he is attributing a partitive use of the term to the Anabaptists, but in the second, he is very clearly using “soul” in an aspective way – as a spiritual and relational dimension of human life. With few exceptions all of the uses of the terms “soul,” “spirit,” and “flesh” that he cites in his historical survey are aspective. Hiebert says that for Sattler and Menno, flesh and spirit are scriptural terms contrasting the *life* devoted to Christ with that yoked to Belial, and that both Sattler and Menno referred to the soul as “a believer’s openness to God” (Hiebert 2008, 189). One notable exception in the early history was Balthasar Hubmaier, one of a few theologians who taught that humans are composed of three substances – flesh, soul, and spirit – that separate at death. His *is* clearly a partitive use of the terms.

I hope that this quick look at a small piece of church history not only clarifies Dunn’s concept of aspective uses of anthropological terms, but will also promote a strategy for reading terms such as “soul,” “mind,” and “spirit” in Scripture and theology that does not automatically commit Christians to substance dualism. This conclusion would certainly mean that Christians are free to consider physicalism as an option.

### 20.3 The Convergence of Physicalism with Cognitive Neuroscience

The impact of cognitive neuroscience on current theories of human nature can be summarized roughly as follows: all of the human capacities once attributed to the mind or soul are now being fruitfully studied as brain processes. As is apparent from the first two parts of the present volume, there is no general agreement on what the mind/soul is. If I were to hold any concept of the soul, it would be that of Thomas Aquinas (although it is often questioned whether he should be counted as a *substance* dualist). His is, I believe, the most elaborate and sophisticated account of the functions of the soul in Christian history (in his *Summa Theologiae*, part I, articles 75–102).

As did Aristotle, Thomas attributed souls to plants and animals. The human soul could be said to be tripartite, incorporating all the capacities of both plants and animals, as well as

our particular rational capacities. The vegetative soul is the life principle, and current biologists agree roughly with Thomas that the basic conditions for life are growth, nutrition, and reproduction.

Thomas's list of the capacities or faculties of the animal (or sensitive) soul and the rational soul closely match what contemporary cognitive scientists and philosophers would describe as *mental* capacities, and these are now being related to neural functions and systems. Thomas's list of the sensitive capacities includes locomotion, the five senses, four "internal senses," as well as emotion and appetite. The capacities of the rational soul are intellect and will.

It is enlightening to see how many of these capacities, which Thomas attributed to the soul, are now studied by neurobiologists. The most significant developments are localization studies – that is, research indicating not only *that* the brain is involved in specific mental operations, but that very specific regions or systems correlate with very specific cognitive abilities. Some of the most intriguing research involves what Thomas called the interior senses. Here is but one example. One of the interior senses is the *sensus communis*, which is the ability to collate the deliverances of the five external senses in order to recognize a single object. This is exactly what neuroscientists now call the binding problem. Another is the *vis aestimativa* (estimative power), which is the ability to recognize something as useful (e.g., straw for building nests) or friendly or dangerous. What Joseph LeDoux writes about "emotional appraisal" is relevant to distinguishing this estimative power from the *sensus communis*:

When a certain region of the brain is damaged [namely, the temporal lobe], animals or humans lose the capacity to appraise the emotional significance of certain stimuli [but] without any loss in the capacity to perceive the stimuli as objects. The perceptual representation of an object and the evaluation of the significance of an object are separately processed in the brain. [In fact] the emotional meaning of a stimulus can begin to be appraised before the perceptual systems have fully processed the stimulus. It is, indeed, possible for your brain to know that something is good or bad before it knows exactly what it is. (LeDoux 1996, 69)

So to put LeDoux's findings in Thomas's terms, the *vis aestimativa* is indeed a separate faculty from the *sensus communis*, and it works faster.

Volumes could be written detailing current scientific research on "soulish" capacities, and it is clear that the burden of proof has shifted to the dualists to explain the need to postulate an additional entity, the mind/soul, when accounts in terms of brain activity are becoming increasingly powerful.

## 20.4 Anti-Reductionism and Complex Systems

So far I have provided (brief) arguments for the convergence between biblical and scientific theories of human nature. The point at which many cognitive neuroscientists and Christians would part ways regards reductionism. In fact, some scientists argue that neurobiological study of human capacities would make no sense unless human behavior is reducible to biology (while I claim that their *arguments* would make no sense if reductionism is true).

Neurobiological reductionism is an instance of the broader thesis of *causal* reductionism – the claim that the behavior of all complex entities is determined by

the behavior of their parts. This *is* the case in many of the systems we understand, such as mechanical clocks. These are designed so that the movement of the parts determines the behavior of the whole. The problem is that there has been a tendency throughout the modern period to assume that, when we turn to entities that are too complex to understand in detail, such as organisms, they also must be determined by their parts. For the purpose of defending nonreductive physicalism, the significant parts are brain components, and so there is a very sensible worry that the laws of neurobiology are inevitably determining all of our thoughts and actions.

An obvious answer to the problem of neurobiological reductionism would be to develop and apply a concept of downward causation or whole-part causation. That is, if causal reductionism is the thesis that all causation is from part to whole, then the complementary alternative would be whole-part causation. Alternatively, if we describe a more complex system, such as an organism, as a higher-level system than its biological parts, then causal reductionism is bottom-up causation, and the alternative is top-down or downward causation.

While downward causation was first defined by Donald Campbell in 1974, it received little attention by philosophers until the 1990s, when it began to be employed in philosophy of mind. Jaegwon Kim (1995) has convinced many that the concept is useless for defining nonreductive physicalism. However, I claim that Robert Van Gulick's (1995) account is quite helpful. He pointed out that complex systems or entities are *patterns* that maintain themselves over time, often despite variations or exchanges in their underlying physical constituents. Consider the very simple example of a whirlpool going down a drain. The pattern remains while the water that makes it up is constantly changing. Many patterns are self-sustaining or self-reproducing in the face of perturbing forces that might otherwise destroy them (e.g., DNA patterns). That is, selective activation of the causal capacities of the pattern's parts may contribute to the maintenance and preservation of the pattern itself. These points illustrate that higher-order patterns can have a degree of independence from their underlying physical realizations and can exert downward causal influences without altering the underlying laws of physics.

Van Gulick's position, however, is open to the following objection: The reductionist will ask *how* the larger system affects the behavior of its constituents. To affect it must be to cause it to do something different than it would have done otherwise. Either this is causation by the usual physical means, or it is something "spooky." If it is by usual physical means, then those interactions must be governed by ordinary physical laws, and so all causation is bottom-up after all. Furthermore, even if one uses downward causation to argue against neural determinism, this still does not provide an account of human *agency*. That is, how are we to explain why humans are not just passive players influenced from "below" by biology and from "above" by the environment? Answering this question is the point at which we need to shift to the perspective of complex (dynamical, self-organizing, adaptive) systems theory. In what follows I shall give a brief overview of systems theory and in the process suggest how it describes the emergence of agency.

Alwyn Scott was a mathematician who applied nonlinear mathematics to understand neural processes. He claimed that the development of complex systems theory represents a paradigm shift across all of the sciences, amounting to a new conception of the very nature of causality (Scott 2004). Harold Morowitz (2002) describes systems theory as a revolution comparable in importance to the discovery of language or mathematics. Francis Heylighen (2010) goes further in arguing that "systems thinking" will provide the basis for an entirely new worldview.

Systems thinking has been developing over the past half century, although it has only recently begun to have a significant impact. It draws from a number of sources. There are significant roots in “general systems theory,” developed from the 1950s through the 1970s by thinkers such as Ludwig von Bertalanffy. He was interested in explaining why it was the case that equations of the same form turned out to be applicable in widely different areas of science. Another source was “cybernetics,” so named by Norbert Wiener, which began in the 1940s as the study of feedback control processes in mechanical systems, and has turned out to be essential for understanding regulatory and goal-directed processes in biology. Additional contributions come from information theory, nonlinear mathematics, the study of chaotic and self-organizing systems, and nonequilibrium thermodynamics. Examples of the systems of interest range from autocatalytic processes – that is, self-organizing chemical reactions – at the most basic, to weather patterns, insect colonies, social organizations, and, of course, human brains.

I shall set out some of the essential concepts involved in this change. Several authors call for a shift in ontological emphases. Alicia Juarrero (1999) says that one has to give up the traditional Western philosophical bias in favor of *things*, with their intrinsic properties, for an appreciation of processes and relations. So, for example, from a systems perspective, a mammal is composed of a circulatory system, a reproductive system, and so forth, *not* of carbon, hydrogen, and calcium.

Systems have permeable boundaries, allowing for the transport of materials, energy, and information. The boundary is a matter of the tighter coupling of its components with one another relative to their coupling with entities outside of the system.

Systems are different from both mechanisms and aggregates in that the properties of the components themselves are dependent on their being parts of the system in question. Philosophers distinguish between internal and external relations. External relations do not affect the nature of the *relata*, but internal relations are partially constitutive of the characteristics of *relata*. An assumption of the predominant modern worldview was that the world is composed of *things* related to one another *externally*. Systems theory takes the relations among the constituent *processes* of a system to be *internal*. They also need to be distinguished from entities or systems that are very complicated but do not have these interrelations of functional systems.

Systems range from great stability to wild fluctuation. This is because complex systems are nonlinear, that is, the current state affects the development of each future state. The difference in stability is due to the extent to which the system is sensitive to slight variations in initial conditions, and also to the extent to which there are feedback processes that either do or do not dampen out fluctuations. Systems at the extremes of this spectrum of stability are not of great interest to systems theory; the systems of interest are those in the middle of the spectrum. Chaotic systems are now widely familiar. They result from having a very high sensitivity to initial conditions and their behavior fluctuates wildly, but within a predictable *range* of states. This is often described as the “butterfly effect” – the flapping of the wings of a butterfly on one continent can (supposedly) affect weather elsewhere in the world. This is actually not an appropriate example, however, since there are negative feedback loops that dampen the fluctuations.

More interesting are systems at the edge of chaos. Here the system has the freedom to explore new possibilities and may “jump” to a new and higher form of organization. They are characterized by goal-directedness, at least insofar as they operate in order to maintain themselves. In the process of self-maintenance they create their own components. Evan



Thompson says that a living cell is a paradigm case. Its constitutive processes are chemical; “their recursive interdependence takes the form of a self-producing, metabolic network that also produces its own membrane; and this network constitutes the system as a unity in the biochemical domain and determines a domain of possible interactions with the environment” (Thompson 2007, 44).

We reach a new level of complexity in systems that operate not on the basis of predetermined goals and feedback loops (for example, the homeostatic systems in an organism) but also have the capacity to select their own goals, and thereby adapt to new circumstances. These are called complex *adaptive* systems. When such systems also have some sort of memory, a way of storing information about what has or has not worked in the past, there is heightened ability for the system to increase its adaptation over time. A primary example here is the storage of information in the genome that results in adaptation of the species. The capacity for memory in individual organisms brings us to the point of being able to speak of information and meaning. This adaptive selection opens the possibility of learning and the emergence of novel behavior, based in neural plasticity and the ongoing influence of events outside of the organism.

Complex adaptive systems theory has dramatic consequences for understanding causation. While ordinary efficient causation is presupposed, systems theory developed specifically because such causation is inadequate to describe complex systems. This is in part because they operate on information as much as on energy and matter. The relations among the components of a system need to be thought of as *constraints*. An efficient cause makes something happen. A constraint *reduces* the number of things that can happen, due to the fact that the components are internally related. Thus, a change in one automatically changes the other. Juarrero says that the concept of a constraint in science suggests “not an external force that pushes, but a thing’s connections to something else . . . as well as to the setting in which the object is situated” (Juarrero 1999, 132). So, constraints are relational properties rather than primary qualities in the object itself. Objects in aggregates do not have constraints; constraints only exist when an object is part of a unified system.

From information theory Juarrero employs a distinction between context-free and context-sensitive constraints. For example, in successive throws of a die, the numbers that have come up previously do not constrain the probabilities for the current throw; the constraints on the die’s behavior are context-free. In contrast, in a card game the constraints are context-sensitive: the chances of, say, drawing an ace at any point in the game are sensitive to history because the rules of the game, the number of cards in the deck, and so forth, create relations among the possible outcomes such that the probability of one occurrence is related to all of the others. This account suggests that a better term in place of “downward causation” is Arthur Peacocke’s term, “whole-part constraint” (Peacocke 1995). The “higher-level” system, the whole, does not exert efficient, forceful causation on its components. Rather, global features of the system are such that a change in one component changes the probabilities of the occurrence of other lower-level events. So here is an explanation of Van Gulick’s selection.

Due to the role of probability in complex systems, it is necessary to do away with the sharp distinction between determinism, on one hand, and indeterminism (that is, quantum indeterminacy or complete randomness) on the other. The appropriate middle term is “propensity,” coined by Karl Popper to mean “an irregular or non-necessitating causal disposition of an object or system to produce some result or effect” (Sapire 1995, 657, referring to Popper 1990).

An understanding of the concept of a propensity has been aided by the study of nonlinear mathematics and especially chaotic systems. It begins with a visual or imaginary “state space” or “phase space,” which is an  $n$ -dimensional space. In this space, a trajectory represents possible transitions from one state of the system to another. Think first of a simple two-dimensional graph, say, plotting temperature against time. Now imagine graphs in three or more dimensions.

Chaotic systems theory introduced the concept of a “strange attractor” to describe the development of chaotic systems over time. This is a “shape” in phase space that depicts the boundaries within which the system can be found during its evolution.

From the concept of a strange attractor the idea of an “ontogenic landscape” has been developed. This is a “topographical map” in which valleys represent areas in phase space in which the system is likely to stay. Peaks represent states in which the system will only be found as a result of a major perturbation, such as the injection of a great deal of energy. So the system has a propensity to remain within the valleys. The topography represents a summation of the general effects of a vast number of contextually constrained interactions among the system’s component processes.

In sum, complex adaptive systems theory postulates that such systems become causal players in their own right, partly independent of the behavior of their components, selectively influenced by the environment, and capable of pursuing their own goals. So causal reductionism in general has been called into question by philosophers and scientists in the past generation. The most cogent arguments against causal reductionism are those showing that in many complex systems the whole has reciprocal effects on its constituents. I argue that this set of new concepts, particularly that of context-sensitive constraints, gives us the conceptual tools to explain how downward “causes” cause without violating the causal closure of the physical and without postulating causal overdetermination.

## 20.5 Conclusion

This chapter presents a partial argument for a Christian version of nonreductive physicalism. Its structure is based on the view that a Christian anthropology at a minimum must be: (1) consonant with Scripture and at least a part of the Christian tradition; (2) not in conflict with widely accepted science, and preferably supported by science; and (3) internally coherent. I have provided what I believe are two general ways of reading Scripture and (some) tradition physicalistically: Dunn’s distinction between aspective and partitive uses of anthropological terms; and a likely tale of how Christians came to translate biblical terminology in terms of various conceptions from Western philosophy.

Next, I provided but one example of the way cognitive neuroscience explains faculties once attributed to mind or soul. Finally, I devoted the bulk of the chapter to resources for showing that nonreductive physicalism is a more coherent philosophical position than reductive physicalism. I claimed that the modern bias in favor of reductionism is rapidly giving way to recognition of the existence and nonreducibility of complex systems capable of agency and recognition of meanings.

Along the way I noted that “nonreductive physicalism,” while the term most often used in philosophy, is perhaps not the best for purposes of Christian anthropology because, at least by connotation, it places disproportionate stress on the *aspect* of our physicality. Theologian Veli-Matti Kärkkäinen (in personal conversation) proposes “multi-aspect

monism.” A good way to illustrate the need for a term that recognizes the complexity of human life, and a good way to conclude this chapter is with a quotation from Dunn summarizing Paul’s conception of human nature:

Paul’s conception of the human person is of a being who functions within several dimensions. As embodied beings we are social, defined in part by our need for and ability to enter into relationships, not as an optional extra, but as a dimension of our very existence. Our fleshness attests our frailty and weakness as mere humans, the inescapableness of our death, our dependence on satisfaction of appetite and desire, our vulnerability to manipulation of these appetites and desires. At the same time, as rational beings we are capable of soaring to the highest heights of reflective thought. And as experiencing beings we are capable of the deepest emotions and the most sustained motivation. We are living beings, animated by the mystery of life as a gift, and there is a dimension of our being at which we are directly touched by the profoundest reality within and behind the universe. Paul would no doubt say in thankful acknowledgement with the psalmist: “I praise you, for I am fearfully and wonderfully made” (Ps. 139.14). (Dunn 1998, 78)

## Note

1. Nancey Murphy raises an important point here, and the editors would like to explain that in their original plan, the current volume would be complemented with a further one devoted entirely to the relevance of findings in neuroscience (and other sciences) to the debate between proponents of substance dualism and their rivals.

## References

- Brown, Warren S. 2004. “Resonance: A Model for Relating Science, Psychology, and Faith.” *Journal of Psychology and Christianity*, 23(2): 110–120.
- Campbell, Donald. 1974. “‘Downward Causation’ in Hierarchically Organised Biological Systems.” In *Studies in the Philosophy of Biology*, edited by F. J. Ayala and T. Dobzhansky, 179–186. Berkeley: University of California Press.
- Dunn, J. D. G. 1998. *The Theology of Paul the Apostle*. Grand Rapids, MI: Eerdmans.
- Gillman, Neil. 1997. *The Death of Death: Resurrection and Immortality in Jewish Thought*. Woodstock, VT: Jewish Lights Publishing.
- Goetz, Stuart. 2005. “Substance Dualism.” In *In Search of the Soul: Four Views of the Mind-Body Problem*, edited by J. B. Green and S. L. Palmer, 33–74. Downers Grove, IL: InterVarsity.
- Heylighen, Francis. 2010. “Self-Organization of Complex, Intelligent Systems: The ECCO Paradigm for Transdisciplinary Integration.” Presented at International Symposium, “Research across Boundaries.” University of Luxembourg.
- Hiebert, Terry G. 2008. “Is the Search for the Anabaptist Soul a Dead End? Historic Anabaptism Meets Nancey Murphy’s Nonreductive Physicalism.” *Direction*, 37(2): 185–200.
- Juarrero, Alicia. 1999. *Dynamics in Action: Intentional Behavior as a Complex System*. Cambridge, MA: MIT Press.
- Kim, Jaegwon. 1995. “The Non-Reductivist’s Troubles with Mental Causation.” In *Mental Causation*, edited by J. Heil and A. Mele, 189–210. Oxford: Clarendon Press.
- LeDoux, Joseph. 1996. *The Emotional Brain: The Mysterious Underpinnings of Emotional Life*. New York: Simon & Schuster.
- Morowitz, Harold J. 2002. *The Emergence of Everything: How the World Became Complex*. Oxford: Oxford University Press.
- Murphy, Nancey. 2006. *Bodies and Souls, or Spirited Bodies?* Cambridge: Cambridge University Press.

- Murphy, Nancey, and Warren S. Brown 2007. *Did My Neurons Make Me Do It? Philosophical and Neurobiological Perspectives on Moral Responsibility and Free Will*. Oxford: Oxford University Press.
- Peacocke, Arthur. 1995. "God's Interaction with the World: The Implications of Deterministic 'Chaos' and of Interconnected and Interdependent Reality." In *Chaos and Complexity: Scientific Perspectives on Divine Action*, edited by R. J. Russell, N. Murphy, and A. R. Peacocke, 263–288. Vatican City: Vatican Observatory Press.
- Popper, Karl R. 1990. *A World of Propensities*. Bristol, UK: Thoemmes.
- Sapire, David. 1995. "Propensity." In *The Cambridge Dictionary of Philosophy*, edited by R. Audi. Cambridge: Cambridge University Press.
- Scott, Alwyn. 2004. "A Brief History of Nonlinear Science." *Revista del Nuovo Cimento*, 27(10–11): 1–115.
- Searle, John R. 1996. "Contemporary Philosophy in the United States." In *The Blackwell Companion to Philosophy*, edited by N. Bunnin and E. P. Tsui-James, 1–22. Oxford: Blackwell.
- Stout, Jeffrey. 1981. *The Flight from Authority: Religion, Morality, and the Quest for Autonomy*. Notre Dame, IN: University of Notre Dame Press.
- Thompson, Evan. 2007. *Mind in Life: Biology, Phenomenology, and the Sciences of Mind*. Cambridge, MA: Harvard University Press.
- Van Gulick, Robert. 1995. "Who's in Charge Here? And Who's Doing All the Work?" In *Mental Causation*, edited by J. Heil and A. Mele, 233–256. Oxford: Clarendon Press.

# Against Nonreductive Physicalism

JOSHUA RASMUSSEN

## 21.1 Introduction

What sort of thing are you? One way to think about this question is in terms of your *fundamental nature*. What are the fundamental elements which make you *you*?

The answers on the market divide into two main options. First, there is the *physicalist* option: at the foundation of your nature are physical states of physical things – like particles or fields. On this view, all your thoughts and feelings are either identical with or grounded in physical states. Second, there is the *basic mentality* option: at the foundation of your nature is a conscious substance. On this view, your fundamental nature is witnessed most clearly and accurately from a first-person perspective. You are just what you seem to be when you focus inwardly upon yourself as the bearer of your conscious experiences.<sup>1</sup>

My purpose in this chapter is to develop an argument in support of the basic mentality thesis. I will focus on the nature of *thinking*. I'll argue that thoughts are fundamentally mental – that is, not physical nor grounded in the physical. I will develop my argument over the course of two sections. In the first section, I'll construct a “counting” argument that poses a problem for the identity (mental = physical) thesis. Then, in the second section, I'll extend the “counting” argument in a way that exposes a problem for the dependence (mind *grounded in* physical) thesis. In the final section, I will defend the basic mentality thesis by considering the main arguments for standard physicalism (reductive or nonreductive). The upshot will be that the usual considerations in support of standard physicalism are fully compatible with the basic mentality thesis, whereas standard physicalism is incompatible with the counting principles in my argument. These results, if correct, add weight to the classical option that you are fundamentally a mental substance rather than a physical phenomenon.

## 21.2 A Problem with Identity

In this section, I shall pose a *counting problem* for the identity thesis that mental reality (thinking, feeling, or intending) is physical.

Before we get into the argument, let us get clearer on what it might mean to say that mental reality is *physical*. When I use the term “physical reality,” I have in mind a complete characterization of every physical property (relational and nonrelational) of everything that exists. The term “physical property” is a term of art, but as a first pass, I intend to mean *whatever it is* that physicalists mean when they use the term “physical property.” Here is a fuller definition of “physical property”: a physical property is any property that (i) can be sensed – in principle – from a “third-person” perspective, using any of the five senses, or (ii) is analyzable wholly in terms of such properties. On this definition, physical properties include, for example, shape, size, motion, quantity, vibration, force, and so on. I also include logical constructions (such as conjunctions and disjunctions) of physical properties as physical. This definition accounts for paradigm cases of physical properties. And, as far as I see, my definition successfully excludes paradigm nonphysical properties.

My definition of “physical property” is intended to account for a central usage of the term in the context of the philosophy of mind. That isn’t to say, however, that there aren’t other legitimate usages of “physical property.” John Searle (2004), for instance, uses the term to include macrolevel mental features which he thinks are irreducible to third-person “physical” properties (not even to third-person *macro* physical properties). Thus, Searle’s notion of “physical” includes mental properties which dualists have typically labeled “nonphysical.” To avoid talking past each other, I stipulate that my use of the term “physical” applies more narrowly to *third-person* physical properties (as explained above), whether macro or micro.<sup>2</sup>

Let us turn now to the question at hand: is our mental life physical? Elsewhere, I introduced a counting argument against reducing mental properties to physical properties (Rasmussen 2015). I will offer a new, expanded version of that argument here.

The basic strategy of a counting argument is to show that there is a greater quantity of members of the one category than of some other. To illustrate, consider the categories *integers* and *reals*. These categories both have infinitely many members. But as Cantor famously showed, there is a mathematical sense in which there are more reals than integers: the infinity of reals is greater than the infinity of integers (Halmos 1960). So, although it may already seem intuitively obvious that reals and integers form nonidentical categories, we also have a counting argument against their identity.

Interestingly, we can develop a counting argument against the identity of *mental* and *physical* properties. The argument has two steps. The first step is to show that there are more mental properties than physical properties. This step allows us to infer that the class of mental properties is not the same as the class of physical properties. The second step is to motivate a principle of uniformity, which says that mental properties, whether instantiated or not, are categorically alike. From uniformity, we infer the target conclusion: mental properties are categorically different from – and so irreducible to – physical properties.

We may outline the argument as follows:

A1 There are more mental properties than physical properties, where “properties” ranges over all conceivable properties.

A1(a) There are more *plurals* of physical properties than physical properties.

- A1(b) There is at least one mental property for each plural of physical properties.  
 A2 If there are more mental properties than physical properties, then some mental properties are nonphysical.  
 A3 Therefore, some mental properties are nonphysical.  
 A4 If some mental properties are nonphysical, then all mental properties (or all of a certain class) are nonphysical.  
 A5 Therefore, all mental properties (or all of a certain class) are nonphysical.

Before we consider reasons in support of the premises, a few clarifications are in order. First, when I say there are more mental properties, I do not mean that there are more *actually instantiated* mental properties. So far I'm only concerned with the *nature* of mental properties themselves, not with their exemplification conditions. Consider by comparison that there are more real numbers than integers *whether or not* all reals are actually exemplified. In fact, we don't even need to assume that unexemplified numbers actually exist (such as in Plato's heaven) in order to run a counting argument with respect to numbers; their conceivability, or definability, is enough. Similarly, we don't need to assume that unexemplified mental properties exist in order to run a counting argument with respect to them; their conceivability, or definability, is enough for our purposes.

Second, when I talk of properties, I intend to be neutral with respect to debates over the existence of abstract objects. Identity theorists who think that mental properties are physical properties presumably have some way of understanding property talk, whether or not they happen to believe in abstract objects. Nominalists are welcome to plug in their favorite nominalist translation or paraphrase.

Let us now consider why one might accept the premises of my counting argument. Start with A1: why think there are more mental properties than physical properties? My reason is based upon a procedure for constructing complex mental properties out of more basic ones. To achieve more mental properties than physical properties, we use a building procedure for identifying conceivable mental properties in terms of physical properties. Here is an example: given any plurality of physical properties, let *M* be the property of *thinking that those properties are physical*.<sup>3</sup> The result is that there are at least as many mental properties as *plurals* of physical properties. Next, we observe that there are more plurals of physical properties than particular physical properties. This result follows from a "plural" version of Cantor's theorem according to which there are more *plurals* than *particulars*.<sup>4</sup> Combining these results gives us A1: there are more mental properties than physical properties.

Let us examine the argument more closely. The key which unlocks the argument is a building procedure for constructing mental properties out of plurals of more basic properties. Specifically, for any arbitrary plurality of *physical* properties one might conceive, we can define a mental property in terms of that plurality – such as *thinking that that plurality is physical*. The motivation for this principle is that we see by reflection on our own thinking that thoughts can be *about* anything. For example, I can think about cheese, cornflakes, and bananas, while also pondering a political election. It is the nature of a thought to be about things – any things.

We need to be careful, however, to avoid a certain paradox that can arise from unrestricted building procedures. Take, for example, the following principle:

Construction principle (CP): for *any* properties, the *xs*, there is a mental property of thinking that the *xs* are physical.

Trouble arises because CP allows cases of *self-inclusion*: some xs include the very mental properties which are defined in terms of those xs. Such properties result in paradoxes. For example, a mental property which includes all *non-self-including* mental properties includes itself if and only if it doesn't. Fortunately, where self-inclusion is not in play, the resulting mental properties are unproblematic. Therefore, we may safely sidestep these paradoxes by narrowing our scope to cases which do not involve self-inclusion.

We can give a general building principle which avoids self-inclusion and which entails the specific building principle in our counting argument. Here is a way to do that. Call any property which has, or includes, the form *thinking that such and such* a "thinking-that" property. Then we may work with the following paradox-free principle:

Construction principle 2 (CP<sub>2</sub>): for any non-thinking-that properties, the xs, there is a mental property of thinking that the xs are properties.

This principle puts our focus on *conceivable* mental properties, since it avoids paradoxical cases of self-including mental properties.<sup>5</sup>

We may use CP<sub>2</sub> against the identity thesis as follows. Assume for the sake of argument that the identity theorist is right: every mental property is a physical property. Then the class of physical properties divides into mental properties (P<sub>M</sub>) and nonmental properties (P<sub>N</sub>). Now let M be the class of all mental properties. Then:

- 1 M > P<sub>N</sub> (from CP<sub>2</sub>).
- 2 P<sub>M</sub> ≤ P<sub>N</sub>.
- 3 Therefore: M ≠ P<sub>M</sub>.

Regarding Premise 1, we infer from CP<sub>2</sub> that there are at least as many mental properties as plurals of nonmental properties. Recall next that plurals outnumber particulars (by the plurals version of Cantor's theorem). Therefore, there are more mental properties than nonmental properties. From the identity theorist's perspective, by contrast, there are *not* more mental properties than nonmental physical properties (Premise 2). On their view, physical reality is far more abundant than mental reality, for mental reality is analyzable in terms of a *specific* arrangement or combination of more basic physical states – for example, *being a firing C-fiber within a neural network*. From these premises, it follows that the class of mental properties is not the same as the class of physical mental properties. In other words, not every mental property is physical.

I should emphasize that the counting argument doesn't presuppose a dualist perspective. The argument gets its life from our common ability to see that certain mental properties are distinct from each other. We see, for example, that *thinking a triangle is physical* is distinct from *thinking a square is physical*. Consider, by contrast, that when one compares first-person properties with third-person properties, there is the "opaque context" problem: the problem, basically, is that the appearance of a distinction could be thought to arise from our seeing the same property from two fundamentally different perspectives, first-person and third-person. That worry doesn't bite the counting argument because we are comparing properties from the *same* first-person perspective. So, for example, when comparing *thinking a triangle is physical* with *thinking a square is physical*, we aren't behind an opaque door: we can see clearly that they are distinct.<sup>6</sup>

The final step in the argument takes us to a further conclusion: *no* mental property is a physical property. This step is not as intimidating as it might initially seem. It is firmly



supported by a principle of *categorical uniformity*. Consider that the mental properties under consideration are categorically alike: they differ merely in terms of complexity of psychological content. By contrast, the divide between *physical* and *nonphysical* properties involves much more than a mere difference with respect to complexity of psychological content.

To further illustrate the categorical difference between the physical and nonphysical, imagine building a Lego tower. As you stack more and more Legos, your goal is to transform your Lego tower into something very special: you want to make its shape *nonphysical*. Could you do it? I hope it is clear that the answer is *obviously not*. If (say) *being a stack of  $n$  Lego blocks* is a physical property, then clearly so is *being a stack of  $n + 1$  Lego blocks*, for any  $n$ . The *physicality* of the Lego's spatial structure doesn't turn on the number or arrangement of its components, whether finite or infinite. More generally, the physicality of a physical property doesn't turn on its internal complexity.

Here, then, is a principle of uniformity:

PU: The divide between any two mental properties is narrower than the divide between physicality and nonphysicality.<sup>7</sup>

With PU in hand, it follows that no mental properties are physical.<sup>8</sup> For suppose there were a *physical* mental property  $M_P$  and a *nonphysical* mental property  $M_N$ . Then the difference between  $M_N$  and  $M_P$  would be categorical. In other words, the divide between these two mental properties would *not* be narrower than the divide between physicality and nonphysicality – contra PU. So if you accept PU, then you may infer that all mental properties alike are nonphysical, if any are.

Note that I have not yet said anything about the *grounds* or *realizers* of mental properties. For all that our argument shows so far, it could be that all our mental properties are physically grounded. We'll look at the nature of the grounds of mental properties in the next section.

## 21.3 A Problem with Dependence

In this section, I'll extend the counting argument to reach an even bolder conclusion. I'll explain why I think that mental properties are not even *grounded by* (fixed or necessitated) by physical properties. Note here that this bolder claim is a denial of psychophysical supervenience, which is generally regarded as a "lowest common denominator" commitment of all forms of physicalism (reductive and nonreductive).

When I say that a mental property  $M$  is not grounded by a physical property, I mean minimally this: there is no physical property (possible or actual, micro or macro, local or global) that *entails*  $M$ , where entailment is a relation of metaphysical necessity. More precisely, for any given mental property  $M$ , there is no physical property  $P$ , such that necessarily if  $P$  is instantiated, then  $M$  is instantiated. For ease of presentation, let us say that a mental property *lacks a physical grounding* (possible or actual, micro or macro, local or global) if and only if no physical property grounds it. My thesis, then, is that mental properties lack a physical grounding. Or to be more positive about it: mental properties enjoy fundamentality.

I should clarify at the outset that I am not challenging the premise that the mental properties which you and I instantiate are importantly *connected* to physical properties of

our brain. I propose, rather, that the connection between the mental and physical is *contingent*. As far as my arguments go, it could be that while the instantiation of a physical property can cause the instantiation of a certain mental property, the causal relationship itself depends upon certain contingent psychophysical laws.

For ease of presentation, I will focus on mental properties akin to the ones we considered earlier. In particular, let  $M_{\text{PROPERTIES}}$  be all mental properties of the following form: *thinking that  $x_1$  or  $x_2$  or  $x_3$  . . . is my favorite*. I'll argue first that some mental properties in  $M_{\text{PROPERTIES}}$  lack a physical grounding. Then I'll use a principle of uniformity to generalize the result for all mental properties, or at least for all properties in  $M_{\text{PROPERTIES}}$ .

Here is an outline of the argument:

- B1 Independence: no member of  $M_{\text{PROPERTIES}}$  entails any other member.
- B2 If no member of  $M_{\text{PROPERTIES}}$  entails any other member, then some mental properties lack a physical grounding.
- B3 Therefore, some mental properties lack a physical grounding.
- B4 If some mental properties lack a physical grounding, then all mental properties alike lack a physical grounding.
- B5 Therefore, all mental properties alike lack a physical grounding.

Start with B1: no member of  $M_{\text{PROPERTIES}}$  entails any other member. My reason for thinking that B1 is true is based upon my awareness of individual members of  $M_{\text{PROPERTIES}}$  in my own mind. I begin by noticing that some members of  $M_{\text{PROPERTIES}}$  are individually exemplifiable. Take, for example, this property: *thinking that being a square is my favorite*. Call it " $M_1$ ." I have the privilege of instantiating  $M_1$  right now. Perhaps you do, too. I am thereby able to see that  $M_1$  is *actually* exemplified, from which I infer that  $M_1$  is *possibly* exemplified. I notice next that I can instantiate  $M_1$  without thereby instantiating *other* properties in  $M_{\text{PROPERTIES}}$ . In general, I can think about one set of properties without thereby thinking about some *other* set. That's true even when the "other" set is a subset. For example, I can obviously think *that P or Q is my favorite* without thereby also thinking that *P is my favorite* (and vice versa). I infer, therefore, that each member of  $M_{\text{PROPERTIES}}$  is individually exemplifiable: no member entails any other.

We may further display the above reasoning via analogy. Consider Legos. Suppose you see a small stack of three blue Legos. You infer from your sight of the three blue Legos that it is *possible* for there to be a stack of three blue Legos without there also being an adjacent stack of three *red* Legos. You thus see that red Legos and blue Legos are independent. But could there be a stack of *four* blue Legos without a stack of *four* red Legos? Your ability to imagine the blue stack without a red stack is perhaps one reason to think so. There is another reason: you can see that a mere difference in the number of Legos is manifestly *irrelevant* to the independence between blue and red Legos. Independence doesn't turn on Lego complexity. By the same reasoning, we can see that members of  $M_{\text{PROPERTIES}}$  are independent no matter their complexity. Take any two members,  $M_1$  and  $M_2$ . Both have the same form: *thinking that A or B or . . . is my favorite*. They differ merely in terms of psychological content, but that difference is manifestly irrelevant to their mutual independence.

You might wonder whether there could be a relevant difference between *infinitely* complex and *finitely* complex mental properties. Maybe nothing can be infinitely complex. In that case, the infinitely complex members of  $M_{\text{PROPERTIES}}$  are not exemplifiable, and hence not all members of  $M_{\text{PROPERTIES}}$  are individually exemplifiable.

However, the case of infinite complexity doesn't threaten the heart of my argument. For even if nothing can be infinitely complex, Cantor's results show that infinite complexity is at least mathematically conceivable. Thus, we can still consider whether infinitely complex mental states *would* be independent of each other were they exemplifiable. Compare: even if no Lego structure could be infinitely complex, we can still consider whether an infinitely complex *blue* Lego structure would require the existence of a *red* Lego structure. I think we can see clearly enough that the independence of red and blue Lego structures doesn't turn on their degrees of complexity, regardless of whether there is a degree of complexity which cannot be exemplified. Similarly, I think we can see clearly enough that the independence of members of  $M_{\text{PROPERTIES}}$  doesn't turn on their degrees of complexity, regardless of whether there is a degree of complexity which cannot be exemplified.

The next step is B2: if no member of  $M_{\text{PROPERTIES}}$  entails any other, then some mental properties lack a physical grounding. My reason for accepting B2 is based upon the problem of *too few grounds*. Suppose every mental property has a physical ground. Then since there are more mental properties than physical grounds (per the counting argument), some mental properties must share the *same* physical grounds. In other words, there are too few physical grounds for each mental property to have its own physical ground. It follows that some mental properties must be physically grounded *together*, if they are physically grounded at all. This result is in sharp tension with independence. For suppose  $M_1$  and  $M_2$  must be physically grounded together. Then they are physically *dependent*: any physical state which grounds the one *thereby* grounds the other. Technically, these mental properties could still be independent when *not* physically grounded. But that doesn't help the physicalist. The problem here is that *anyone* can entertain individual members of  $M_{\text{PROPERTIES}}$  individually. When we reflect on our own thoughts in particular, we see that members of  $M_{\text{PROPERTIES}}$  are independent: we can entertain one without the other. Therefore, since the mental properties are independent for minds like ours (i.e., we can think one without the other), there cannot be physical grounds for all of the thoughts that minds like ours can have; there are too few grounds.<sup>9</sup>

The final step is B4: if some mental properties lack a physical grounding, then all mental properties alike lack a physical grounding. My reason for B4 is based again upon categorical uniformity. The idea is that mere differences in complexity, whether finite or infinite, have nothing to do with *having a physical grounding*. It would be really weird (absurd) if, for example, all finite mental properties must have a physical grounding, while all infinite ones must lack a physical grounding. Why would the physical nature of a property turn merely on its complexity? By comparison, suppose there were infinitely many red Legos stacked on top of each other. The structure of that stack clearly wouldn't be nonphysical or nonphysically grounded merely on account of its great complexity. Physicality is a matter of *category*, not *complexity*. If that is correct, then mental properties are uniform with respect to having, or lacking, a physical grounding, no matter how complex they might be.

Let us recap the argument. We observe that members of  $M_{\text{PROPERTIES}}$  are independent of each other: you can think of one member without thinking of another. So, *if* our thinking requires a physical grounding, there must be a unique physical grounding (at least one) for each member of  $M_{\text{PROPERTIES}}$  – so that each one can be thought individually. But there are *not* as many physical properties as members of  $M_{\text{PROPERTIES}}$  (per the counting argument). So, there cannot be a unique physical grounding for each member of  $M_{\text{PROPERTIES}}$ . Therefore, our thinking doesn't have a physical grounding. The final premise is based

upon categorical uniformity: the members of  $M_{\text{PROPERTIES}}$  are alike with respect to having, or lacking, a physical grounding. It follows that if some member of  $M_{\text{PROPERTIES}}$  lacks a physical grounding, then all mental properties in  $M_{\text{PROPERTIES}}$  lack a physical grounding.<sup>10</sup>

## 21.4 In Defense of Basic Mentality

The central thesis that emerges from the counting argument is the *basic mentality* thesis – that at least some of our mental properties are not identical to nor grounded in any physical properties. I will now consider how the basic mentality thesis bears on standard physicalist theories, reductive and nonreductive.

We may classify different versions of physicalism in terms of different kinds of reduction. Here are three kinds:

Microphysical reduction ( $R_M$ ): every mental property of a human person is a micro physical property.

Explanatory reduction ( $R_E$ ): every mental property of a human person is sufficiently explained (causally or ontologically) by micro physical properties.

Ontological reduction ( $R_O$ ): every mental property of a human person is a physical property (global or local).

A variety of physicalist views are definable in terms of the rejection or acceptance of the above reductions. Nancey Murphy (2006), for example, rejects  $R_M$  and  $R_E$ , while she apparently accepts  $R_O$ . Her view, if I understand it, is that mental properties include *macrolevel* properties (“contextualized brain properties,” she calls them) which are physical (i.e., ultimately analyzable in terms of “scientific” third-person properties, such as systems or functions physical states) but not determined by *micro-level* properties. Determination can go the other way – from the “top” down. By contrast, John Searle (2004) rejects  $R_O$  but accepts  $R_E$ .

The basic mentality thesis is incompatible with all three versions of reductive physicalism. According to basic mentality, no mental property is a physical property; hence,  $R_O$  and  $R_M$  are false. Furthermore, basic mentality says – *contra*  $R_E$  – that no mental property has a physical grounding.

Basic mentality is also incompatible with the usual forms of *nonreductive* physicalism. All nonreductive physicalisms have this in common: *physicalism*. Although “physicalism” is notoriously difficult to define, the standard physicalist views include the thesis that mental reality is identical with, or is grounded by, some physical reality (Stoljar 2015). That thesis is incompatible with basic mentality, since basic mentality implies that some mental reality (mental properties) are neither identical with, nor grounded by, any type of physical reality.

So if the basic mentality thesis is true, then standard physicalism (reductive or nonreductive) is false. We have seen one argument for basic mentality. Let us now consider arguments against it. We may classify the main arguments in terms of the following five problems for dualism:

- 1 The problem of psychophysical correlation: how can dualism account for the findings of modern neuroscience, which has a track record of finding purely physical explanations for mental states?

- 2 The problem of complexity: why complicate our ontology beyond necessity by positing undetectable nonphysical substances if we can explain everything in terms of a physicalist ontology?
- 3 The problem of causation: how can there be a causal connection between a purely physical state and a purely nonphysical mental state?
- 4 The problem of pairing: in virtue of what is a nonphysical mental substance paired with *its* physical body, rather than with some other body or no body at all?
- 5 The problem of causal closure: how can nonphysical mental states contribute causally to the physical world if every physical effect already has a completely sufficient physical cause?

Each problem has been discussed extensively in the literature, and each deserves far more attention that I can give to it here. My aim will be modest, then: I will simply express succinctly why I do not personally find any of these problems to be particularly troublesome for basic mentality.

Start with the problem of psychophysical correlation. Basic mentality rules out *necessary* psychophysical laws. But it doesn't follow that there are no psychophysical laws at all. Basic mentality is actually fully compatible with the existence of psychophysical correlations.

Moreover, the *probability* of psychophysical correlation would not necessarily be any lower on basic mentality than on standard physicalism. In fact, it seems to me that the opposite is so. Given basic mentality, mental reality can be explanatorily *prior to* physical reality. This mental priority fits well with a broadly theistic picture on which some original Mind intentionally organizes a world with psychophysical correlations. On standard physicalism, by contrast, I have no such expectation of psychophysical correlations. On the contrary, I find it less likely a priori that there would be an original Mind *given* standard physicalism than given basic mentality. Furthermore, without an original Mind (or minds), I find it vanishingly unlikely that physical reality would happen to unfold in the specific ways evidently required for there to be psychophysical beings to which psychophysical laws might apply.

Different philosophers will of course differ in their assessment of these probabilities. The more important point is that neuroscience cannot in principle tell us how we should assess the relevant *conditional* or *prior* probabilities. That is to say, nothing in neuroscience can, by itself, tell us whether the findings of neuroscience are more likely on standard physicalism than on basic mentality, or whether basic mentality is unlikely a priori. As far as the empirical data goes, one could think psychophysical correlations actually fit better with basic mentality.<sup>11</sup>

Consider next the problem of complexity. The idea here is that positing nonphysical minds complicates our ontology beyond necessity. In reply, consider first that the problem of complexity doesn't target the basic mentality thesis per se, since basic mentality is strictly compatible with a single category ontology, like idealism. Second, and more importantly, the basic mentality thesis is about the nature of something we already know exists: our thoughts. Rather than posit some extra reality, basic mentality frees us from needing to posit nonmental grounds for the mental reality we know about. Finally, if basic mentality complicates one's ontology, then the counting argument itself is a good reason to think the complexity is necessary.

Consider, next, the problem of causation. My main response is that this problem is not narrowly a problem for psychophysical causation. Causation is mysterious in *general*. So,

for example, even physical–*physical* interaction is mysterious: how does an energy field, which has no mass or definitive shape, cause a particle to move? It's perplexing. But it doesn't follow that energy fields *can't* cause particles to move. They surely can and do. Similarly, although it may be mysterious how a nonphysical mental state can cause a physical state, it doesn't follow that it *can't*.

I should add that basic mentality gives us special resources for causation: it allows for the option that mental states have causal powers which are not grounded in the powers of any physical states (local or global). This option, even if mysterious in certain ways, accounts well for our apparent ability to make choices via mental intentions which affect the physical world.

Turn to the problem of pairing: what makes *my* body *mine*? Here is one story I find plausible. I am a single substance with fundamentally mental and physical sides. My body is not itself a substance in its own right. Rather, my body is a *state* of me, and its existence and identity depend in part upon how certain physical systems (heart, lungs, brain, etc.) are functioning. These physical systems are themselves defined teleologically in terms of their role in contributing to my overall well-being. On this theory, a given particle is part of my body when it is caught up in one of these systems which constitute my bodily state. Thus, my body is paired with me by being a state of me. On this account, pairing isn't a problem.<sup>12</sup>

Finally, there is the problem of causal closure, according to which every physical effect has a physical cause. I wonder, though: why think every physical effect has a physical cause? The best arguments I've seen for causal closure are inductive generalizations from our track record of finding physical causes of physical events. These generalizations seem to me to be hasty, however, when applied to human brains. A brain is precisely where I'd expect a nonphysical mental substance to act if I *were* a nonphysical mental substance. Moreover, the evidence we have from neuroscience about the neuroplasticity of the brain is what I'd expect if an agent with nonphysical mental states *could* affect its brain states.<sup>13</sup>

My brief survey of the problems posed against dualism is by no means exhaustive or conclusive, but it does suggest that basic mentality can account for much of the data that is often thought to motivate standard physicalist theories.

## 21.5 Conclusion

We have seen that the usual evidence for physicalism is compatible with the basic mentality thesis. Consider, by contrast, the arguments on the market *against* physicalism. They include (among others): the replacement argument, the zombie argument, Leibniz's Mill argument, and now my counting argument. These arguments have premises whose justification is supposed to be based upon a nonempirical, rational sense, such as the sense that I possibly exist without my body, or the sense that no mere change in motion could by itself give rise to a thought, or the sense that *thinking about a triangle* is not identical to *thinking about a square*. Maybe this sense is unreliable or misguided in each case. My observation here is just that while the basic mentality thesis seems to fit just fine with the usual data offered in support of physicalism, there are reason-based considerations for basic mentality which don't fit well with the standard forms of physicalism (reductive or nonreductive). For that reason, my mind is pressured to think that the basic mentality thesis is true, while standard physicalism is not.<sup>14</sup>

## Notes

1. I have defined “physicalism” here in terms of your fundamental nature. I take this characterization to cover standard physicalist theories (reductive and nonreductive). But see Andrew Bailey (2016) for an account of “materialism” on which your fundamental layer need not be physical.
2. My stipulation is in line with typical formulations of the identity theory. See, for example, Polger (2004).
3. I use “plural” talk instead of “set” in case there is no set of all physical properties. (Some plurals fail to form a set. For example, as Cantor famously showed, the plural of all *sets* doesn’t itself form a set.)
4. To be technically precise, the theorem states that for any plurality, the *xs*, there is no mapping from individuals to the *xs*, such that (i) each member of the *xs* is mapped to at most one subplurality of the *xs*, and (ii) for every subplurality of the *xs*, there is a member of the *xs* that is mapped to it. I give a version of Cantor’s diagonal argument for this theorem in Rasmussen (2015, 194).
5. For a more detailed investigation of parallel paradoxical arguments, see Rasmussen (2015, 238–240).
6. In fact, it seems we can even recognize a distinction between thoughts with *equivalent* contents: for example, one could think *that’s a closed three-sided figure* without thinking *that’s a closed three-angled figure*.
7. For the sake of modesty, we could restrict our scope to the mental properties which are of the form *thinking that P is a property*. The result will still be that many mental properties, including ones you and I have, are not physical. For ease of presentation, I will leave the scope unrestricted.
8. That isn’t to say that no mental properties *could* be physical. My argument doesn’t require the premise that physical properties are essentially physical. I thank Alexander Pruss for drawing my attention to this consideration.
9. Someone might resist this conclusion by supposing that there is some necessary law which applies to physical beings but not nonphysical beings. The idea here is that physical beings are required to think certain of the thoughts in question *together*. This idea is ad hoc, however. Moreover, it conflicts with our experience of our own thoughts. It is apparent, for example, that I can think some state P is my favorite without thereby thinking that some other state Q is my favorite. By categorical uniformity, independence holds for any such thoughts.
10. Alexander Pruss pointed out to me that the conclusion of my argument is compatible with the hypothesis that each mental property is grounded in a *plurality* of physical properties. That’s true. But my conclusion still rules out standard physicalism, since it rules out the standard physicalist theory that the complete physical profile – itself a global physical property – grounds any mental property. Moreover, we can give a counting argument against Pruss’s proposal, since we can generate a superclass of mental properties in terms of *plurals* of  $M_{\text{PROPERTIES}}$  – take, for example, the properties of thinking *disjunctively* about plurals of  $M_{\text{PROPERTIES}}$ .
11. I have been assuming for the sake of argument that every mental state of a human being is indeed correlated with a physical state. Some neuroscientists have challenged that assumption by suggesting that some mental states may have *mental* correlates without any physical basis. See, for example, Beauregard and O’Leary (2008).
12. For a fuller critique of the pairing problem, see Bailey, Rasmussen, and Van Horn (2011).
13. See, for example, Schwartz and Begley (2002) and Beauregard (2007).
14. I am especially grateful to Alexander Pruss for penetrating comments on earlier drafts.

## References

- Bailey, Andrew. 2016. “Materialism Through and Through.” Unpublished manuscript.
- Bailey, Andrew, JOSHUA Rasmussen, and Luke Van Horn. 2011. “No Pairing Problem.” *Philosophical Studies*, 154: 349–360.
- Beauregard, Mario. 2007. “Mind Does Really Matter: Evidence from Neuroimaging Studies of Emotional Self-Regulation, Psychotherapy and Placebo Effect.” *Progress in Neurobiology*, 81(4): 218–236.
- Beauregard, Mario, and Denyse O’Leary. 2008. *The Spiritual Brain: A Neuroscientist’s Case of the Existence of the Soul*. New York: HarperCollins.
- Halmos, Paul. 1960. *Naive Set Theory*. Princeton, NJ: D. Van Nostrand Company.
- Murphy, Nancey. 2006. *Bodies and Souls, or Spirited Bodies?* Cambridge: Cambridge University Press.
- Polger, Thomas W. 2004. *Natural Minds*. Cambridge, MA: MIT Press.

- Rasmussen, Joshua. 2015. "Building Thoughts from Dust: A Cantorian Puzzle." *Synthese*, 192: 393–404.
- Schwartz, Jeff, and Sharon Begley. 2002. *The Mind and the Brain: Neuroplasticity and the Power of Mental Force*. San Francisco, CA: HarperCollins.
- Searle, John. 2004. *Mind: A Brief Introduction*. Oxford: Oxford University Press.
- Stoljar, Daniel. 2015. "Physicalism." In *Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. Accessed August 21, 2016. <http://plato.stanford.edu/entries/physicalism/>.



# Debating Constitutionalism

# Constitutionalism

## *Alternative to Substance Dualism*

LYNNE RUDDER BAKER

Many Christians, as well as Jews and Muslims, have traditionally held that a human person is (or has) an immaterial soul and a material body. Holy Scripture portrays human beings as spiritual entities, and one obvious way to be a spiritual entity is to be (or to have) an immaterial soul that can exist independently of any body. Despite the popularity of the doctrine of immaterial souls, I do not believe that it is required either by the Bible or by Christian doctrine as it has developed through the centuries. I want to show that there is a Christian alternative to immaterialism. I call this alternative “person-body constitutionalism,” or just “constitutionalism” for short.

What I want to do here is to set out my view of persons, according to which there are no immaterial souls, and to show how this view is congenial to a doctrine shared by Christians and Muslims – bodily resurrection and Purgatory or an “intermediate state” (*barzakh*). I conclude with a brief word about the Christian doctrine of Incarnation.

### 22.1 Human Beings without Immaterial Souls

One reason, I believe, that Christians have been drawn to immaterialism is that they think that if we did not have souls, we would be mere bodies, not spiritual beings at all. But if we understand “spiritual beings” as beings capable of having inner lives, this does not follow. I think that it is obvious to each of us that we are capable of having an inner life. What I want to do is to set out a view according to which the capacity of having an inner life does not require that we have immaterial souls.

According to constitutionalism, we are fundamentally persons – whole persons – not minds, souls or brains. Constitutionalism holds, in the first place, that human persons are necessarily constituted by bodies: to have a body is essential to a human person, but it is not essential to have the particular body that one has at some particular time. In the second place, human persons, though not identical with the bodies that constitute them, have no

immaterial parts, and hence no immaterial souls that could exist separately from any and every body. What distinguishes persons from their bodies is that persons are of a kind – the only kind – that has a second-order capacity for a robust first-person perspective essentially. Human persons are spiritual beings – they are capable of having inner lives – in virtue of having robust first-person perspectives.

A robust first-person perspective is the ability to think of oneself without the use of any name, description or demonstrative; it is the ability to conceive of oneself as oneself, from the inside, as it were. In English, linguistic evidence of a first-person perspective comes from use of first-person pronouns embedded in sentences with linguistic or psychological verbs – for example, “I wonder how I will die,” or “I promise that I will stay with you.”<sup>1</sup> If I wonder how I will die, or I promise that I’ll stay with you, then I am thinking of myself as myself; I am not thinking of myself in any third-person way (e.g., not as LB, nor as the person who is thinking, nor as that woman, nor as the only person in the room) at all. Anything that can wonder how *she herself* will die *ipso facto* has a first-person perspective and thus is a person.

A first-person perspective is a dispositional property that has two stages: a rudimentary stage, exemplified both by higher animals and persons, which requires only consciousness and intentionality; and a robust stage, exemplified only by persons with complex linguistic abilities. The robust stage of the first-person perspective is a conceptual capacity that is unique to persons. (For brevity, I shorten “robust stage of the first-person perspective” to “robust first-person perspective.”) The robust first-person perspective is the ability to conceive of oneself from the first-person, as a subject of thought and action. This ability is exercised every time I think, “I’m pleased that I was invited to contribute a paper.” What pleases me is that I myself was invited, not that Lynne Baker was invited or that the oldest woman in the philosophy department at the University of Massachusetts was invited. No, I can entertain thoughts that are self-consciously about myself without using any names or descriptions. This ability manifests my robust first-person perspective.

A person has a first-person perspective essentially. She comes into existence with a rudimentary first-person perspective, and a second-order capacity to acquire a robust first-person perspective, which she develops as she learns a complex language.

Although I could not exist without some body or other, what makes me a person is not a particular kind of body. Rather, what makes me a person is exemplifying a first-person perspective essentially, and what makes me the person I am is being *this* exemplifier of a first-person perspective.<sup>2</sup> Exemplification of a first-person perspective essentially is what unifies a person as an individual, not just a system of parts. The brain makes possible our first-person perspectives, but we have no idea how – any more than we know how it makes possible any of our other mental capacities. The important point is this: what makes any person a person is his or her first-person perspective, not the “stuff” of which he or she is made.<sup>3</sup>

If we are persons with no immaterial souls and we are not identical to our bodies, how are we related to our bodies? On my view, we are *constituted* by our bodies, just as statues are constituted by pieces of bronze, or rugs are constituted by sums of threads. The threads become frayed, a few come loose altogether, but the rug may remain in existence. Therefore, it follows that the rug is not identical to the sum of threads that constitutes it at a certain time. The rug can survive many changes of thread. Similarly, a human person can survive numerous changes. Not only are our cells continually being replaced, but also we can walk

on artificial legs, see with artificial eyes; cochlear implants allow deaf people to hear. A totally paralyzed person with a brain implant can move a computer cursor merely by thinking. A human person can survive enormous changes in her body.

So, according to constitutionalism, a human person is constituted by a body, but is not identical to the body that constitutes her. Someone may ask: If a human person is not identical to a body or to a soul or to a body-plus-a-soul, what is she identical to? This question does not make sense. A person is identical to herself and not another thing. (That's what it means to be nonreductive about persons.)

Let me explain constitutionalism about persons with an analogy: Michelangelo's *David* is essentially a statue. It is not identical to the *David*-shaped piece of marble that Michelangelo carved. If the *David*-shaped piece of marble had spontaneously coalesced in outer space, it would not have been *David*; it would not even have been a statue. But *David* could not have failed to be a statue. No non-statue is identical to *David*. *David* is essentially a statue. And of course, the piece of marble that constitutes *David* is not essentially a statue; it was not a statue when it came out of the quarry. *David* is a statue nonderivatively; the piece of marble that constitutes *David* is a statue derivatively – in virtue of constituting something that is a statue nonderivatively. The piece of marble is a statue during the period of time that it constitutes *David*.

The analogy to persons and their bodies is that persons are related to their bodies as statues are related to pieces of marble, bronze, wood, and so on. Persons are essentially persons (i.e., they essentially have first-person perspectives); statues are essentially statues (i.e., they were intended for display, etc.). During the period that a particular piece of marble constitutes *David*, the piece of marble is a statue derivatively – in virtue of constituting something that is a statue nonderivatively. If *x* constitutes *y* at *t* and *x* is wholly material, then *y* is wholly material. (For details, see Baker 2000, ch. 2.) The human body is wholly material and the human body constitutes the human person. Therefore, the human person is wholly material. A human person is as material as Michelangelo's *David* is.

Whether we are talking about human persons, statues, rivers, or countless other constituted things, the basic idea of constitution is this: when certain things of certain kinds (human organisms, pieces of marble, aggregates of water molecules) are in certain circumstances (different ones for different kinds of things), then new entities of different kinds come into existence. The circumstances in which an aggregate of water molecules comes to constitute a river have to do with the relation of the water molecules to each other; the circumstances in which a piece of marble comes to constitute a statue includes an artist's intention. The circumstances in which a human organism comes to constitute a human person include the organism's developing a brain that can support a (rudimentary) first-person perspective. But in each case, new things of new kinds – rivers, statues, persons – with new kinds of causal powers, come into being.

To summarize this discussion of the idea of constitution: constitution is a very general relation throughout the natural order. Although it is a relation of real unity, it is short of identity.<sup>4</sup> (Identity is necessary; constitution is contingent.) Constitution is a relation that accounts for the appearance of genuinely new kinds of things with new kinds of causal powers. If pieces of marble constitute statues, then an inventory of the contents of the world that includes pieces of marble but leaves out statues is incomplete.<sup>5</sup> And if human bodies constitute persons, then an inventory of the contents of the world that includes human bodies but leaves out persons is similarly incomplete.

The important distinction is between persons and bodies, not between minds and bodies. What we call “minds” are not entities at all, but collections of properties and capacities that we call “mental.” We persons share some of these properties – those whose bearers need only have rudimentary first-person perspectives (namely, intentionality and consciousness) with higher animals. Other of these properties – like wondering how one will die, or being grateful that one is healthy (properties that require robust first-person perspectives) – are exemplified only by persons. So, my solution to the mind-body problem is to say that there are no entities that are minds, no finite immaterial entities that are parts of persons or that can exist apart from bodies. There are rather persons (and some animals) who have all kinds of mental properties.

Now let us turn to the question of whether constitutionalism, this view of persons without immaterial souls, is consistent with Christian doctrines of Resurrection, and of intermediate states between death and resurrection? I believe so. Let us turn to the doctrine of the Resurrection of the body.

## 22.2 The Doctrine of the Resurrection of the Body

All the great monotheistic religions – Judaism, Christianity, and Islam – recognize doctrines of an afterlife. I shall focus on doctrines of resurrection of the dead, and in particular on Christian doctrines. Christian doctrines have two sources. The first source is Second-Temple Judaism, which contributed the idea of resurrection of the body. (The New Testament mentions that the Pharisees believed in bodily resurrection, but that the Sadducees did not believe in an afterlife. Jesus endorsed the former, which was fixed as Christian doctrine by his own bodily resurrection.) The second source was Greek philosophy, which contributed the idea of the immortality of the soul (Cullman 1973).

In what follows, I shall consider Christian views on resurrection and souls and bodies. I shall discuss St Thomas Aquinas’s views in particular, and point to two difficulties it has. Then, I shall show how constitutionalism can avoid Thomas’s difficulties and provide an understanding of the doctrine of resurrection without immaterial souls.

To the early Church Fathers, belief in the immortality of the soul was connected with belief in resurrection of the body. The belief that Jesus rose from the dead was the belief that his soul survived death of the body and was “reinvested with his risen body” (Wolfson 1956–1957, 8). The belief in a general resurrection was the belief that surviving souls, at the end of time, would be “reinvested” with risen bodies. During the interval between death and the general resurrection, a soul would have a life without a body, but a person’s final state would be re-embodied. In this general picture, belief in resurrection includes belief in immortal souls and belief in postmortem bodies (of some sort).

The Christian doctrine of an afterlife is pieced together out of hints and metaphors in Scripture. Jesus’s resurrection is the paradigm case. According to Christian doctrine, Jesus was the Son of God, who was crucified, dead, and buried. The third day he rose again from the dead and ascended into Heaven. Although Jesus’s resurrection is the ground of the Christian doctrine of Resurrection, many questions are left open. Perhaps the most explicit, but still sketchy and metaphorical, account of an afterlife in the New Testament is in 1 Corinthians 15, with its “seed” metaphor. Our bodies are said to be sown in corruption, and raised in incorruption; sown in dishonor, raised in glory; sown in weakness, raised in power; sown a natural body, raised a “spiritual” body. But this passage is notoriously open to

several interpretations. What is a “spiritual body”? Is it made of the same flesh-and-blood particles as the premortem body? Of the same kind of particles if not exactly the same ones? Of some entirely different kind of stuff? There is no unanimity.

The doctrine of Resurrection raises further questions. For example, is there immediate resurrection at the instant of death, or is there a temporary mode of existence (an intermediate state) before a general resurrection at the end of time? There is no general agreement. But whatever the details of the conception of an afterlife, there are three characteristics of the Christian view of resurrection: First, it is miraculous. Unlike the classical Greek doctrine of the immortality of the soul, life after death does not occur naturally, and is not subject to natural law. It occurs only by the Grace of God. Second, life after death concerns the identity of the human being, the person. The very same individual person is to exist in the afterlife as exists today. The person does not merge with the universe, or with an eternal mind. “Survival as” in a sense of, say, psychological similarity is not enough. The person retains her particular identity after death. Third, resurrection is bodily. Resurrected people are embodied. St Paul in 1 Corinthians says that resurrection bodies will be “spiritual” or “imperishable” or “incorruptible,” depending on the English translation.

Philosophically speaking, the question of personal identity in particular stands out: In virtue of what is a person in an afterlife identical to a certain person in a premortem state? Let us begin by considering the view of personal identity of one of the great Christian philosopher-theologians, St Thomas Aquinas.

## 22.3 Thomas Aquinas on the Afterlife

There are many schools of thought about Thomas Aquinas and Aristotle’s influence on him. One of Aquinas’s contributions to Christian thought was to give an account of what happens between death and resurrection in terms of the subsistence of the rational soul. On one traditional interpretation, Aquinas’s view has the advantage over the substance dualists like Plato and Descartes in that it gives a reason why resurrection should be bodily resurrection: the body is crucial for a complete substance.

On the traditional interpretation, Aquinas took over Aristotle’s framework for understanding human beings, modifying it as little as possible to accommodate Christian doctrine. According to Aristotle, a human being is a substance; a substance is formed matter. The body supplied the matter, the soul the form. On Aristotle’s view, a rational soul can no more exist apart from the body whose form it was than shape of a particular ax could exist apart from that ax. The soul is the form of the body. So, Aristotle had no place for an unembodied soul.

Aquinas followed Aristotle in holding that the soul is the form of the body (Aquinas 1945, 1.75). A soul is not a human being. The soul provides the form for the material body: A human being is a substance; it is formed matter. So far, Aquinas agreed with Aristotle.

However, unlike Aristotle, Aquinas held that the soul is a substantial form that could “subsist” on its own; it is “something immaterial” (Aquinas 1945, 1.75.2). Aquinas believed that there would be a general resurrection at the end of time. The human being – the substance, the individual – does not exist as such between death and the general resurrection. What continues through the interim is only the rational soul which “subsists” until reunited with the body, at which time the human being is fully recovered. While the soul is disembodied, the soul is *not* the person who died. It is merely a remnant of the

person, awaiting reunion with the person's body. It is only when the soul is reunited with the person's body (the same one) that the person resumes life.

However, a philosopher may worry that Aquinas's account commits him to a new ontological category of being: the rational soul as a subsisting entity that is not a substance. The rational soul is not really an individual, but a kind of individual-manqué. We can say very little about this new kind of entity except that it seems to fill Aquinas's need to combine Aristotle's ideas with the Christian doctrine of an afterlife. It would be desirable to make sense of a Christian doctrine of Resurrection without appealing to a new and strange kind of entity – a “subsisting” entity that is not a substance – as I shall try to do in a moment.

Another question that immediately arises about the idea of a disembodied soul concerns the question of individuating disembodied souls at a time. What makes a disembodied soul Smith's soul, rather than Brown's soul? If human beings (Smith or Brown) were at issue, the answer would be clear: Smith and Brown are substances, and, on Aristotle's view, substances are individuated by their matter. But Smith's or Brown's disembodied soul is not a substance, but a substantial form.

Aquinas gives a way to distinguish between Smith's and Brown's souls between death and the general resurrection. Separated souls are individuated by the bodies that they long for. Each separated soul has an affinity for the body with which it was united during earthly life. Even when Smith's soul is disembodied, what makes Smith's soul *Smith's* soul – and not Brown's soul, say – is that Smith's soul has a tendency and potential to be reunited with Smith's body, and not with Brown's body.<sup>6</sup>

There remains a difficulty: If Smith's body is individuated by its being the body that Smith's soul longs for, what makes a body Smith's body? The answer cannot be that a body is Smith's body because it is the body that Smith's soul yearns for, when we just said that a soul is Smith's soul because it yearns for Smith's body. To put it another way, we cannot both identify a body as Smith's by being the body that Smith's soul longs for, and identify Smith's soul as the soul that longs for that body. That is a very tight circle.

Aquinas cannot get out of this circle by appealing to his (and Aristotle's) view that matter individuates. If matter individuates, what makes a soul Smith's soul is the body that Smith's soul longs for. But what body is that? Since the soul is the form of the body, on Aristotle's view, there exists no body without a soul. Without a form, there would just be unformed matter. Unformed matter is not actual; it is only potency – only potentially a body. So, on Aristotle's view, there is no actual thing for a subsisting soul to long for. If matter individuates, an immaterial soul cannot be Smith's soul in virtue of its yearning for Smith's body since, in the absence of a soul, there is no body that is Smith's body. The Aristotelian view that matter individuates seems not to be reconcilable with the view that the soul is a substantial form that can “subsist” – and experience God – apart from a body. (The problem does not arise for Aristotle since on Aristotle's view, there are no separated souls, just human beings.)

If matter individuates, then an immaterial soul separated from a body is not Smith's or anybody's soul. However, Aquinas could reject Aristotle's view that matter individuates. He could appeal to God's divine action in creating Smith's soul: He could say that what makes a soul (separated or not) Smith's soul is that God created it to be Smith's soul before Smith was born. Nevertheless, I do not believe that Aquinas actually rejected the Aristotelian view of individuation.

## 22.4 Constitutionalism and Resurrection

I believe that constitutionalism can handle the two problems that fall out of Aquinas's view. First, constitutionalism does not need a new ontological category of disembodied souls as subsisting entities that are not substances; and, second, that constitutionalism can say which body is Smith's body in the afterlife without appeal to a soul: God, who traditionally freely decrees every contingent state of affairs, freely decrees that a certain resurrection body exemplify Smith's first-person perspective. Since constitution is a contingent relation, which body constitutes Smith in the afterlife is a matter of God's free decree (Baker 2011).

Moreover, if I am right about the three features that characterize the doctrine of Resurrection – miracle, identity of person, and embodiment – then constitutionalism with respect to human persons provides a good metaphysical backdrop for the doctrine of Resurrection, more satisfactory than mind-body dualism.

First, consider essential embodiment. Being essentially embodied does not imply that we essentially have the bodies that we in fact have. We could have different bodies, and if we are resurrected, we will have different bodies. This is implied by St Paul when he says, "What I mean, my brothers, is this: flesh and blood can never possess the kingdom of God, and the perishable cannot possess immortality" (1 Corinthians 15:50). Our bodies now are perishable, but in the resurrection we will have imperishable bodies. This leads to a simple argument, letting Smith be a person who will be resurrected:

- 1 The body Smith has now is perishable.
- 2 The body Smith will have in the resurrection is imperishable.
- 3 If  $x$  is perishable and  $y$  is imperishable, then  $x$  is not identical to  $y$ .
- ∴ 4 The body Smith has now is not identical to the body that Smith will have in the resurrection.

Let me defend this simple argument. It is valid: the premises entail the conclusion. But are the premises true? First, consider (1): The body that Smith has now is a biological body – a carbon-based organism – and all carbon-based organisms are subject to decay and hence are perishable. Consider (2): Resurrected bodies are supposed to be eternal, and whatever is eternal is imperishable. Consider (3): (3) is likely to be more controversial. Can God not transform Smith's body that is perishable now into a body that is imperishable? Certainly, he can. But to do so is to effect a substantial change: Smith's new imperishable body would not be the same body as Smith's current perishable body. Why not? Perishability and imperishability are persistence conditions. Objects have their persistence conditions essentially: a single object cannot be perishable at one time and imperishable at another time. Hence, the perishable body that Smith has now is not identical to the imperishable body that Smith will have in the resurrection. I think that it follows that Smith's resurrection cannot coherently be understood as a matter of rejoining Smith's body with Smith's soul.

Here I just want to draw attention to the point that a resurrection body cannot be the same body as a biological body, and that constitutionalism can allow for a change of body without appeal to an immaterial soul.

Now consider identity of a person over time. On the constitution view, identity of a person is identity of exemplification of a first-person perspective. There is no informative



criterion for identity of a first-person perspective over time. It is just a brute fact about some future person that I am (or am not) she. I do not think that this is a shortcoming of my view. If there were an informative criterion of identity over time of persons, it would be in nonpersonal terms. That is, it would be reductive (e.g., continuity of organic functioning or continuity of psychological states or continuity of brain states). But there is a strong religious reason to hold that there is no reduction of persons to nonpersonal entities.

If Christ died for our sins, or if God punishes us for our sins, the object of attention is the sinner – that is, the person, not some subpersonal features to which a person can be reduced. For example, suppose that Smith sinned by lusting after Mrs Jones and that the lust was constituted by some complex brain state. God does not punish the brain state. It is the whole person who is subject to punishment. And if the person is not reducible to subpersonal features, then there is no informative, noncircular criterion of personal identity over time. So, we can hold that personal identity consists in sameness of first-person perspective, while recognizing that this is not an informative, noncircular criterion of sameness of first-person perspective.

Finally, consider the miraculous nature of resurrection. In the natural course of affairs, human bodies decay and are not replaced by, or changed into, resurrection bodies. However, the domain of natural laws is nature. And God is supernatural – omnipotent, omniscient, and perfectly good. So, there is no conflict between natural laws and God’s power to bring about resurrection. As one scholar put it, “God can make the body of Peter out of the dust that was once the body of Paul” (Walker Bynum 1995, 260).

## 22.5 Between Death and Resurrection

Some – but not all – Christians believe that there is a kind of existence after death and before resurrection. Roman Catholics believe in Purgatory, a time of purification and punishment for sin, during which the disembodied soul exists until it is reunited with “its” body. We have just seen the difficulties of supposing, as Aquinas does, that apart from a soul there is a body that is “its” body (that soul’s body) to which the soul can be reunited. So, it seems that mind-body dualism with the possibility of disembodied souls does not help us understand Purgatory.

Without appealing to Purgatory, some Protestants hold that there is an “intermediate state” between death and a general resurrection at the end of time, and at least one Protestant theologian – John W. Cooper – has argued that an intermediate state entails mind-body (or soul-body) dualism.<sup>7</sup> However, there is no reason – Biblical or philosophical – to suppose that the intermediate state must be a *disembodied* state. For all we know, persons in the intermediate state (assuming that there is one) are constituted by intermediate-state bodies. If God can so transform or replace our bodies once – as he does in resurrection – he can do it twice. So, the arguments about the intermediate state provide no reason to prefer soul-body dualism to constitutionalism.

## 22.6 A Brief Word about the Christian Doctrine of Incarnation

The doctrine of the Incarnation, which takes Jesus Christ to be a person fully human and fully divine, requires a slight modification of constitutionalism as I have presented it. Christ

is not essentially embodied. He became embodied at a certain point in time. So, I should restrict my claim to saying that all beings that *begin existence* as human persons (i.e., are constituted by bodies at the beginning of their existence) are essentially embodied.

Constitutionalism seems to have an advantage over mind-body dualism about Christ's nature: His human nature is wholly material and his divine nature is wholly immaterial. By contrast, mind-body dualism, which holds that human persons have immaterial minds, seems to imply that Christ has two immaterial minds – one human and one divine. It is surely more straightforward and elegant to treat Christ's human nature as wholly material and Christ's divine nature as wholly immaterial.

## 22.7 Conclusion

We are human persons, not because we have immaterial souls, but because we have first-person perspectives essentially. Biologically, we are continuous with nonhuman animals, but ontologically, we are unique. So, if constitutionalism is right about our place in nature, I think that it would be good news for Christians – and perhaps for Muslims and Jews as well.

## Notes

1. Hector-Neri Castañeda developed this idea in several papers (see Castañeda 1966, 1967).
2. For details on what makes you the particular person you are, see Baker (2013, 154–156 and 179–182).
3. The kind of first-person perspective that I have just described is robust; it is tied to language. A human infant who lacks a language is nonetheless a person; the infant is born with a rudimentary first-person perspective that typically develops into a robust first-person perspective.
4. Some philosophers have held that the idea of unity without identity is incoherent. In *Persons and Bodies*, I give a completely general definition of “constitution” that is coherent.
5. There is much more to be said about the idea of constitution (See Baker 1999, 2000, especially ch. 2).
6. This reply is not available to proponents of immaterial souls – like Plato or Descartes – who take a human person to be identical to a soul.
7. The immaterial soul is taken immediately to Christ, and later when all the saved are resurrected, it will be reunited with “its” body – a phenomenon that we just saw to be problematic (Cooper 1989).

## References

- Aquinas, Thomas. 1945. *Summa Theologica I: Questions 75–89*, translated by Anton C. Pegis. New York: Random House.
- Baker, Lynne Rudder. 1999. “Unity without Identity: A New Look at Material Constitution.” In *New Directions in Philosophy: Midwest Studies in Philosophy 23*, edited by Peter A. French and Howard K. Wettstein, 144–165. Malden MA: Blackwell.
- Baker, Lynne Rudder. 2000. *Persons and Bodies: A Constitution View*. Cambridge: Cambridge University Press.
- Baker, Lynne Rudder. 2011. “Christian Materialism in a Scientific Age.” *International Journal for Philosophy of Religion*, 69(1): 1–12.
- Baker, Lynne Rudder. 2013. *Naturalism and the First-Person Perspective*. New York: Oxford University Press.
- Castañeda, Hector-Neri. 1966. “He: A Study in the Logic of Self-Consciousness.” *Ratio*, 8: 130–157.
- Castañeda, Hector-Neri. 1967. “Indicators and Quasi-Indicators.” *American Philosophical Quarterly*, 4: 85–100.

- Cooper, John W. 1989. *Body, Soul and Life Everlasting: Biblical Anthropology and the Monism-Dualism Debate*. Grand Rapids, MI: Eerdmans.
- Cullman, Oscar. 1973. "Immortality of the Soul or Resurrection of the Body." In *Immortality*, edited by Terence Penelhum, 53–85. Belmont, CA: Wadsworth Publishing.
- Walker Bynum, Caroline. 1995. *The Resurrection of the Body in Western Christianity, 200–1336*. New York: Columbia University Press.
- Wolfson, Harry A. 1956–1957. "Immortality and Resurrection in the Philosophy of the Church Fathers." *Harvard Divinity School Bulletin*, 22: 7–40.

# Against Constitutionalism

ROSS INMAN

## 23.1 Constitutionalism

As a metaphysic of human persons, constitutionalism in its most general form is the view that human persons are constituted by their bodies, but are not strictly identical to them. The relation between human persons and their bodies is that of *constitution*, a type of unity relation whose relata are strictly nonidentical; “constitution is not identity,” as the phrase goes.

Proponents of constitutionalism maintain that the constitution relation is a very general relation that is ubiquitous in the natural world. Bronze statues are constituted by pieces of bronze; dollar bills and diplomas are constituted by pieces of paper; rivers are constituted by an aggregate of water molecules. Constitutionally related objects, such as the human person, one’s body (which many construe as identical to a living human animal), and the aggregate of particles that (immediately) constitutes one’s body, are nonidentical material objects that all share the same matter – indeed are atom-for-atom qualitative duplicates – and are thereby spatially coincident at the times at which they are so related. Each of these objects all belong to distinct primary kinds and, as a result, each differ with respect to their modal profiles, that is, their modal properties and persistence conditions. As the literature on constitutionalism is plentiful – the proponents and critics of the view are many – I will interact principally (though not exclusively) with the brand of constitutionalism defended by Lynne Rudder Baker (Baker 2000, 2007, 2013) in what follows.

For Baker, human persons are constitutionally related to their bodies (which are strictly identical to living, human animals) without being strictly identical to them. As a genuine relation of unity, constitution facilitates a kind of mutual property-sharing between the *constituting* human animal and the *constituted* human person, each of which belong to distinct primary kinds (which answers the question “What most fundamentally is *x*?”). I borrow the property of *being 5 feet 11 inches* from the human animal that constitutes me, and the human animal borrows the property of *thinking about the metaphysics of human persons* from me, the person. In Baker’s terminology, the human animal has the property of *being 5 feet 11 inches* nonderivatively at *t* in the sense that it does not have it in virtue of

being constitutionally related to me at  $t$ . More generally, if  $x$  constitutes  $y$  at  $t$ , and  $x$  is  $F$  at  $t$  independently of its standing in constitution relations to  $y$  at  $t$ , then  $x$  has  $F$  *nonderderivatively* at  $t$ . I, on the other hand, have the property of *being 5 feet 11 inches derivatively* at  $t$  in that I have the property in virtue of my being constitutionally related to an object at  $t$  that has the property nonderivatively at  $t$ , that is, independent of constitution relations.

So what exactly marks the fundamental difference in kind between human persons and human animals for Baker? For Baker, it is the possession of a first-person perspective *essentially* and *nonderderivatively* that sets human persons apart from animals. Persons alone have as part of their nature “the ability to conceive of oneself as oneself, from the inside, as it were” (Baker 2007, 69). When a human animal or organism develops a first-person perspective in this sense, a new entity comes into existence, a human person.<sup>1</sup> Human persons are *human* in virtue of being constituted by human animals (*Homo sapiens*), and are *persons* in virtue of exemplifying a first-person perspective nonderivatively. Unlike the biological persistence conditions of human animals, human persons have unique first-person persistence conditions such that they exist as long as their first-person perspective is exemplified.

In my estimation, the chief virtue of Baker’s constitutionalism is that it aims to uphold a view of human persons as fundamental in some sense; persons carve out genuinely unique and irreducible joints in reality. This emphasis on the fundamentality or metaphysical priority of human persons is to be commended in my view, particularly in light of the ever-increasing penchant for ascribing metaphysical priority to the very small (particles as on microphysicalism) or the very large (the cosmos as a whole as on priority monism).<sup>2</sup>

## 23.2 Against Constitutionalism

I will assume going forward that constitution-without-identity is a coherent metaphysical concept, and thus leave aside discussion regarding the adequacy of the machinery standardly used in formulating constitutionalism.<sup>3</sup> Instead, in this section I want to explore what I consider to be a few of the most trenchant criticisms of constitutionalism, with a particular emphasis on the brand of constitutionalism defended by Lynne Rudder Baker as well as her rejoinders to such criticisms.<sup>4</sup>

### 23.2.1 The grounding problem

Perhaps the most common objection leveled against constitutionalism is what is known as the *grounding (or indiscernibility) problem* (Olson 2001a; Zimmerman 1995). Recall that on constitutionalism you (“Person”), the human animal (“Animal”) that constitutes you, and the aggregate of particles that constitutes the human animal (“Aggregate”) are atom-for-atom qualitative duplicates, they share precisely the same physical profile (same size, shape, structure, weight, texture, and made of the same bits of matter); as such, each object is empirically indistinguishable and shares precisely the same nonmodal profile. Yet, on constitutionalism, Person, Animal, and Aggregate each have radically different modal profiles, including different persistence and identity conditions. But if these objects share precisely the same physical and nonmodal profiles, then what accounts for such radical modal differences between them, say, an Animal’s ability to continue to exist despite a brain injury that brings about the cessation of Person? Or what, for example, explains the modal

differences in the way these objects exhibit a first-person perspective, Person having a first-person perspective essentially and nonderivatively where Animal and Aggregate do not, given that they are empirically indiscernible neuron-for-neuron duplicates? Herein lies the grounding problem for constitutionalism.

Baker's primary response to the grounding problem is to reject the latent assumption in the objection that a difference in modal profiles entails a difference in nonrelational (nonmodal) properties.<sup>5</sup> More specifically, Baker contends that the modal profile of material objects can be fixed by relational as well as nonrelational properties. She remarks that the objection "is a worry only on the assumption that the nature and identity of a thing are determined by its actual intrinsic physical properties. But we have independent reason to think that ordinary things – like statues, flags, carburetors, and passports – have relational properties essentially" (Baker 2001, 170). If constitutionally related objects have at least some relational properties essentially, then they need not differ in their modal properties and persistence conditions in virtue of any nonrelational property, as the objector assumes.

What are we to make of this appeal to essential relational properties to avoid the grounding problem? While the appeal to essential relational properties might carry some weight for *intentional* objects like artifacts or artworks (objects the existence and identity of which are intention-dependent), the move lacks force when it comes to grounding the relevant difference in modal profile for *nonintentional* objects like Person and Animal in particular. What relational property of Animal, for instance, accounts for the radical differences in modal properties and persistence conditions between it and Person? As far as I can tell, Baker does little to fill out the details as to how exactly her appeal to relational properties might help ground the difference in modal profile for non-artifact kinds such as *human person* and *human animal* in particular.

But consider the following reason to think that Baker's appeal to essential relational properties falls short even in the case of intentional objects like artifacts and artworks. One can construct a case where constitutionally related objects share all nonrelational *and* relational (nonmodal) properties. If so, then it remains difficult to see what grounds remain for thinking that constitutionally related objects differ in their modal profiles by way of differing with respect to some relational property.

Taking a cue from Alan Gibbard's (1975) famous thought experiment, suppose an artist fashions from two distinct lumps of clay a statue of the bottom half of the giant Goliath at  $t$ , and a statue of the top half of the giant at  $t_1$ , where each piece of clay constitutes each respective statue at each time. Now suppose, as seems plausible, that in bringing the two statues together at  $t_3$ , the artist brings into existence a new statue, *Goliath*, as well as a new lump of clay, *Lumpl*. The artist then puts the new statue on display in his local art gallery where it is admired for its exquisite proportion and symmetry, and in time is considered by the majority of art critics as one of the most beautiful clay statues ever produced. Now suppose the statue is completely incinerated at  $t_4$  by the artist's closest friend in a fit of envy, simultaneously destroying *Goliath* and *Lumpl*.

Here we have a case of constitutionally related objects, *Goliath* and *Lumpl*, that not only share precisely the same temporal boundaries ( $t_3$ – $t_4$ ) and are exact atom-for-atom qualitative duplicates during their careers, but they also share all *historical* properties and stand in precisely the same *relations* to the artworld. Both *Goliath* and *Lumpl* are brought into existence by the same artist, at the same time, for the same aesthetic purposes, and both stand in the same relations to the wider artworld for the same temporal duration.

Nevertheless, according to Baker, *Goliath* and *Lumpl* are constitutionally related objects that differ radically in their modal profile, in virtue of having different relational properties throughout their spatiotemporal careers. But if *Goliath* and *Lumpl* are qualitatively identical both with respect to their physical construction as well as their relational and historical properties at each time they exist, what grounds remain for claiming that some relational property determines the modal profile of one and not the other? Consequently, the radical difference in modal properties and persistence conditions lacks a difference-maker.<sup>6</sup> At the very least, the above case suggests that Baker's rejoinder in terms of essential relational properties fails even to generalize to all cases of constitution regarding intentional objects like artifacts and artworks.

But the move to essential relational properties is just as untenable when it comes to grounding the modal differences for nonintentional objects such as Person and Animal. What relational properties of Animal, for example, might account for its differing so radically from Person in its modal profile, its modal properties and persistence conditions?

Perhaps the most promising route for Baker here would be to argue that Animal, as a member of the biological species *Homo sapiens*, is what it is in virtue of specific relational and historical facts. In fact, Baker (Baker 2007, 63) explicitly rejects the notion that kind-membership for biological organisms amounts to exemplifying "underlying intrinsic properties" (Baker 2007, 63). By her account, kind-membership for organisms is fixed purely by relational properties, in particular, genealogical lineage. The view that biological species are purely relational and historical in that they are constituted exclusively by evolutionary lineage on the phylogenetic tree is known as "cladism" in contemporary philosophy of biology.<sup>7</sup> Thus Baker might call on a cladistic concept of biological species to help ground the difference in modal profile between Animal and Person.

While cladism rightly emphasizes the importance of ancestral lineage for the classification of biological organisms, it is susceptible to several notable objections, only one of which I am able to rehearse here.<sup>8</sup> Even if we were to uncritically grant that relational or historical properties concerning ancestral lineage are necessary for biological species membership, such properties are arguably not sufficient; the nature and identity of biological organisms cannot be fully captured by where they come from.

To see this, we can once again construct a scenario resembling Gibbard's above thought experiment, although this time adapted to the generation and extinction of genealogical lineages and their respective biological members (see Oderberg 2007, sec. 9.2). Take the event of an existing genealogical lineage  $l$  splitting into two lineages  $l_1$  and  $l_2$ , where the splitting constitutes a genuine speciation event,  $e_1$ ; on cladism, the splitting of  $l$  into two distinct lineages  $l_1$  and  $l_2$  in  $e_1$  suffices to generate two distinct biological species. Now suppose that some time after the occurrence of  $e_1$ , the members of  $l_1$  and  $l_2$  simultaneously cease to exist as a result of a single extinction event,  $e_2$  (due to a meteorite perhaps). In this case, we have two distinct lineages  $l_1$  and  $l_2$  consisting of distinct members, but nevertheless share the very same ancestral lineage ( $l$ ) and are bounded by the very same speciation and extinction events  $e_1$  and  $e_2$ .

But if biological species are purely relational and historical as on cladism, it is difficult to see what grounds the difference in biological species between members of  $l_1$  and  $l_2$  given that they share precisely the same genealogical nexus on the evolutionary tree. As with the case above with *Goliath* and *Lumpl*, the case suggests that what fixes biological species membership amounts to something more than mere historical or relational properties.

A purely relational or historical conception of biological species, then, is not sufficient to account for Animal's modal profile. And on the operational assumption that Person and Animal are qualitatively indistinguishable with respect to their nonrelational (nonmodal) properties, no properties remain to act as the relevant difference-maker regarding their difference in modal profile.

But one can argue that not only are historical or relational properties not sufficient for biological species membership, they are not necessary either. For suppose, borrowing from Donald Davidson's (Davidson 1987, 443–444) famous Swampman thought experiment, that a bolt of lightning destroys Animal while, at the same time, another bolt strikes the surface of a swamp and spontaneously produces an atom-for-atom duplicate of Animal. The newly generated Animal shares all of the same qualitative features, including sub-systems, functions, and behaviors as the original Animal, yet lacks a causal history and an ancestral lineage. If ancestral lineage is necessary for biological species membership, the newly generated Animal would not only fail to be a human animal and a member of the species *Homo sapiens*, it would fail to be a member of *any* biological species at all. But given that the original Animal and the newly generated Animal are atom-for-atom duplicates, this seems deeply problematic.

Consequently, Baker's appeal to essential relational properties as a solution to the grounding problem seems untenable for both intentional and nonintentional objects alike.

### 23.2.2 *Constitutionalism and the zombie problem*

Eric Olson (2016) has recently argued that certain forms of constitutionalism have the untoward consequence that zombies are both actual and ubiquitous. Currently roaming the earth are several billion (7.4 billion to be precise) exact physical duplicates of human persons that are nevertheless devoid of a mental life. Where zombies were once thought to be exclusively denizens of possibility-space (if at all) and acclaimed residents of philosophers' thought experiments, Olson argues that it is a consequence of most forms of constitutionalism that there is at least one living, breathing zombie within the spatio-temporal boundaries where each human person is located.

Why think so? Olson points out that if the human animal that constitutes the human person is in fact conscious, then it would be puzzling to learn, given that both are atom-for-atom (and neuron-for-neuron) physical duplicates, that the person and the animal diverge significantly when it comes to their mental lives (say, one's possessing simple, low-grade sentience and the other higher-order rational thought). It is for this reason, argues Olson, that most defenders of constitutionalism are not inclined to attribute mentality to human animals or organisms at all.<sup>9</sup> If no physical difference-maker is forthcoming to account for sharp differences in mentality between human persons and human animals, then it is best to withhold mentality from human animals altogether.

The implications of this brand of constitutionalism are strange indeed. If human animals are atom-for-atom physical duplicates of human persons and yet are completely devoid of conscious experience, it follows that for each human person there exists a zombie that constitutes that person.

Olson recognizes that his argument fails to implicate Baker's brand of constitutionalism, precisely because Baker (Baker 2000, 103) maintains that human animals exhibit at least some conscious, mental properties in both a derivative and nonderivative sense. In particular, she allows for human animals to be nonderivatively sentient, and derivatively



possess a first-person perspective. But in so far as Baker's view entails that the human animals that constitute human persons really do exhibit conscious experience, her view raises a difficulty to which we now turn.

### 23.2.3 *The duplication problem and the problem of too many thinkers*

Informally, the *duplication argument* against constitutionalism runs as follows. According to constitutionalism, there is a human animal that is located precisely where you (the person) are that is physically indistinguishable from you in every respect, your atom-for-atom duplicate. As a result, for every physical activity you undertake, the animal undertakes the very same activity. If you stub your toe and experience pain, the animal that constitutes you likewise stubs its toe and experiences the very same instance of pain; if, in holding your newly born child in your arms, you experience a deep sense of joy and wonder, the animal likewise experiences the very same sense of joy and wonder, and holds the very same child in its arms. But many find these overcrowding implications of constitutionalism deeply problematic, even absurd, and consider them a serious mark against the view.

One particularly thorny version of the *duplication problem* is known as the *too many thinkers* argument and can be formulated against constitutionalism as follows:

- 1 If an animal constitutes me, and the animal is a thinker and I am a thinker, then there are two thinkers occupying the region where I am located.
- 2 It is false that there are two thinkers occupying the region where I am located.
- 3 I am a thinker, and the animal that constitutes me is a thinker.
- 4 Therefore, an animal does not constitute me.

The argument trades on the implausibility of there being more than one conscious thinker in any region of spacetime where a human person is located. As we noted in Section 23.2.2, many prominent defenders of constitutionalism reject (3) and deny that human animals have mental lives at all.<sup>10</sup> For this brand of constitutionalism, we might once again press the following grounding worry as above: if persons and animals are exact atom-for-atom (as well as neuron-for-neuron) physical duplicates, then what grounds the fact that each exhibits such a radically different mental profile?<sup>11</sup>

As Baker affirms (3), her own response to the *too many thinkers* argument is to argue that (1) begs the question against her brand of constitutionalism. How so? Part and parcel of Baker's constitutionalism is a form of numerical sameness without strict identity. This point is often underappreciated in extant criticisms of Baker's response to the problem of *too many thinkers* (Olson 2001b; 2016, 4; Zimmerman 2004). More specifically, Baker rejects the following standard analysis of number concepts and counting in terms of strict, absolute identity:

(Counting<sub>Identity</sub>) There is exactly one F =<sub>def.</sub>  $\exists x (Fx \ \& \ (y) (Fy \text{ iff } y=x))$

In words: there is exactly one F just in case there exists something,  $x$ , where  $x$  is F and any  $y$  is F if and only if  $y$  is strictly identical with  $x$ . On the standard view, we count the number of Fs by strict, absolute identity; according to Counting<sub>Identity</sub>, if  $x$  is F and  $y$  is F and  $x$  is strictly nonidentical to  $y$ , then there are at least two Fs.

Baker's brand of constitutionalism rejects Counting<sub>Identity</sub> and opts instead for an analysis of number concepts and counting in terms of strict identity-or-constitution. That is:

(Counting<sub>Identity-or-Constitution</sub>): There is exactly one  $F =_{\text{def.}} \exists x (Fx \ \& \ (y) (Fy \text{ iff } y = x \text{ or } y \text{ is constitutionally related to } x))$

In words: there is exactly one  $F$  just in case there exists something,  $x$ , where  $x$  is  $F$  and any  $y$  is  $F$  if and only if  $y$  is either strictly identical to  $x$  or is constitutionally related to  $x$ . Baker argues that if we gloss number concepts and counting in terms of Counting<sub>Identity-or-Constitution</sub>, then the problem of *too many thinkers* is a nonstarter as there is exactly one thinker occupying the region where I am located. While the animal and I are strictly nonidentical, we are to be counted as *one and the same* thinker in virtue of the fact that we are constitutionally related. If Counting<sub>Identity-or-Constitution</sub> is adopted, then there is no duplication of thinkers and the problem dissipates.

Baker seems wholly unmoved by the counterintuitiveness of the notion of two strictly nonidentical objects being counted as numerically one and the same object. Where some who rely on numerical sameness without strict identity in their metaphysical theorizing are quick to recognize that it cuts against widespread and well-entrenched intuitions about sameness and counting, Baker is undeterred by the incredulous stares (see Brower and Rea 2005; Brower 2014).

Why, then, does Baker hold fast in spite of the incredulous stares? As far as I can tell, Baker's motivation for wedding her version of constitutionalism to the above revisionary analysis of number and counting is primarily twofold, neither tenet of which is tenable in my estimation.

First, Baker advertises the relation of constitution as one of deep unity that falls short of strict identity. In her own words, "Although you are a person and your body is a person, there are not two persons where you are. This is so because constitution is a unity relation. If  $x$  constitutes  $y$  at  $t$ , and  $x$  is an  $F$  at  $t$  derivatively and  $y$  is an  $F$  at  $t$  nonderivatively – or vice versa – then there are not thereby two  $F$ s" (Baker 2007, 38). But note that deep unity without strict identity does not *in and of itself* require the revision of well-entrenched intuitions about number and counting by strict identity. The relation of constitution stands among a family of what E. J. Lowe (2006, ch. 3) calls "formal ontological relations" such as *grounding*, *dependence*, *instantiation*, *composition*, *identity*, *determination*, and *realization*. Composition, for example, is an asymmetric and irreflexive, one-many "building relation" (to use Bennett's 2011 terminology) that generates genuine unity out of plurality, which many take to fall short of strict identity. Contemporary metaphysicians who reject composition-as-identity maintain that the composition relation generates a single, unified whole out of the many parts, without in any way being strictly identical to those parts. While Baker is quick to distance herself from a mereological conception of the constitution relation (in particular many of the formal axioms of classical mereology), the point stands that the desire to secure deep unity without strict identity is not a sufficient motivation *in its own right* for overthrowing standard counting practices in terms of strict identity.

Second, Baker takes her particular brand of constitutionalism, with Counting<sub>Identity-or-Constitution</sub> at its core, to have certain theoretical virtues with respect to (a) solving longstanding puzzles in material objects, and (b) rival accounts of the human person such as animalism and substance dualism. I have already noted that I think Baker is to be commended for her overall nonreductive approach to ordinary objects and persons. But of course it is well beyond

the scope of this chapter to offer a full-scale cost-benefit analysis of Baker's constitutionalism with respect to (a) and (b). Any full-scale critique of Baker's view must weigh such theoretical virtues against the costs and benefits of alternative solutions to such puzzles, in addition to rival views regarding the metaphysics of human persons.

In my estimation, the virtues of constitutionalism (of which there are many) are outweighed by the high price of departing with what many consider to be other more deeply entrenched philosophical views about the world.<sup>12</sup> These include: (i) an analysis of number concepts and counting solely in terms of strict identity (Counting<sub>Identity</sub>); (ii) that ordinary, strictly nonidentical material objects cannot be in the same place at the same time; (iii) that nothing that is not strictly identical to me thinks my thoughts (see below); and (iv) that I do not borrow any aspect of my mental life from any thinker that is not strictly identical to me (see Section 23.2.4). A good many philosophers have taken the intuitive plausibility of (i)–(iv) as evidence against any philosophical model of the human person that entails their denial. Of course, many very capable philosophers have not done so, Baker being a prime example. Perhaps it is true, as one philosopher aptly put it, that serious metaphysics ultimately proceeds not from the head, but from the gut (Heil 2012, 9).

But for those who identify as orthodox Christians (as Baker herself does), there are arguably additional, *prima facie* theological costs, namely, the departure from a long-standing, historically entrenched view of human persons in explicitly *nonmaterialist* terms, along with certain Christian doctrines that are corollaries to such a view (Crisp 2009, 153; van Inwagen 1995, 487). These include the existence of a disembodied intermediate state between death and the final resurrection of the saints (grounded in Pauline and non-Pauline Scriptural passages, for example, 2 Cor. 5:1–10; Matt. 22:23–33), as well as a creedally orthodox Christology which consists of Christ's having a "rational soul" in accord with a straightforward reading of The Council of Chalcedon (451 CE).<sup>13</sup> Christians who remain committed to these two historic positions will find additional theological motivation to reject Baker's brand of constitutionalism.

Even still, the underlying intuition driving the original *Too Many Thinkers* argument has not been completely put to rest by Baker's use of Counting<sub>Identity-or-Constitution</sub>. The fact remains that on Baker's view there exists an object that is strictly not identical to me that thinks my thoughts and shares my mental life. So instead of (1) of the *too many thinkers* argument we might offer the following revised (nonquestion begging) premise:

- 1\* If an animal constitutes me, and the animal is a thinker and I am a thinker, then in the region that I occupy there is a thinker that is strictly not identical to me that thinks my thoughts.

We can then go on to amend the argument by revising (2) to:

- 2\* It is false that in the region that I occupy there is a thinker that is strictly not identical to me that thinks my thoughts.

The argument can then be run the same as before to the conclusion that an animal does not constitute me. Unlike (1), (1\*) makes no claim about the *number* of thinkers in the region I occupy, and thus begs no questions against Baker's view. And many will find (2\*) just as compelling as (2) of the original *too many thinkers* argument. If so, then despite Baker's

protestation to the contrary, her brand of constitutionalism remains saddled with a version of the *too many thinkers* argument.

### 23.2.4 Constitutionalism and the priority principle

Constitutionalism runs afoul of an independently plausible principle regarding the relationship between persons and their mental lives, what Andrew Bailey (2015) has recently called the *priority principle*:

*Priority principle:* We human persons possess all our mental properties in the primary and nonderivative sense. We think each of our thoughts in the primary and nonderivative sense.<sup>14</sup>

By my lights, the principal motivating factor behind the *priority principle* is that human persons are the ultimate source or originator of their mental lives. While Baker argues that some mental properties are had by persons in the primary and nonderivative sense, namely, a first-person perspective, she maintains that persons possess certain mental properties (e.g., *being in pain*) in a derivative or secondary sense, that is, in virtue of being constitutionally related to animals that possess such states in a nonderivative sense (Baker 2000, 100–101). So, when I stub my toe and find myself in a state of pain, I possess the mental property of *being in pain* in virtue of being constitutionally related to a (non-identical) human animal that is itself in pain in a nonderivative sense. On Baker's view, while I really do possess the property of *being in pain*, I am in pain *in virtue of* the animal's being in pain, and not conversely (ignore whether this entails that pain is extrinsic). As a result, Baker's view cuts against the *priority principle*.

But why think the *priority principle* true? For starters, the principle is independently plausible and is arguably a "down-home observation" (Baker 2002, 31) regarding the mental lives of persons. Intuitively, it is facts about me alone that ground facts about the specific contour and qualitative character of my mental life. It would be odd if a person provided an explanatory account of their mental life – *why* the person is in pain, say – by appealing to the mental life of some object that is not strictly them. The notion of borrowing or possessing a mental property in a secondary or derivative sense cuts against the intuitive idea that we are the ultimate source or originator of our mentality.

In addition to its independent plausibility, Bailey offers the following quick argument in favor of the *priority principle*. If I think some of my thoughts in a derivative sense, and doing so is a genuine way for me to think my thoughts, then there are two things thinking my thoughts in so far as the animal that constitutes me thinks some of my thoughts in a nonderivative sense, for example, *being in pain*; but the notion that there are two things currently thinking my thoughts is absurd; hence, I think all of my thoughts in the primary or nonderivative sense (*priority principle*). Yet, in light of our previous discussion (Section 23.2.3), to wield this particular argument for the *priority principle* against Baker would be to beg the question against her view in that it assumes a standard gloss on number and counting (solely) in terms of strict identity. Thus Bailey's argument in this context neglects Baker's use of a revisionary number-concept and counting principle that allows her to outright reject the claim that there are *two* things thinking my thoughts; the person and the animal are numerically the same thinker in so far as they are constitutionally related.

Be that as it may, the *priority principle* itself strikes many as a well-grounded datum regarding the relationship between persons and their mental lives. Baker will no doubt rejoin that the *priority principle* itself begs the question in that it is part and parcel of her view that not all of a person's mental properties are had in the primary or nonderivative sense. Yet for those who remain perplexed by the notion of borrowing a portion of one's mental life from something that is not strictly identical to *oneself* (even if we are to count them as one and the same object), the fact that Baker's constitutionalism runs roughshod over the *priority principle* is a steep cost for the view.

### 23.2.5 Constitutionalism and the first-person perspective

As noted above in Section 23.1, at the heart of Baker's brand of constitutionalism is the notion of the first-person perspective that sets human persons apart from animals and organisms. In her more recent work on the first-person perspective, Baker (2013) glosses the first-person perspective as a complex, irreducibly dispositional property that consists of two distinct stages, the *rudimentary* and the *robust* stage, where each stage is accompanied by the exemplification of a distinct dispositional property.

For Baker, the rudimentary stage of the first-person perspective is nothing more than the exemplification of the disposition or "capacity to interact consciously and intentionally with one's environment" (Baker 2013, 173). Both higher-order nonhuman animals as well as human infants exemplify this particular dispositional property in so far as they have the power to interact with their environment in such a way. By contrast, the robust stage of the first-person perspective consists in the exemplification of the dispositional property of *being self-conscious* or the "capacity to conceive of oneself as oneself" in the first person" (Baker 2013, 174).<sup>15</sup> As persons alone possess self-concepts in general, persons uniquely exemplify the robust stage of the first-person perspective. For Baker, a human person is any being that essentially exemplifies the disposition corresponding to either the rudimentary or robust stage of a first-person perspective (Baker 2013, 40).

There are, however, several notable objections to Baker's dispositional ontology of the first-person perspective.<sup>16</sup> First, while Baker contends that the exemplification of a first-person perspective (essentially) is the sole distinguishing mark of human personhood, she also adamantly maintains the deep ontological divide between the personhood of human infants and the nonpersonhood of higher-order nonhuman animals, despite the fact that both exemplify a rudimentary first-person perspective (essentially).<sup>17</sup> In fact, the ontological uniqueness of human persons is one of the two key desiderata that Baker employs for an explanatorily adequate account of human persons (Baker 2007, 87). Baker herself states the worry here as follows,

If having a first-person perspective is what distinguishes a person from everything else, and if a human infant and a chimpanzee both have rudimentary first-person perspectives, how can a human infant be a person if a chimpanzee fails to be a person? (Baker 2007, 79)

Baker goes on to answer her own query,

What distinguishes the human infant from the chimpanzee is that the human infant's rudimentary first-person perspective is developmentally *preliminary* to having a robust first-person perspective, but a chimpanzee's rudimentary first-person perspective is not preliminary to anything further. (Baker 2007, 79)

On Baker's view, then, the following are necessary and sufficient conditions for the constitution of a human person:

- (HP)  $x$  constitutes a human person at  $t$  iff (i)  $x$  is a human animal (nonderivatively) and (ii)  $x$  has either a robust first-person perspective at  $t$ , or a rudimentary first-person perspective at  $t$  that is *developmentally preliminary* to a robust first-person perspective.<sup>18</sup>

By a human infant's having a rudimentary first-person perspective that is "developmentally preliminary" to a robust first-person perspective, Baker intends to underscore the unique dispositional properties exemplified by the human infant that "developmentally ground or underpin" the development of the robust first-person perspective, properties not possessed by nonhuman animals. In Baker's own words, "a being with a rudimentary first-person perspective is a person *only if it is of a kind that normally develops robust first-person perspectives*" (Baker 2007, 79; emphasis in original). On her view, severely mentally impaired individuals are persons nevertheless because they are the sorts of beings – members of the kind *human animal* – that are naturally disposed to develop a robust first-person perspective, in contrast to higher-order nonhuman animals. According to Baker, then, the deep ontological divide between the personhood of human infants and the nonpersonhood of higher-order nonhuman animals amounts to a fundamental ontological difference in kind-membership.

Baker's appeal to kind-membership as marking the fundamental ontological divide between persons and nonpersons is important in light of her response to what she calls a "regress" or "slippery slope" argument to HP (Baker 2007, 80). As Baker herself points out, the following worry arises for those who aim to secure the ontological uniqueness of human persons by way of HP: "Once we consider a being with a preliminary to a robust first-person perspective to be a person, why not consider a being with a preliminary to that preliminary also to be a person?" (Baker 2007, 80). At bottom, the worry here is that once the appeal to kinds or sortals has been made as the principal way of carving the deepest metaphysical divide between human persons and higher-order nonhuman animals, there remains no principled reason to exclude from the class of human persons beings that belong to the same kind as human infants, *human animal*, yet fail to (occurently) exemplify a first-person perspective.

The worry threatens to undermine HP as a way of securing the ontological uniqueness of human persons. To illustrate, consider the example of a human embryo after implantation (about fourteen days after fertilization).<sup>19</sup> By Baker's lights, the human embryo in this stage of development is not a person in any sense in so far as it lacks a first-person perspective. Nevertheless, as a member of the kind *human animal*, the human embryo is a genetically distinct, albeit radically underdeveloped human organism at this particular developmental stage. As an organism belonging to the kind *human animal*, the human embryo is the kind of organism that exemplifies the requisite dispositions that are developmentally preliminary to both rudimentary and robust first-person perspectives; given a suitable environment, nutrition, and unless hindered by some extrinsic condition (preventers, masks, etc.), the human embryo is precisely the *sort* of organism (in contrast to other living organisms) that is naturally *disposed* to develop the capacity for a first-person perspective, whether rudimentary or robust. Consequently, both the human embryo and the human infant alike are the *sorts* or *kinds* of beings that have the requisite dispositions that developmentally underpin a robust first-person perspective. If so, then according to HP there are no

principled grounds for including human infants in the class of human persons while excluding human embryos after implantation from that same class.

As noted, Baker foresees this objection to HP and formulates the objector's proposed revision to HP as:

- (HP\*)  $x$  constitutes a human person at  $t$  if and only if  $x$  is a human animal (non-derivatively) and either (i)  $x$  has a robust first-person perspective or (ii)  $x$  has capacities that, in the normal course of development, produce a being with a robust first-person perspective.<sup>20</sup>

Baker's chief argument against HP\* relies on the distinction between an "in-hand" and a "remote" capacity, which tracks the notion of a *first-order* and a *higher-order* capacity (i.e., a capacity to develop a capacity), respectively. A hammer, she notes, "has an in-hand capacity at  $t$  for driving nails whether or not it is actually driving nails; you have an in-hand capacity at  $t$  for digesting food whether or not you are actually digesting food. Unassembled hammer parts (a wooden handle and a metal head) have only a remote capacity at  $t$  for driving nails; an embryo has only a remote capacity at  $t$  for digesting food" (Baker 2007, 80).

With the distinction between first-order and higher-order capacities in place, Baker claims that the first-person perspective (whether rudimentary or robust) is a first-order capacity; a subject's exemplifying a first-person perspective "awaits nothing for its exercise other than a subject's thinking a certain kind of thought" (Baker 2007, 80). Along the lines of HP, then, personhood depends exclusively on the exemplification of first-order capacities, since "remote capacities do not suffice for making *anything* the kind of thing that it is" (Baker 2007, 80). As Baker takes the human embryo just after implantation to be a being with absolutely no first-order capacities (2007, 80) but only capacities to develop capacities, it follows that no human embryo at that stage of development qualifies as a human person.

What are we to make of Baker's rejoinder to HP\*, and by extension her defense of the ontological uniqueness of human persons in terms of HP? For one, note that HP *itself* assumes that the determinants of personhood are not exclusively first-order capacities. In contrast to nonhuman animals, human infants are ontologically unique precisely *because* the kind to which they belong is characterized by the higher-order capacity to develop a robust first-person perspective. The possession of a higher-order capacity in this case is what cuts the deepest ontological divide between the human persons and nonpersons, and explains why a human infant and not a chimpanzee counts as a human person on Baker's view; nonhuman animals like chimpanzees are not the kinds of beings that, by nature, possess the higher-order capacity to develop the first-order capacity for a robust first-person perspective. While human infants do not have a robust first-person perspective in hand as a first-order capacity, they nevertheless are still persons in so far as they have the higher-order capacity to develop and exercise a first-order capacity for self-conscious awareness.

In summary, the following dilemma applies to Baker's preferred way of securing the ontological uniqueness of persons in terms of HP. Either the above regress or slippery slope argument against HP is sound or it is unsound. If (a) the argument is sound, then HP is false and either (i) HP\* is true and human embryos after implantation are human persons (which Baker denies), or (ii) Baker rejects HP\* and is left without a principled way to secure the ontological uniqueness of persons, and thereby unable to satisfy one of her key desideratum for an explanatorily adequate account of human persons. If (b) the regress argument is unsound, precisely because first-order or in-hand capacities are exclusively determinative of

personhood as Baker argues, then HP is undermined in so far as it relies on the fact that the having of at least one higher-order capacity, in this case having (in virtue of belonging to the kind *human animal*) the higher-order capacity to develop and exercise the first-order capacity of a robust first-person perspective, is at least partially determinative of personhood. On either (a) or (b), then, HP is undermined and Baker is without a principled way to secure the ontological uniqueness of persons in contrast to the rest of the natural world.

Consider a second objection to Baker's conception of the first-person perspective, one that stems from the use of first-person indexicals. Recall that Baker maintains that human animals possess a first-person perspective *derivatively* in virtue of constituting human persons who possess a first-person perspective *nonderderivatively*. Recall further that for Baker the notion of having a property in a derivative sense is a *genuine* form of property-exemplification, and thus amounts to more than the having of a property "by courtesy." By extension, to have a first-person perspective derivatively is to *genuinely* or *really* have a first-person perspective.<sup>21</sup>

Originally introduced by Paul Snowdon (1990, 83–107), I take the following version of Snowdon's argument to cut against Baker's claim that derivative property-exemplification is genuine form of property-exemplification (and thus a genuine form of having a first-person perspective). For *reductio*, assume constitutionalism – that human persons are not strictly identical to human animals – and call the animal that constitutes you "A."<sup>22</sup>

- 1 If constitutionalism is true, then on occasion O when you, a human person, utter the indexical, "I am strictly identical to an animal," that remark is false.
- 2 Animals – and hence A – have acquired the capacity to use first-person indexicals and thus genuinely possess a first-person perspective.
- 3 If (2), then indexical remarks made through the mouth of A are remarks in which A is the subject.
- 4 Therefore, indexical remarks made through the mouth of A are remarks in which A is the subject.
- 5 If (4), then the indexical remark on occasion O made through the mouth of A (i.e., "I am strictly identical to an animal") is true.
- 6 Therefore, you are strictly identical to an animal; constitutionalism is false (1 and 5).

In response, Baker argues that (4) and the consequent of (3) are ambiguous and are false if read along the lines of:

- (a) Indexical remarks made through the mouth of A are remarks in which A is the subject *nonderderivatively*.

If (2) is true and one adopts reading (a), then the consequent of (3) is false and (4) does not follow and the *reductio* fails. This is precisely because on Baker's view the animal (A) is the subject of indexicals only in a derivative sense, that is, in virtue of constituting a human person who is the subject of indexicals in a nonderivative sense. On the other hand, if (2) is false, then the *reductio* is straightforwardly unsound; either way, says Baker, the *reductio* fails on reading (a).

Alternatively, Baker argues that if (4) and the consequent of (3) are read along the following lines of

- (b) Indexical remarks made through the mouth of A are remarks in which A is the subject *derivatively*.



then (5) is false and the *reductio* once again fails. While (4) and the consequent of (3) may be true if read along the lines of (b), the utterance “I am strictly identical to an animal” in the mouth of the animal is false, precisely because the indexical “I” always refers nonderivatively to the human person. In her own words:

When “I am identical to an animal” issues from the mouth of the animal that constitutes me, I refer to myself (the person) nonderivatively, and say of myself that I am identical to an animal . . . There are not two referents of “I” – any more than there are two persons or two animals – where I am. So, when “I am identical to an animal” issues from an animal’s mouth, there are not two statements – a true one by the animal and a false one by the person. There is only one statement (a false one) made nonderivatively by the person. (Baker 2001, 172)

Baker’s response to the argument here brings to light a peculiar feature of her brand of constitutionalism, one that calls into question whether derivative property-exemplification is *real* property-exemplification and not merely exemplification *by courtesy*. If the animal (A) is the genuine subject of indexicals and the bearer of a first-person perspective albeit in a derivative sense (as per (4) and the consequent of (3) along the lines of (b)), then why is the indexical remark “I am strictly identical to an animal” false in the mouth of the animal? If none of the animal’s first-person remarks or thoughts suffice to generate remarks or thoughts with the animal as the subject or referent, that is, remarks with the content “I (the animal) am strictly identical to a human animal,” then it is difficult to see how the animal really does possess a first-person perspective in the first place.

On Baker’s view, even indexicals that issue from the mouth of the animal have the human person as the subject or referent. While the animal is said to genuinely possess a first-person perspective, the animal is never strictly the subject or referent of first-person indexicals, even when uttered by the animal. The fact that Baker takes (5) to be false on reading (b) strongly suggests that the animal is not *really* a subject of indexicals and a *genuine* bearer of a first-person perspective, even though it is true on Baker’s view that the animal has these *derivatively* in virtue of constituting a human person.

Consequently, if Baker reads (4) and the consequent of (3) along the lines of (b), then her rejection of (5) as a way to block the *reductio* threatens to undermine derivative property-exemplification as *genuine* or *real* property-exemplification. If the animal *really* is the subject of indexicals and the *genuine* bearer of a first-person perspective, then (5) seems unassailable and the *reductio* against constitutionalism sound.

## Notes

1. See Baker (2007, 49) for her distinction between a *rudimentary* and a *robust* first-person perspective, the possession of either (to be exact: a rudimentary first-person perspective that is *developmentally preliminary* to a robust first-person perspective) is both necessary and sufficient for a human animal to constitute a human person. I unpack this distinction more in Section 23.2.4.
2. As recently defended by Schaffer (2010), “priority monism” is the view that the maximal whole – the fusion of all concrete reality, the cosmos – is metaphysically fundamental and prior to each of its proper parts. All ordinary concrete objects such as persons, tables, and trees are derivative and thus posterior to the cosmos. In classical terms, the cosmos is the sole substance.
3. For criticisms of Baker’s formal notion of constitution see the contributions by Rea and Zimmerman in Baker *et al.* (2002), as well as Jedwab (2013).

4. For other important objections to Baker's brand of constitutionalism in particular see Moreland (2009, 130–137) and Zimmerman (2002).
5. There is a straightforward, albeit controversial, response to the grounding problem available to the constitutionalist, yet one that Baker outright rejects given her "quasi-naturalist" (Baker 2007, 87) desideratum that persons are *wholly* material. The controversial response is this: while persons and animals share all of their material parts in common, they might nevertheless differ with respect to some *immaterial* part (see Koslicki 2008). If persons, say, have immaterial, rational souls as proper parts, then while they might share all of their material parts in common with animals, they nevertheless differ from animals with respect to their *total* compositional base. Yet given her quasi-naturalist desideratum, that is, persons are *wholly* part of the natural world, Baker will have no truck with immaterial souls, or even immaterial parts of persons; any adequate account of human persons ought to construe them as material through and through (2007, 68).
6. A standard four-dimensionalist move here, see Sider (2001, 114), is to argue that *Statue* and *Lumpl* stand in distinct modal counterpart relations. However, Baker is no four-dimensionalist, and she emphatically rejects modal counterpart theory (2007, 210–213).
7. Two contemporary proponents of cladism include Okasha (2002, 200) and Sober (1993, 148).
8. For more objections to cladism and purely relational accounts of biological species see Goodwin and Webster (1996, 44–47) and Devitt (2008).
9. Olson cites Johnston (2007, 55), Lowe (1996, 1; 2010), Shoemaker (2008) as proponents of constitutionalism that explicitly deny mentality to the bodies that constitute human persons (but see Hawthorne and McGonigal (2008) for the view that bodies are psychologically indistinguishable from persons).
10. See note 8.
11. The objection here seems particularly problematic for materialist brands of constitutionalism: if persons are *wholly* physical, and persons and animals are atom-for-atom and neuron-for-neuron duplicates, then what grounds the radically different mental profiles between the two?
12. A quick note on (a). If Baker's principal way of solving puzzles in material objects (e.g., the problem of material constitution) is to put to use Counting<sub>Identity-or-Constitution</sub>, then what she considers to be a theoretical benefit of her view will be deemed by many to be a steep theoretical cost.
13. While constitutionalists like Baker (2004, 336) and Corcoran (2006, 140) argue that Scriptural passages that teach the reality of an intermediate state are *logically consistent* with a *bodily* intermediate state (Baker uses the phrase "for all we know" suggesting that such a reading is epistemically possible), it is no accident that such texts have historically been straightforwardly interpreted as, in fact, teaching a *disembodied* intermediate state (see *Westminster Confession of Faith*, ch. 32 and *The Heidelberg Catechism*, answer 57). While deference to Christian tradition on theological matters is defeasible in my opinion, I see no overriding reason to depart from the longstanding, traditional interpretation of these passages.
14. This is a slightly altered version of Bailey's principle. As originally stated, his principle is consistent with Baker's constitutionalism in so far as she maintains that some of our mental properties are had in the primary and nonderivative sense. As such, the original principle lacks an exclusivity clause specifying that *all* of our thoughts are had in the primary and nonderivative sense.
15. Following Gareth Matthews (1992), Baker employs terms accompanied by an asterisk such as "I\*" or "oneself\*" to signal that the user of such terms has a robust first-person perspective.
16. For further objections to Baker's notion of the first-person perspective see Moreland (2009).
17. Baker underscores the fact that "human infants are persons and higher nonhuman mammals are not persons (or probably not)" (Baker 2007, 79).
18. This is adapted from Baker (2007, 79).
19. Baker (2007, 75) is clear that she rejects the view that the embryo is a human organism prior to implantation. While I disagree that the embryo prior to implantation is not a human organism, I grant Baker's view for the sake of argument in what follows.
20. This is slightly adapted from Baker (2007, 80).
21. Along these lines, Baker (2007, 175) remarks: "I take (G) – If *x* has a property *F* at *t*, then *x* has *F* nonderivatively or *x* has *F* derivatively – to be a substantive axiom. So, it is simply not the case that to have a property derivatively is not to have it at all. If you take the constitution relation seriously as a unity relation, then to have a property derivatively is not just to have it by courtesy."
22. This particular formulation of the argument has been adapted from Baker (2000, 201–202).

## References

- Bailey, Andrew. 2015. "The Priority Principle." *Journal of the American Philosophical Association*, 1(1): 163–174.
- Baker, Lynne Rudder. 2000. *Persons and Bodies: A Constitution View*. Cambridge: Cambridge University Press.
- Baker, Lynne Rudder. 2001. "Materialism with a Human Face." In *Soul, Body, and Survival: Essays on the Metaphysics of Human Persons*, edited by Kevin Corcoran, 159–180. Ithaca, NY: Cornell University Press.
- Baker, Lynne Rudder. 2002. "Making Things Up: Constitution and its Critics." *Philosophical Topics*, 30(1): 31–51.
- Baker, Lynne Rudder. 2004. "Christians Should Reject Mind-Body Dualism." In *Contemporary Debates in Philosophy of Religion*, edited by Michael L. Peterson and Raymond J. Van Arragon, 327–338. Malden, MA: Blackwell.
- Baker, Lynne Rudder. 2007. *The Metaphysics of Everyday Life*. New York: Cambridge University Press.
- Baker, Lynne Rudder. 2013. *Naturalism and the First-Person Perspective*. New York: Oxford University Press.
- Baker, Lynne Rudder, Dean Zimmerman, Michael Rea, and Derk Pereboom. 2002. "Book Symposium on *Persons and Bodies*." *Philosophy and Phenomenological Research*, 64: 592–635.
- Bennett, Karen. 2011. "Construction Area (No Hard Hat Required)." *Philosophical Studies*, 154: 79–104.
- Brower, Jeffrey. 2014. *Aquinas's Ontology of the Material World: Change, Hylomorphism, & Material Objects*. Oxford: Oxford University Press.
- Brower, Jeffrey, and Michael Rea. 2005. "Material Constitution and the Trinity." *Faith and Philosophy*, 22: 57–76.
- Corcoran, Kevin. 2006. *Rethinking Human Nature: A Christian Materialist Alternative to the Soul*. Grand Rapids, MI: Baker Academic.
- Crisp, Oliver. 2009. *God Incarnate*. New York: T&T Clark International.
- Davidson, Donald. 1987. "Knowing One's Own Mind." In *Proceedings and Addresses of the American Philosophical Association*, 61: 441–458.
- Devitt, Michael. 2008. "Resurrecting Biological Essentialism." *Philosophy of Science*, 75: 344–382.
- Gibbard, Alan. 1975. "Contingent Identity." *Journal of Philosophical Logic*, 4: 187–221.
- Goodwin, Brian, and Gerry Webster. 1996. *Form and Transformation: Generative and Relational Principles in Biology*. Cambridge: Cambridge University Press.
- Hawthorne, John, and Andrew McGonigal. 2008. "The Many Minds Account of Vagueness." *Philosophical Studies*, 138: 435–440.
- Heil, John. 2012. *The Universe as We Find It*. Oxford: Oxford University Press.
- Jedwab, Joseph. 2013. "A Critique of Baker's Constitution View." *Metaphysica*, 14(1): 47–62.
- Johnston, Mark. 2007. "'Human Beings' Revisited: My Body is Not an Animal." In *Oxford Studies in Metaphysics*, edited by Dean Zimmerman, vol. 3, 33–74. Oxford: Oxford University Press.
- Koslicki, Katherin. 2008. *The Structure of Objects*. New York: Oxford University Press.
- Lowe, E. J. 1996. *Subjects of Experience*. Cambridge: Cambridge University Press.
- Lowe, E. J. 2006. *Four-Category Ontology*. New York: Oxford University Press.
- Lowe, E. J. 2010. "Substance Dualism: A Non-Cartesian Approach." In *The Waning of Materialism*, edited by Robert Koons and George Bealer, 439–461. New York: Oxford University Press.
- Matthews, Gareth B. 1992. *Thought's Ego in Augustine and Descartes*. Ithaca, NY: Cornell University Press.
- Moreland, J. P. 2009. *The Recalcitrant Imago Dei: Human Persons and the Failure of Naturalism*. London: SCM Press.
- Oderberg, David. 2007. *Real Essentialism*. New York: Routledge.

- Okasha, Samir. 2002. "Darwinian Metaphysics: Species and the Question of Essentialism." *Synthese*, 131: 191–213.
- Olson, Eric. 2001a. "Material Coincidence and the Indiscernibility Problem." *The Philosophical Quarterly*, 51(204): 337–355.
- Olson, Eric. 2001b. "Review of Persons and Bodies." *Mind*, 110: 427–430.
- Olson, Eric. 2016. "Zombies Among Us." *Noûs*, 50(2): 1–11.
- Schaffer, Jonathan. 2010. "Monism: The Priority of the Whole." *Philosophical Review*, 119(1): 31–76.
- Shoemaker, Sydney. 2008. "Persons, Animals, and Identity." *Synthese*, 163: 313–324.
- Sider, Theodore. 2001. *Four-Dimensionalism: An Ontology of Persistence and Time*. Oxford: Clarendon Press.
- Snowdon, Paul. 1990. "Persons, Animals, and Ourselves." In *The Person and the Human Mind*, edited by Christopher Gill, 83–107. Oxford: Clarendon Press.
- Sober, Elliot. 1993. *Philosophy of Biology*. Oxford: Oxford University Press.
- van Inwagen, Peter. 1995. "Dualism and Materialism: Athens and Jerusalem?" *Faith and Philosophy*, 12(4): 475–488.
- Zimmerman, Dean. 1995. "Theories of Masses and Problems of Constitution." *Philosophical Review*, 104: 53–110.
- Zimmerman, Dean. 2002. "The Constitution of Persons by Bodies: A Critique of Lynne Rudder Baker's Theory of Material Constitution." *Philosophical Topics*, 30(1): 295–338.
- Zimmerman, Dean. 2004. "Christians Should Affirm Mind-Body Dualism." In *Contemporary Debates in Philosophy of Religion*, edited by Michael L. Peterson and Raymond J. Van Arragon, 315–327. Malden, MA: Blackwell.

# Debating Emergent Individualism

# For Emergent Individualism

TIMOTHY O'CONNOR

*Persons* are knowing subjects who intentionally act. More carefully, persons are those individuals who have or have a natural potential for the capacities of subjective awareness, intrinsic intentionality and cognition, and intentional action. In this chapter, I want to consider persons primarily through their capacity for intentional action, and more specifically still through the freedom of will or choice that we commonly suppose mature, intact human persons to manifest. The philosophical problem of free will is in large measure the problem of understanding how causally conditioned, physically composed agents can be loci of *ultimate* control: how can human agents freely originate, and not merely conduct, causal influence upon the world? That our acts be causally undetermined is necessary for us to exert such ultimate control – or so say I – but it is not nearly enough. Again, truly free agents would be originators, not merely conductors, of probabilistic causal influence. Elementary particles would be instances of such originators, if they turn out to be basic (uncomposed) ingredients of the physical world. If not, then this will be true of whatever supplants them from that status. (For myself, I can scarcely credit the notion entertained by some metaphysicians that *nothing* would have that status: causality and so-called “atomless gunk” cannot coexist if we maintain, as I do, that causation is an ontologically basic relation.)<sup>1</sup>

Maybe certain varieties of elementary particle are mereologically basic, but it appears that human persons are not. It appears that we are physically composed – *pace* mind-body dualism on which we are simple, purely mental entities that are causally, but not mereologically, bound up with our bodies. I don't regard this as just *obvious*. The conscious mind is a very striking and puzzling affair, enough so as to make mind-body dualism an option to be seriously explored, even in 2018. However, for what it's worth, my own judgment is that the schematic philosophical “theory” of minded human persons that best accounts for relevant natural-historical, organismic-developmental, neurophysiological, *and* introspective evidence has it that we are wholly physically composed.

Suppose that this thesis of human material composition is correct. Does it *follow* that we are not originators of causal influence upon the world – that our influence at any given moment is instead merely the “sum” of the influences of our basic constituents, themselves exerting nonrational forces? I think the answer to this question is No. There is *conceptual* space for fundamental causal influence exerted by nonbasic entities. What is more, it seems to me to be a wholly open question empirically whether this might *actually* be the case concerning ourselves.

My central purpose here is to show how causally conditioned and physically composed entities (as I take us to be) may nonetheless be fundamental, nonderivative causes. My point of departure will be recent work in metaphysics trying to articulate a notion of metaphysical fundamentality and its associated notion of grounding, whereby non-fundamental entities are grounded in fundamental entities. (An advantage of focusing on the notion of grounding is that it permits one to avoid the matter of “reducibility” that has become quite complicated and vexed in recent discussion. Doubtless the same will be true ten years from now for a proliferated range of notions of grounding.) This will suggest a notion of emergence on which emergent entities are *fundamental* though *nonbasic* – causally sustained by, but not grounded in, entities that are fundamental *and* basic.

A full articulation of emergence in terms of the nonbasic but fundamental requires us to take a stand on the ontological categories of individual (substance) and property. I will state with only minimal defense my favored variety of substratum-attribute theory. I will then apply the theory to the thesis that human persons are composite, emergent individuals.

## 24.1 “Ontological” Emergence as Fundamental-but-Nonbasicality

I am a full-bore ontological realist. I believe that there is an objective, wholly nonconventional and nonperspectival way the world is. Within this realist perspective, it is natural to interpret the analytic sciences as pointing to a kind of *priority* structure: macroscopic objects typically derive their existence and features from, are grounded in, their proper parts, and ultimately from a privileged set of basic objects that are not so grounded in anything and that belong to kinds that are ubiquitous.

(An aside: suppose that classical theism is true, and the totality of physical reality is purposed and causally sustained at all times by an immaterial Creator. Would this entail that *no* physical reality is ontologically basic? We have to be careful here. Setting aside Berkeleyan idealism, on which physical objects ultimately consist in ideas in the mind of God, classical theism entails that physical reality is, in its being, wholly other than God. God is a sufficient *cause* for the existence of physical entities. What is more, as God is a metaphysically necessary being, it follows that physical entities *supervene* of metaphysical necessity on God’s willing them. They are beings that are *essentially* causally dependent on God. Even so, they are not, in the intended sense, ontologically derivative from or grounded in God, as the dependency is causal, not constitutive. The moral here is that we cannot capture the basic/derivative ontological grounding structure in purely formal terms.)

So we have a schematic picture of physical reality on which there are ontologically basic physical entities that fall into a small set of kinds, at least some of which are instanced wherever any nonbasic physical kinds are instanced. What the defining features of these

kinds are is of course an open question. They may be particle-like, field-like, or of some further kind bandied about within the speculative reaches of current physics. They may be momentary or enduring. And so on. (My adherence to *scientific* realism has its limits, getting most tenuous when contemplating the more *outré* reaches of theoretical physics, but I try to carry it as far as I can.)

I have used the term “basic” to pick out those objects that have no objects as parts. I will also say that the perfectly natural, intrinsic *properties* of such basic objects, and the perfectly natural relations into which they enter, are likewise basic. Basic objects, properties, and relations are not grounded in anything. Are *only* basic entities ungrounded? Let us use the term “fundamental” to characterize that which is ungrounded, that which is ontologically primitive. Our question then is whether there can be nonbasic but fundamental objects, properties, and relations.

Let us start by considering intrinsic natural properties. If such a property were nonbasic but fundamental, it would be (1) had by a nonbasic object, that is, an object that itself has parts and (2) ungrounded, or “ontologically primitive.” How we understand the notion of a *property’s* being ungrounded will depend on how we think of properties generally. Some contend that fundamental natural properties are “*qualities*” or characters that make for objective, intrinsic similarity of objects – and that nonredundantly, that is, in a way that is not “already” fully captured by the objects’ having other properties, or by the objects’ parts having certain properties and standing in certain relations. A second kind of property theorist thinks of natural properties as “*pure powers*.” For her, a fundamental natural property is a nonredundant causal power. A still third kind of theorist in effect brings these two conceptions together, taking properties to be “*powerful qualities*.”<sup>2</sup> I am myself tempted by this last conception. Insofar as I understand it, such a theorist will say that we can capture the idea of a fundamental natural property in each of the first two ways, and the verdict that they deliver will necessarily coincide.

All of these recently popular ways of trying to get at the notion of property-fundamentality depend on thinking of natural properties as *immanent* to objects. And immanence, in turn, leads to property *constituency*, whether of universals or tropes. But since John Heil (2012) and E. J. Lowe (2012) demur on this last point, let me elaborate. In their terminology, there are “modes” (or “ways” that objects are) that are immanent to but not constituents of objects. They argue that, because these modes are inherently dependent entities, not capable of independent existence, they cannot be constituents. This reason is unconvincing. They allow that modes are “aspects” of objects, particular, perceivable, located where the objects are, and partly responsible for objects having causal powers of various kinds. I do not see how anything could fill these theoretical roles without being (proper) *constituents* of the objects. Granted, if we also contend, plausibly, that these property-like aspects of things are inherently *dependent* entities, not capable of existence apart from the objects they characterize, then the way they enter into the makeup of an object importantly differs from the way basic *objects* enter into the makeup of a composite object, enough so that we should sharply separate property constituency from part-whole mereology. But the notion of constituency is not analytically connected to that of real separability.<sup>3</sup>

In any case, I myself do think of natural properties as constituents of things. Given this constituent ontology, we have perhaps another, more straightforward way of characterizing fundamental intrinsic properties: they are those properties the having of which does not consist in the having of any other properties, or in the having of other properties by one’s parts and any of the natural relations among them.



Understood in any of these ways, it seems to me perfectly intelligible that some of the properties of composite objects might yet be fundamental. Note that fundamentality is consistent with the having of such a property's being causally dependent in some way on something else, such as the states and arrangements of the object's parts being some more or less particular way. And that is precisely how I want to think of an *emergent property*: a fundamental natural property of a composite object that is causally generated and sustained (at least in part) by the object's basic parts in virtue of their own intrinsic properties and configuration.<sup>4</sup> Like all fundamental natural properties, it is a powerful quality whose possession enables an object to make a nonredundant causal contribution to its context.

Nonbasic fundamentality for objects, or what you might think of as emergent individuality, is less straightforward because of the contested, obscure issue of whether even basic material objects have ontological structure and its connection to other fraught matters such as the nature of time and persistence. Here I am just going to chart my own way through these thickets, setting off philosophical tempests in different ones of my readers at various points. I already noted that I favor an ontology on which natural properties are constituents of objects. I will write as if such properties are universals, but everything of importance that I will claim in what follows could (I think) be stated, with minimal change in words, in terms of tropes instead.

Consider then basic objects – those objects that have no (object) *parts*. According to the constituent ontologist, they nonetheless have *constituents* that are properties. How should we understand the constituency-structure of these objects? We have two broad options: we may say that they are (mere) bundles of properties, unified by a primitive relation of some kind. But bundles of universals, it seems, do not an individual object make.<sup>5</sup> We might simply add into the mix primitive thisness, a property – only one that is nonqualitative, noncausal, and nonshareable. I prefer instead to reject bundle theory in favor of a substratum-attribute structure. Individual basic objects consist in there being a substratum (aka bare particular, thin particular) being “tied” in a primitive way to a cluster of natural properties, the object's powerful qualities.

It is often said that the substratum-attribute theory of objects is incoherent. One might object: “The theory says that substrata ‘in themselves’ have no properties, and nothing can be like *that*. (And isn't the very assertion that a substrata lacks properties self-refuting?)” This objection rests upon confusion, though the critics have been abetted in some cases by proponents of the theory, in particular David Armstrong (1997). Yes, we can truly *say* some things about substrata. (If we couldn't, we wouldn't have a theory.) We can say, for example, that substrata are inherently incomplete, particular entities, that they necessarily join in a primitive way with natural properties, and that these natural properties belong to a distinct category from that of substrata. We don't say that these statements are true in virtue of substrata being joined “in themselves” to a special class of universals. Instead, the theory posits them as primitive truths concerning the theory's primitive categories – truths that identify their theoretical role.<sup>6</sup> (We do not appeal to property-constituents in a misguided attempt to reductively analyze all predication; their theoretical work is elsewhere.)<sup>7</sup> Indeed, the claim that substrata are “tied to” constituent universals may be taken as such a primitive truth.<sup>8</sup>

Since the objection is persistently raised, let me dwell on it a bit further. Consider an electron that we might call “Eleonore.” Eleonore is negatively charged; equivalently, Eleonore has the property of (unit) negative charge. From the standpoint of metaphysics, if not of physics, that is a pretheoretical truth, a true predication. Furthermore, it is plausible

that the *concepts* of “property” and of “an object’s having of a property” are basic, part of our conceptual toolkit that does not admit further analysis. Nonetheless, according to the substratum theory, not *every* property term picks out a metaphysically basic kind of entity (at least not of an immanent sort) and *having a property* in the pretheoretical sense is not a metaphysically basic relation. That is why, to avoid confusion, it is best to sharply segregate the terms of ordinary discourse and those of the theory when it comes to what grounds the truth of ordinary claims. Lewis’s “natural property,” which I have invoked, is a half-step in this direction. It marks a theoretical role that specific theories such as the substratum-immanent universals theories fill out in particular ways.

Pretheoretic claim: Eleonore the electron has the property of unit negative charge. A theoretical account of what the truth of this claim consists in: Eleonore has a substratum and several immanent universals as interdependent (“tied”) constituents, and among the latter is the immanent universal of negative charge. As to what to say concerning this substratum itself, the theory bids us to apply a few sortal concepts: for example, particular, inherently incomplete, and “taking” immanent universals, thereby constituting an individual substance. And it also maintains that these predicates are true of substrata without being grounded in further immanent universals. (You might worry: but if we accept primitively true predications in some cases, why not go all the way and accept a form of nominalism? Answer: All analysis must end somewhere, but the theory stops in a better place than the nominalist does. Unlike nominalism, the theory can give an informative account of the many varieties of objective qualitative similarity and causal/dispositional similarity among objects, and this advantage is especially telling when we note that such similarity can cut across substance kinds – for example, both electrons and protons have mass. We can thus reduce the number and kind of unexplained similarities we must embrace. The only “similarities” among bare particulars are categorial. Furthermore, the stopping point is not arbitrary: given the theory’s machinery, we cannot go further with ontological analysis, since universals cannot be ingredients of bare particulars.)

We might go on to ask, is Eleonore’s substratum, like Eleonore, negatively charged? Critics (e.g., Bailey 2012) sense a dilemma here, but it is spurious. We could, if we *liked*, say that the substratum, too, is negatively charged. If we do so, we should say that there are two ways to have a property: by having a substratum that is tied to a suitable underlying universal and by being such a substratum. This is not a distinction that marks anything in the theory, as there is just one metaphysical truthmaker here. But partly for this very reason, I see no reason why we should want to say this. *Eleonore is negatively charged* may be a pretheoretical truth, but *Eleonore’s substratum is negatively charged* is not. Eleonore’s substratum is a creature of metaphysical theory, and we should look to the theory alone to tell us what we should say about it. So, I do not say that Eleonore’s substratum is charged, or has mass, or any of the things physical theory leads us to say about Eleonore.

One final point, before I move on to talk about composite objects and emergence. Critics see it as a weakness in the substratum-attribute theory that the sortal concepts we may apply to its posited substrata are so few and so informationally “thin.” But the natural reply is that its simplicity/thinness appropriately matches its job description: mere individuation.<sup>9</sup>

So, a basic object is a structured entity involving a substratum and a small number of immanent universals (perhaps certain quantities of mass, negative charge, spin, and magnetic moment, as the case may be). Given the essential interdependency of the basic object’s constituents, universals, substrata, *and* the basic objects they enter into are *all* fundamental entities.<sup>10</sup> (More precisely, as Stephan Leuenberger has pointed out to me,

here we need to introduce the notion of *partial* grounding: basic objects are partly but not wholly grounded in their constituent universals and substrata.)

A composite object has mereological structure: it has basic objects as its parts. What I'll call a "*garden-variety*" composite object does *not* have universals as constituents; instead, each of *its* intrinsic properties is grounded in the monadic universals had by, and the relations among, its parts. It also lacks a substratum: there is nothing more to the being of such an object, at any given moment, than the mereological sum of its object-parts and their interrelations.

By contrast, a composite object that has fundamental natural properties functions as a true unity, doing fundamental causal work additional to the activity of its parts. What should we conclude from this? Here, I must confess to being uncertain, as I now see a real difficulty that I had not previously appreciated in defending a view somewhat along the lines of the above (O'Connor and Jacobs 2003, 2010).

I am tempted to suppose that such an emergent composite has a proprietary substratum, one that persists just so long as some or other of its nonbasic but fundamental properties do. But this requires supposing a special, fundamental category of substrata, encompassing those that "attach themselves" to objects – those objects that are the composite's parts. That's a significant difference in ontological profile from substrata for *basic* objects, which attach only to universals; but given the framework of the theory, it cannot be explained by the presence of distinctive properties, as other fundamental causal-functional differences are.

Suppose, then, that we reject the notion of a proprietary substratum being tied to a collection of objects constituting the system giving rise to the emergent properties. Now consider the intended application to human persons: knowing subjects and agents that are biological organisms, and thus entities constituted by an ever-changing collection of parts. The problem here is that the (seeming) unity of a *subject* of experience and purposive *originator* of action seems to go ungrounded. There is an emergent experiential state, but only an organized collection of objects that, individually experience-less, are collectively *the* subject of the experience. This is of doubtful coherence.

Insofar as both options seem unsatisfactory, substance dualism beckons: persons not just as fundamental but as basic (albeit causally dependent as emergents).

Surely there is a misstep in one or the other of the horns of this dilemma, but at present I cannot see where.

## Notes

1. Well, maybe there is one scenario where this might work: imagine emergent fundamental properties (in the sense discussed below) at every "level" of structure in a bottomless nesting. I take this to be certainly nonactual, but Rob Koons, who suggested it to me, is not so sure!
2. For a defender of the first option, see Armstrong (1978); for the second, see Molnar (2003); and for the third, see Heil and Martin (1999), Jacobs (2011), and Heil (2012).
3. For helpful discussion on this point, I thank Anthony Fisher.
4. Elizabeth Barnes (2012) suggests an alternative notion of emergence as that which is fundamental and dependent. I indicated above that I think there is no one "right" way to think of emergence, since different notions may have applications to different organized phenomena. That said, I think that Barnes's proposal needs some tightening up if it is to be theoretically useful. If there are fundamental entities that are interdependent, they will come out as emergent on her account, and this seems the wrong result for most any theoretical purpose. Emergent properties are quite generally conceived as properties of complex systems. We might also worry that Barnes's account precludes the possibility that, for example, some mental properties are

emergent in us (because we are dependent, nonbasic entities) but not for (possible) other subjects who are fundamental.

5. Laurie Paul (2017) is willing to countenance “primitive individuation” for bundles of universals, such that there might be two or more bundles of the very same universals that are “brutely” different. That there be ungrounded brute individuation of basic entities is unobjectionable, maybe even inevitable. That this be so for *bundles* of such entities, by contrast, is bizarre. Such a position belies Paul’s claim to rely only on a “well understood” notion of the part–whole relation.

If we embrace an account of immanent natural properties as particulars (tropes), then of course there is no problem of *particularity*. However, it remains implausible that *substantiality* can result from bundling of inherently dependent entities, as property-particulars are plausibly taken to be. That is, if there is an immanent instance of unit negative charge belonging to a particular electron, we should not think of it (*pace* Paul, once again) as a kind of thing or individual substance, not even on the way to being a thing (Ayer’s “junior substance”). It is essentially a characteristic of that electron, one way that it is. Supposing that the bundling of several such nonindividuals could result in an individual is to embrace a kind of metaphysical alchemy.

6. For more on this, see Sider (2006).
7. See Lewis (1983).
8. Hence, I reject David Armstrong’s (1997, 2004) “truthmaker” argument for states of affairs. But even if we posited “tie” as a fundamental ontological relation, this would not lead to incoherence, as again, the theory is not seeking to analyze all predication in terms of immanent universals.
9. But the worry may be pressed as follows: lacking thick natures, substrata cannot have essential properties. But then it should be possible that the substratum of an electron has been the substratum of a quark instead. In other words, substrata wouldn’t seem to have anything to do with the essential natures of the things that they particularize. And that just seems bizarre. (Something along these lines is pressed by Loux 1998.) Though I will not develop the response here, I suggest that this sort of problem might naturally be solved by positing that an object’s substratum and essential attributes constitute an indissoluble basic unit, somewhat akin to Peter Simons’s (1994) “nuclear” bundle theory.
10. Since the constituency relation has different formal characteristics than the part–whole relation, we need to maintain their distinctness, and composite objects have both kinds. Laurie Paul (2017 and earlier writings) takes it to be an advantage of her bundle-of-universals view that it makes do with mereology alone. As I see it, the parsimony is achieved by brute force, in her declaration that universals are perfectly “thingy” in their own right.

## References

- Armstrong, David. 1978. *A Theory of Universals: Universals and Scientific Realism*, vol. 2. Cambridge: Cambridge University Press.
- Armstrong, David. 1997. *A World of States of Affairs*. Cambridge: Cambridge University Press.
- Armstrong, David. 2004. *Truth and Truthmakers*. Cambridge: Cambridge University Press.
- Bailey, Andrew. 2012. “No Bare Particulars.” *Philosophical Studies*, 158: 31–41.
- Barnes, Elizabeth. 2012. “Emergence and Fundamentality.” *Mind*, 121: 873–901.
- Heil, John. 2012. *The Universe as We Find It*. New York: Oxford University Press.
- Heil, John, and C. B. Martin. 1999. “The Ontological Turn.” *Midwest Studies in Philosophy*, 23: 34–60.
- Jacobs, Jon. 2011. “Powerful Qualities, Not Pure Powers.” *The Monist*, 94(1): 81–102.
- Lewis, David. 1983. “New Work for a Theory of Universals.” *Australasian Journal of Philosophy*, 61: 343–377.
- Loux, Michael. 1998. “Beyond Substrata and Bundles.” In *Contemporary Readings in the Foundations of Metaphysics*, edited by S. Laurence and C. Macdonald, 233–246. Oxford: Blackwell.
- Lowe, E. J. 2012. “A Neo-Aristotelian Substance Ontology: Neither Relational nor Constituent.” In *Contemporary Aristotelian Metaphysics*, edited by T. Tahko, 229–248. Cambridge: Cambridge University Press.
- Molnar, George. 2003. *Powers: A Study in Metaphysics*. Oxford: Oxford University Press.
- O’Connor, Timothy, and Jonathan Jacobs. 2003. “Emergent Individuals.” *Philosophical Quarterly*, 53: 540–555.

- O'Connor, Timothy, and Jonathan Jacobs. 2010. "Emergent Individuals and the Resurrection." *European Journal for Philosophy of Religion*, 2: 69–88.
- Paul, L. A. 2017. "A One-Category Ontology." In *Being, Freedom, and Method: Themes from the Philosophy of Peter van Inwagen*, edited by J. Keller, 32–61. Oxford: Oxford University Press.
- Sider, Ted. 2006. "Bare Particulars." *Philosophical Perspectives*, 20: 387–397.
- Simons, Peter. 1994. "Particulars in Particular Clothing: Three Trope Theories of Substance." *Philosophy and Phenomenological Research*, 54 (3): 553–575.

# Against Emergent Individualism

ROBERT C. KOONS

## 25.1 Introduction

In a series of at least ten books and articles over the last twenty-two years, Timothy O'Connor and his collaborators have developed one of the most rigorous, subtle, and influential accounts of the relation between mind and body, which for present purposes we can call "emergent individualism." My own work has been shaped and enriched by this body of work. Consequently, the critique I offer here is a decidedly friendly one, intended to advance our understanding of the mind while building on the contributions of O'Connor and his coauthors (Wong, Churchill, Theiner, and Jacobs).

In recent years, I have been working on the articulation and defense of a version of the hylomorphism of Aristotle and Thomas Aquinas (see Koons 2014). There is much common ground between such a Thomistic version of hylomorphism and emergent individualism. Both theories include a rejection of physicalism, in both its reductive and nonreductive versions, based on physicalism's failure to account adequately for qualia, intentionality, normativity, and mental causation. Both embrace an incompatibilist version of free will and both adopt the model of agent causation (in fact, hylomorphists would extend this model to cover all causal interactions, treating the early modern model of event causation as capturing a derivative level of metaphysical reality). Both count mental causation as real and irreducible. Both incorporate causal powers as a fundamental element of ontology, and both endorse a sparse ontology of properties. Both reject nominalism and conceptualism, and include some form of realism about properties, either in the form of immanent universals (that are literally parts of their instances) or trope-like abstract particulars (individual forms).

Where, then, do the differences lie? This question will be taken up in Section 25.2 of this chapter, in which I will locate each of four antireductionist positions on a conceptual map, namely: nonreductive physicalism, Cartesian dualism, Thomistic hylomorphism, and emergent individualism. In Section 25.3, Section 25.4, and Section 25.5, I will argue for the superiority of hylomorphism over emergent individualism on each of three issues: the nature of the causes of the existence of persons (Section 25.3), the possibility of disembodied

personal survival (Section 25.4), and the nature of the influence of mind on body (Section 25.5). I make some concluding remarks in Section 25.6.

## 25.2 A Conceptual Map for the Philosophy of Mind

To begin with, let's exclude the two most extreme views: reductive physicalism (in which only the physical is real) and idealism (in which only the mental is real). That leaves four moderate positions: nonreductive physicalism, Thomistic hylomorphism, emergent individualism, and Cartesian dualism.<sup>1</sup>

How do reductive and nonreductive versions of physicalism differ? We can distinguish the two very simply, if our background theory includes a relation of metaphysical grounding (see Fine 1994, 2012; Schaffer 2009; Rosen 2010) and a sparse ontology of properties and states of affairs. For reductive physicalists, there are nonphysical concepts with nonempty extensions, and true propositions with nonphysical content, but the only properties with instances and the only actual states of affairs (or facts) are entirely physical in nature. Nonreductive physicalists, in contrast, are committed to the real existence of nonphysical properties and facts. They count as physicalists because they hold that all nonphysical facts are wholly grounded in the physical facts alone. Consequently, the class of the truth-values of nonphysical propositions strongly supervenes on the class of the truth-values of the propositions of physics.

The other three positions deny physicalism altogether by denying that the nonphysical facts are wholly grounded by the physical facts. For nonphysicalists, there are fundamental nonphysical properties and facts (including mental properties and facts). This does not necessarily entail a denial of mental-on-physical supervenience, since supervenience is a necessary but not a sufficient condition for physicalism. However, most anti-physicalists (including, I think, all Cartesian dualists) do in fact deny even the weak, global supervenience of the mental on the physical.

It is much harder to maintain the strong, localized supervenience of the mental on the physical, in the absence of the complete grounding of the mental by the physical. For this reason, strong, localized supervenience is (as far as I know) a materially adequate definition of nonreductive physicalism, although I think the definition in terms of grounding does a better job of getting to the heart of the matter.

How can we distinguish Cartesian dualism, Thomistic hylomorphism, and emergent individualism from one another? There are two relatively superficial tests that seem to do an adequate job of sorting anti-physicalist theories into one of the three bins. First, is it possible for a human being (or another entity with mental properties) to begin to exist without having any physical properties (in an immaterial or matterless condition)? If *Yes*, then we have a version of Cartesian dualism. If *No*, then either Thomistic hylomorphism or emergent individualism. Second, is it possible for a human being to reach a condition of immateriality? If *Yes*, then Thomistic hylomorphism. If *No*, then emergent individualism.<sup>2</sup>

Although these questions do give us three mutually exclusive categories that are jointly exhaustive of nonidealistic anti-physicalism, they are not very illuminating about what reasons can be given for these three sets of answers. We have a better chance of gaining such illumination if we look at relations of ontological dependency between the mind and the body. Metaphysical grounding is one species of ontological dependency, but it is not the only species of this genus. It is possible for one metaphysically fundamental (ungrounded)

entity to be ontologically dependent on another (Fine 1994). For example, if we accept origins essentialism, each organism is ontologically dependent on the prior existence of its parents, but that does not mean that the child's existence (now) is grounded by the parents' existence (then). Here's another example: we might think that extended things are ontologically dependent on the existence of space without supposing that the existence of the extended thing is partly grounded by the existence of space.

Ontological dependence can be either synchronic or diachronic. If *A* is synchronically dependent on *B*, then *A*'s existence at each moment *t* depends on *B*'s existence at that moment. Diachronic dependence is weaker: if *A*'s existence is diachronically dependent on *B*, then the existence of *A* at each moment *t* depends on *B*'s existence at some time *t*\* (typically a time no later than *t*).

Can there be synchronic ontological dependency without grounding? I think so. I suppose that metaphysical grounding is a necessitating relation: when fact *F* wholly grounds fact *G*, it is impossible for *F* to exist without *G*'s existing. In contrast, ontological dependency runs in the opposite direction, modally speaking: if object *O* is ontologically dependent on object *P*, it is impossible for *O* to exist without *P*'s existence. Thus, synchronic ontological dependency of one object *O* on *P* is incompatible with the complete grounding of the fact of *P*'s existence by *O*'s existence, even though in each case, the existence of the fact that *O* exists entails the existence of the fact that *P* exists. Metaphysicians have at times spoken as if the impossibility of *F*'s existence without *G*'s existence were a kind of "dependency" of *F* on *G* (even Aristotle spoke this way about "priority"<sup>3</sup>), but this sort of modal "dependency" should be sharply distinguished from true, metaphysical dependency, which is an asymmetric relation between entities, and not merely a fact about covariation across worlds.

We can now ask: Are human beings (and other mental-property bearing entities) in a relation of synchronic ontological dependence to physical things? If the answer is *Yes*, then we have either nonreductive physicalism or emergent individualism. If the answer is *No*, then we have either Thomistichylomorphism or Cartesian dualism. To distinguish hylomorphism from Cartesian dualism, we can ask the follow-up question: Are human bodies in a relation of synchronic ontological dependence to human souls? If *Yes*, then Thomistic hylomorphism. If *No*, then Cartesian dualism.

How then shall we distinguish nonreductive physicalism from emergent individualism? We could try to distinguish them by their answers to the same follow-up question: Is there a synchronic ontological dependency of human bodies on human souls? However, it is likely that both nonreductive physicalists and emergent individualists will answer *No* to this question, on the grounds that there are no entities (on those views) that can reasonably be identified with the term *human souls*. We could try a slightly different form of the question: Is there a synchronic ontological dependency of human bodies on human persons (or individuals)? Emergent individualists should answer *Yes* to this question, on the ground that it is only the emergent human person that supplies the per se unity to the materials that make up the body. Without the emergent human being, the matter that composes the human being would not compose any *one* thing at all, and so nothing that deserves the label of "body" would exist. However, it is not clear to me that the nonreductive physicalists couldn't also give a *Yes* answer for the same reason.

In order to distinguish nonreductive physicalism from emergent individualism, we have to ask a different question, namely: Are there instances of irreducible mental-to-physical causation? The emergent individualists must answer *Yes*, and the nonreductive physicalists



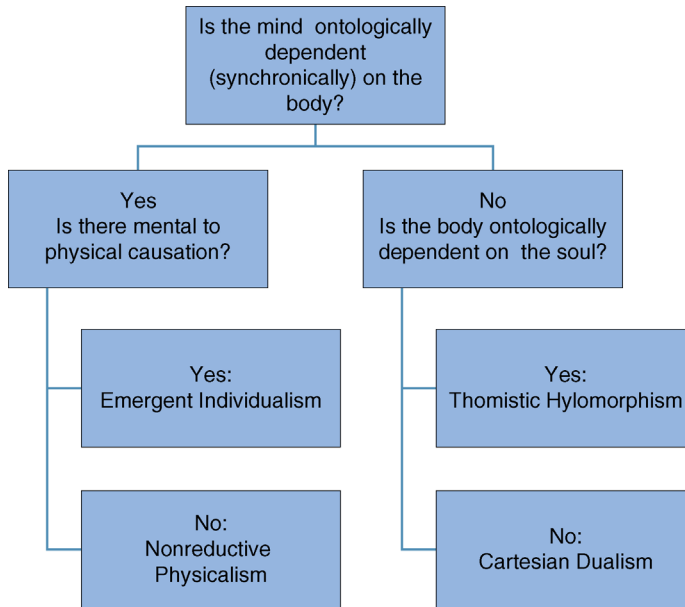
should answer *No*. It is true that some nonreductive physicalists have attempted to answer *Yes* to this question, despite Kim's famous causal exclusion argument, but I agree with O'Connor and Churchill that such attempts ultimately fail (O'Connor and Churchill 2010). This conclusion follows from our shared commitment to both causal powers as fundamental and to a sparse ontology of properties and facts. If mental facts are wholly grounded in the physical facts, then the relation between the mental and the physical is essentially noncausal in nature, and all genuine causal relations tie physical facts and events to other physical facts and events.

What positions do Thomistic hylomorphists and Cartesian dualists take on the question of mental-to-physical causation? It is clear that Cartesian dualists must be interactionists: they must posit direct and fundamental causal ties running from the mind to the body and the body to the mind. Otherwise, they would be forced to embrace idealism (no real causal power in the physical world), epiphenomenalism (no real causal power in the mental world), or Leibnizian pre-established harmony (no real causal power of the mind over the body or *vice versa*). These are not attractive options.

Since both emergent individualists and Cartesian dualists believe in direct and fundamental causation from mental facts to physical facts, and since both think that both the mental and the relevant microphysical facts (i.e., facts about the locations and trajectories of the microparticles) are metaphysically fundamental, both groups of theorists must posit that mental facts can make a real difference to the behavior of physical entities, a difference that cannot be accounted for in terms of the causal powers of those microphysical entities alone, including the powers associated with the four fundamental physical forces (gravitation, electromagnetism, weak and strong nuclear forces). Therefore, they must either posit a fifth fundamental force (a mental or personal force) or posit at least local and temporary violations of mass-energy conservation.<sup>4</sup>

The issue is more complicated when we turn to Thomistic hylomorphists. For hylomorphists, many causal powers of the body and its parts (even its ultimate, microscopic parts) are at least partly grounded in the essential and accidental properties of the soul (or *form*). The formal causation that runs from soul to corporeal organs is a species of metaphysical grounding and not of causation proper (what Aristotelians refer to as *efficient causation*). Thus, the soul acts upon others only indirectly, using corporeal organs as instruments. There need be no direct causation from mind to body. Consequently, there need be neither a fifth fundamental force nor any violation of conservation laws. The nomological completeness of microphysics is no threat to the real and irreducible power of the macroscopic organism, because the hylomorphist no longer supposes the microphysical facts to be fundamental and ungrounded. It is the soul (form) that is responsible (in part) for the relative locations and trajectories of the microphysical parts: the microphysical laws simply take as inputs what is (already) partly grounded in the nature of the whole, living and rational person. Figure 25.1 shows the resulting conceptual map.

I want to make one more attempt at a conceptual map of the territory, this time in terms of metaphysical grounding. Cartesian dualists take both the mind and the body to be complete and independent substances. For such dualists, the facts about the mind are not even partly grounded by the facts of the body, nor are the facts about the body partly grounded by the facts about the mind or soul. In contrast, Thomistic hylomorphists do take many facts about the body (even about its ultimate or simple constituents, if there are any) to be at least partly grounded in facts about the human soul (as the Aristotelian form of the body). Emergent individualists may also take certain mereological facts about the body and



**Figure 25.1** First conceptual map.

its parts (e.g., the fact that these material entities do compose a single thing) to be at least partly grounded in facts about the emergent human individual.

Nonreductive physicalists deny that the body is even partly grounded by the soul, but they insist that the soul is at least partly (and, in fact, wholly) grounded by the body. Thus, we have so far distinguished both Cartesian dualism and nonreductive physicalism from each other and from the remaining two categories, but we have not yet distinguished between Thomistic hylomorphism and emergent individualism.

The difference between Thomistic hylomorphism and emergent individualism seems to lie in the realm of causation. Thomists agree with emergentists in thinking that it is metaphysically impossible for a human being to begin to exist without the synchronic participation of certain physical entities (such as the human ovum). Moreover, it is part of the very essence of human beings that we have such a beginning. However, once we human beings have begun to exist, we are capable of continuing to exist without the cooperation of any physical entity whatsoever. For emergent individualists, in contrast, the existence of a human being depends at each moment on the cooperation of the physical parts that make up the human body, and this causal dependency is itself essential to the persistence (and not just the origination) of a human being.

This scheme would seem to leave us with two possible versions of Cartesian dualism: those who agree with Thomists in thinking that there is an essential causal dependency of the soul on the body at its first moment of existence, and those who deny any such essential causal dependency. In fact, the first sort of Cartesian dualism seems very hard to justify. Hylomorphists have an explanation for the essential causal dependency of the soul on the body at its origin: namely, the thesis that the natural state of the human being is that of a single substance with both mental and physical powers. In rational animals like us, our intellectual powers are essentially dependent on certain of our corporeal powers, namely, our sensory powers. All of our universals and all of our universal knowledge are derived

(by abstraction) from the information received through our senses. For Aristotelians (both ancient and modern), sensory powers are essentially tied to the sensitivities of corporeal sense organs. The corporeal sense organ has among its essential and fundamental passive causal powers systematic sensitivities to the presence of real qualities in the environment. A human being cannot take in such sensory information without a body, and we cannot exercise our intellectual powers without such sensory information.

The physical components of the body do not have these sensory powers, either individually or collectively, except as the living body of an ensouled human being. If a human being were to begin in a disembodied condition, he or she would utterly lack these sensory powers and would lack the capacity to gain them (barring miracles). There would be no soul-less human bodies with the missing powers with which the disembodied human being could be unified. A being in such a condition could not have the human powers of intellect at all and so could not be a human being. Since humanity is essential to us, no human being can begin to exist except as a human being. Thus, human beings cannot begin to exist in a disembodied state. Once a human being has begun to exist and once his or her intellectual powers begin to be exercised in sustained activities of contemplation of universal truth, that human being can persist in existence without the cooperation of either the sense organs or the rest of the body.

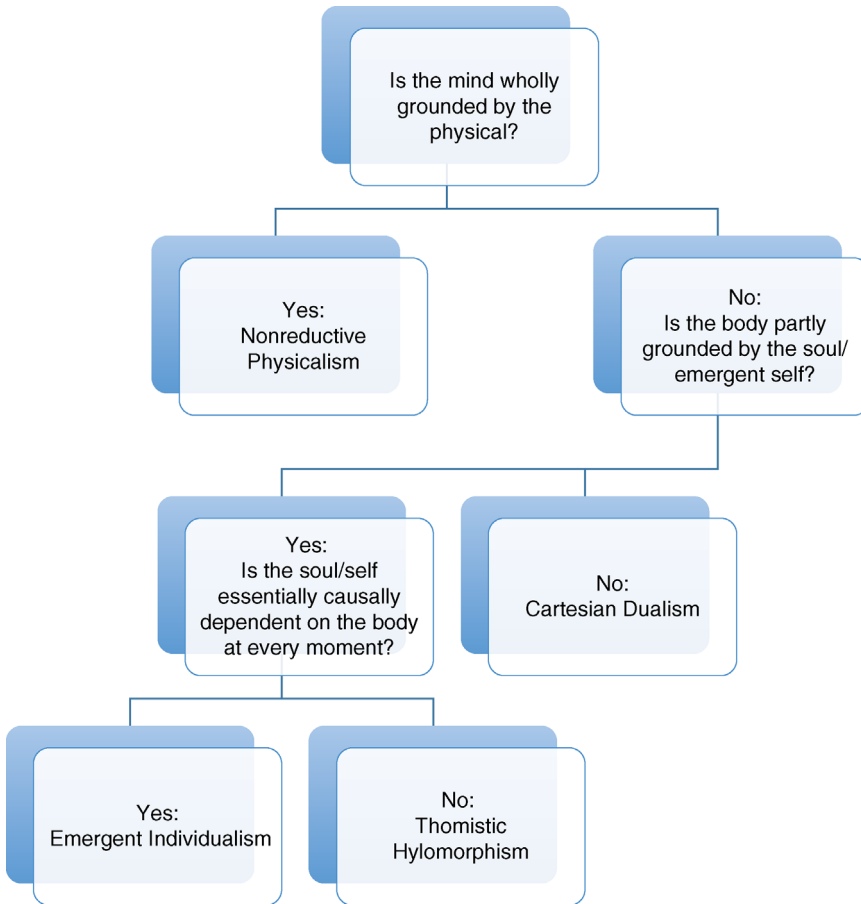
The Cartesian dualist, in contrast, has no such story to tell. If a Cartesian human soul were to begin without a body, it would have all the powers that are essential to being human: it would simply need to be “hooked up” in the right way to a soul-less human body, which would no longer be a metaphysical impossibility. Therefore, the map shown in Figure 25.2 is obtained.

This map provides us with a second way of distinguishing emergent individualism from Thomistic hylomorphism, namely, whether this is an essential causal dependence of the human person on the human body at each moment of the person’s existence. We have seen why the hylomorphist denies the possibility of a disembodied existence of a person at the beginning of his existence, but we have not yet found why the emergent individualist wants to extend this impossibility throughout the person’s life. We must seek a further factor that explains why the emergent individualist wants to make this extension and the Thomist does not.

This further factor would seem to consist in the emergent individualist’s assuming that the physical realm has a certain kind of causal priority over the personal and mental. This subtle priority finds its expression in the term “emergence”: the emergent individualist assumes that the personal or mental *emerges from* the physical, which presupposes that it is the physical (and perhaps only the physical) that has the causal power to produce an instance of personality. Given this picture, it is not surprising that the emergent individualist supposes that there is a causal dependence of the mental on the physical at each moment of a mental or personal thing’s existence. Hylomorphists, in contrast, make no such assumption about the causal priority of the physical. They are more likely to think that the personal or the super-personal has the causal power to generate a person, just as living entities have the power to generate new living things, in which case it would be unsurprising for them to suppose that a thing with intellectual (nonmaterial) powers might have the power to sustain itself in existence, without the cooperation of physical entities.

We have then identified three critical issues that divide emergent individualists from Thomistic hylomorphists:

- 1 Is the power to create and sustain persons possessed by and only by microphysical entities?



**Figure 25.2** Second conceptual map.

- 2 Are disembodied human persons possible?
- 3 How does the mind affect the body: by formal causation (a species of metaphysical grounding) or by direct causal action?

I will take up each of these issues in the following three sections.

### 25.3 Bottom-Up, Top-Down, or On-a-Level: What Causes Human Beings?

Emergent individualists owe much to the British Emergentists, especially Samuel Alexander (1920). Alexander's central theme was the unity of nature, understood diachronically. He embraced an evolutionary picture of the world, with no causal discontinuities in nature. O'Connor and Churchill echo these themes:

It is enough that at every juncture introducing some new kind of causally discontinuous behavior, there is a causal source for that discontinuity in the network of dispositions that

underlie it. In short: unity in the order of the unfolding natural world need not involve causal continuity of behavior, only continuity of dispositional structure. (O'Connor and Churchill 2010, 278)

As a theist (and not a deist), I don't find such a commitment to absolute causal continuity of nature to be obligatory. I wouldn't rule out the occasional occurrence of direct divine interventions – acts of special creation. My openness to such discontinuity is especially clear at two crucial points: the origin of life, and the origin of humanity.

Nonetheless, even if I were committed to natural continuity through time, I wouldn't be persuaded that it is the microparticles that must possess the causal power needed to explain the origin of living organisms and conscious human beings. Why not attribute this power to composite substances that are nonliving? In particular, there are three genuine possibilities: the cosmos, or planets or planetary systems, or complex inorganic systems with holistic chemical and thermodynamical properties.

For emergentists, cosmic history is one of gradual ontological aggregation: to begin with, there were only elementary particles, then (perhaps) substantial atoms, then molecules, then larger systems with various thermodynamic, convective, and/or crystalline structure, and then finally living things (including conscious and free persons). At each stage, the smaller entities exercise at some point their latent powers to combine and form new substances.

There is, however, an alternative story that could be told, one that has prominent metaphysical advantages. This alternative cosmic history is one of gradual disaggregation and splintering. In the beginning, there was a single substance, the cosmos, which eventually broke up into proto-clusters, then galaxies, then stars and planetary systems, then proto-ecological systems with inherent features of a convective and thermal nature, then biotic systems consisting of populations of identical one-celled organisms, and finally individual multicellular organisms. At each stage, existing substances give rise to new substances by division, not aggregation.

The Big Bang model suggests that the world consists of a single substance in the immediate aftermath of the singularity. Although it is true that photons, leptons, and quarks soon appear, it is far from obvious that they constitute complete substances at that point. The phenomena of widespread quantum entanglements, carrying as they do implications of ontological holism or nonseparability, tells in the opposite direction. I call this narrative "de-escalation."

De-escalation has a clear advantage over emergence: it requires at each stage only a single agent of efficient causation, an entity with an active causal power whose exercise results in a multiplicity of new substances. Emergence, in contrast, depends on the collaboration of a large number of independent agents, jointly exercising a set of complementary causal powers. Emergence requires a large-scale conspiracy of mutually agreeing causal powers possessed by the large number of smaller entities that spontaneously join together in forming a new substance. De-escalation, in contrast, involves at most two entities, an agent and patient, the first causing the second to undergo disintegration into a large plurality of new entities.

This is not perhaps a decisive fact, but we should, other things being equal, prefer accounts that avoid brute conspiracies among large numbers of independent agents.

Emergence would require a very improbable and ad hoc preestablished harmony among the powers of the many mutually unifying parts – a coordinated distribution of mutually exercisable powers. This problem ramifies as the number of components to be unified

increases. It becomes quite untenable when billions of components must unite with each other.

De-escalation coheres nicely with hylomorphism, which in turn offers a correspondingly simple account of substantial persistence: hylomorphism locates the source of the persistent unity of each substance in a single agent, the substantial form. The presence of the many material parts serves merely as the patient of the formal action, as enabling conditions for the exercise of the form's formal powers. These enabling conditions are built into the form itself, requiring no prior mutual agreement. In contrast, emergence requires that the same kind of collaborative conspiracy of independent powers needed to bring the composite substance into existence persist throughout the substance's persistence.

How exactly does such de-escalation work? Answering this question would involve some subtle interaction between metaphysics and empirical science. My current proposal is that subatomic particles did not constitute Aristotelian substances in the early history of the universe but did so only much later, as the cosmos cooled and separated. Not all particles separated from the original cosmic substance as distinct substances in their own right: some became instead nonsubstantial, integral parts of other subcosmic substances, such as galaxies and solar systems. Still others eventually ended up as parts of prebiotic proto-ecosystems, and finally as constituting substantial populations of unicellular organisms. The first substantial organisms may have been multicellular in constitution, as discrete and cooperating populations of micro-organisms achieved joint reproduction. At each stage, the substantial form of the larger substance contained within it the potential of generating new substances at a smaller scale.

How does all this speculative natural history relate to our present-day conditions? We now find substances at multiple levels of scale, including perhaps: galaxies and solar systems, ecosystems, unicellular colonies, multicellular organisms, thermal substances, subatomic particles. The substantial forms of composite substances (that is all forms except those of fundamental particles) have the power to take and to expel smaller entities, which exist as substances when separated from the larger composite. If a system is destroyed or suffers amputation, new substances at a smaller scale are created. De-escalation theory differs from emergentism by refusing to locate the ultimate explanation of all these transformations in the forms of the ultimate particles alone.

## 25.4 The Intermediate State: Why not Disembodied Persons?

From a hylomorphic perspective, the persistence of human beings beyond the death of the body is not impossible. If human beings are not caused to begin to exist by the joint action of microparticles, why assume that they are caused to *persist* in existence by such action of microparticles? Why can't the human being persist in existence (and persist in engaging in intellectual activities, like the contemplation of abstract truths) despite the destruction of the body?

Critics of Thomistic Hylomorphism (including O'Connor and Jacobs 2003) offer two principal objections to the disembodied persistence of human persons (the so-called *intermediate* state between death and resurrection): the Cheshire cat objection (forms without matter are impossible in the way that smiles without faces are impossible) and the Dion/Theon objection (the person cannot become identical with one of his own proper parts).

### 25.4.1 *The Cheshire cat objection*

Critics of Thomism often argue that matterless form is as inconceivable as the smile of Lewis Carroll's Cheshire cat, which survives after the rest of the cat has disappeared. The objection assumes that there can't be form without matter to be formed.

We should first note that this is an objection not just to St Thomas but also to Aristotle. Aristotle proposed that the human soul is the form of the body, and yet there exist celestial intelligences (also capable of intellectual activity) that are completely devoid of matter. Defenders of the Cheshire cat objection must suppose that Aristotle was deeply confused about his own notion of form and its relation to the intellect.

The Cheshire cat objection is predicated on the assumption that forms are structures, and that structures are sets of properties or facts, facts about the intrinsic natures of a thing's material parts and about the relations among those parts. A thing without material parts could have no structure, and therefore no form.

However, Aristotelian forms are not structures. They are instead the metaphysical *grounds* of structure. So, it is not impossible for those grounds to exist in the absence of what is grounded. To be precise, forms are *partial grounds* of structure: structure is also partly grounded in the existence of suitable matter. The human soul can exist without being the actual ground of corporeal structure so long as it is still the ground of some activity. In the case of human beings, the soul can be the ground of pure intellectual activity, which does not essentially depend on the existence of a body. A soul that is engaged in such activity can survive the destruction of its body.<sup>5</sup>

### 25.4.2 *The Dion/Theon objection*

How can the human being be composed of both body and soul at one point in time and then be identical to the soul alone at a later point in time? That is, how can a whole become identical to one of its proper parts? This would entail either *relative identity* (the soul is identical to the person *at the later time* but not identical to it *at the earlier time*) or the *denial of the irreflexivity of proper parthood* (the person is always identical to the soul and so is a proper part of itself prior to death), or the *denial of weak supplementation* (the disembodied person after death would have his soul as a proper part, without having any other part that does not overlap with it).

The Stoic philosopher Chrysippus (ca. 280 BCE–ca. 206 BCE) exploited this dilemma (in his Dion–Theon paradox) as a problem for any account in which a substance can lose one of its proper parts. Suppose that the unfortunate Dion loses his left foot to amputation. Let's call the post-amputation person "Theon." If we suppose that Theon is identical to Dion, we face a problem. Consider Dion-Minus: the proper part of Dion (prior to amputation) that includes everything but his left foot. It seems that Dion-Minus is identical to Theon: the two consist of the very same material things arranged in the same way. So, if Dion is identical to Theon, he is (or at least becomes) identical to one of his own proper parts (Dion-Minus). Yet Dion and Dion-Minus are surely distinct entities.

Peter van Inwagen (1981, 123–125) correctly identified the best solution to the Dion–Theon paradox: simply to deny the existence (prior to the amputation) of Dion-Minus, by denying the doctrine of the arbitrary fusion of undetached parts. The parts of Dion compose something (namely, Dion), but proper subsets of those parts (such as the parts of Dion except his left foot) compose nothing whatsoever. Hence, Dion does become Theon without becoming identical to any of his proper parts.

As O'Connor and Jacobs point out (O'Connor and Jacobs 2003), Thomistic hylomorphists face a *prima facie* difficulty in applying van Inwagen's solution: they cannot deny that the soul exists prior to death, and they seem to be committed to the soul's being (prior to death) a proper part of the human being. Therefore, when a human being is reduced at death to a soul without a body, it would become identical to something that had existed as one of its own proper parts.

It is crucial to distinguish between two senses of "part": a broad and a narrow sense. In the narrow sense, the soul is never *part* of the human being (it is not one of the human being's *integral parts*, to use Thomas's language). The soul grounds the existence of the human being at each moment in time, whether the human being is composite (before death) or mereologically simple (after death).

In the broad sense, the soul is *part* of the person at each moment of his or her existence, by way of being one of the metaphysical components of the person, but so too are the person's accidents (including his or her intellectual actions and activities). Before death, the person is constituted by soul, body, and accidents; after death, by soul and accidents alone. By denying the doctrine of arbitrary fusions, the hylomorphist can deny that there is anything constituted by just the soul and accidents prior to death, thus avoiding the Dion-Theon paradox, just as van Inwagen does.

## 25.5 From the Mind to the Body: Formal Causation or Fundamental Force?

Emergent individualists and Thomistic hylomorphists agree that in some sense the body is partly grounded in facts about the whole person. For emergent individualists, it is holistic features of the person (including the whole person's causal interactions with the parts of the body) that provide the body with its *per se* unity. It is by virtue of these emergent facts that the parts of the body compose a single thing. However, O'Connor and his collaborators seem to assume that these mereological or compositional facts about the body are the only ones grounded in the emergent self. Like physicalists, emergent individualists assume that all nonmereological facts about the individual microscopic particles or fields (e.g., facts concerning their causal powers and spatial and spatiotemporal relations) are metaphysically independent and fundamental. In contrast, hylomorphists take all such facts about microscopic parts to be at least partly grounded in holistic facts about the composite substances to which they belong.

Consequently, emergent individualists and hylomorphists have fundamentally different conceptions about the way in which the microphysical parts of the body are affected by the emergent self or Aristotelian form. For emergentists, this influence is primarily causal (in the narrow sense of direct *efficient* causation): the self moves the particles by exerting on them something like a fundamental physical force. For hylomorphists, in contrast, the soul is a *formal* cause, not just of the body as a whole, but also of each of the body's microscopic parts. The soul is the metaphysical ground of the causal powers of the microscopic parts, and the soul acts upon the physical world *indirectly*, through the parts of the body as instruments.

There is some similarity between the hylomorphic conception of the causal role of the soul and the model of strong emergence developed by Carl Gillett (2002, 2003, 2006). In both cases, the microscopic particles and fields have the causal powers they do because of



their inclusion in a whole of a certain kind. In neither case is the whole assigned its own causal power to move its constituent particles. However, there are two key differences between hylomorphism and Gillett's strong emergence. First, Gillett is silent on any relation of metaphysical dependency between the parts and the whole. This leaves open the possibility that the powers of each microscopic part are affected, not by its inclusion in a whole of a certain kind, but simply by the presence in its environment of a large number of other microscopic entities, suitably arranged. Such a possibility is excluded by hylomorphism, since the location and arrangement of the other microparticles is ultimately grounded in the nature of the whole, and not vice versa.

Second, Gillett implicitly excludes the possibility that the whole could instantiate any state or engage in any activity that is not wholly constituted by the arrangement and movement of its constituent particles. Consequently, he does not attribute to the whole any immanent causal powers – any power, that is, to engage in an intrinsic activity that is not strongly and locally supervenient on the successive states and movements of its microparticles. In contrast, Thomistic hylomorphists take the human being to be capable of intellectual activities that are not dependent on any corporeal organ and so need not supervene on the state of the body.

Hylomorphists can easily accommodate the synchronic dependence of much intellectual activity on the state of the brain. Activities such as abstraction, inference, classification, recognition, deliberation, and decision all require the use of internal imaginative representations (St Thomas's *phantasms*), which essentially involve essentially corporeal states. The only human activities that can occur independently of the body are acts of pure contemplation of abstract, wholly general facts and possibilities.<sup>6</sup>

O'Connor and Churchill object that Gillett's model of mind-to-body causation is too limited in its scope:

All we would have embraced are mental properties that play a kind of structuring role in the world's dynamics. They do no distinctive causal work – provide no extra causal oomph. There is, indeed, a strong analogy here to the role played by spatial and temporal relations in Newtonian mechanics, as construed by a causal powers theorist. Such relations, one might say, provide a necessary framework for the interplay of dispositional entities, while themselves having no dispositional nature. Surely our nonreductionist physicalist wants more than this by way of the causal relevance of the mental. More than being local, nondispositional constraints on the way fundamental physical causes operate, our beliefs, desires, and intentions themselves directly contribute to the unfolding dynamics of our behavior. (O'Connor and Churchill 2010, 276)

This overlooks the possibility of action through instrumental intermediaries. Even if the mind cannot act directly on the body, it can act indirectly (through the body) on other physical objects. The whole substance has real causal powers of its own: the causal powers of its parts are metaphysically grounded (at least in part) on those holistic powers, making the action of the microscopic parts merely instrumental in nature (see Koons 2014 for more details). In addition, on the hylomorphic account, the soul has the power to engage in rational activities (resolving itself upon a decision, for example), thereby synchronically altering (via formal causation) the intrinsic features and relations of its microscopic parts.

Won't this alteration of the microscopic parts involve the introduction of new fundamental forces or violations of mass-energy conversion, whether we call this "formal" or

“efficient” causation? No, this would follow only if we assumed that the microparticles have, independently of the exercise of human causal powers, precise locations and trajectories in a common spatiotemporal domain. The quantum revolution of the last one hundred years undermines the Democritean metaphysical assumption shared by physicalists, Cartesian dualists, and emergent individualists.

In the Copenhagen interpretation (developed by Bohr and his collaborators), the microphysical facts consist merely in the attribution to microscopic entities of certain *potentialities*, and these potentialities essentially include causal relations to macroscopic systems. A quantum doesn’t typically have any position or momentum at all (not even a vague or fuzzy one): it has merely the potential to interact with macroscopic systems as if it had some definite position or momentum (or other observable feature) at the moment of the interaction. Thus, the quantum world (so understood) can be neither metaphysically fundamental nor a complete basis for the macroscopic world.

Of course, this situation gives rise immediately to a puzzle: what, then, is the relationship between the macroscopic and quantum worlds? Presumably, macroscopic physical objects are wholly composed of quanta. How, then, can the quanta fail to be a metaphysically fundamental and complete basis for the macroscopic world?

Hylomorphism offers a ready answer to this puzzle. The microscopic constituents of macroscopic objects have (at the level of actuality) only an indirect relation to space and time: they are located (roughly) somewhere at a time only qua constituents of some fundamental, macro- or mesoscopic substance (in the Aristotelian sense). Such microscopic objects are not metaphysically fundamental in their entirety, and their metaphysically fundamental features do not provide a complete basis for the features of the substantial wholes they compose.

Although the Copenhagen interpretation, with its somewhat simplistic dualism of quantum and classical worlds, has fallen out of favor in recent years, Nancy Cartwright has defended a more pluralistic version: the dappled world picture (Cartwright 1999). On this view, the world consists of a variety of domains, each at a different level of scale. Most of these domains are fully classical, consisting of entities with mutually compatible or *commutative* properties. At most one domain is accurately described by quantum mechanics. Since location does not (for quantum objects) “commute” with other observables, like momentum, the quantum objects are only intermittently located in ordinary, three-dimensional space, although they always retain a probability of interacting with classical objects at a definite location. Interaction between quantum properties and classical properties (including those of experimenters and their instruments) precipitates an objective collapse of the quantum object’s wavefunction, as a result of the joint exercise of the relevant causal powers of the object and the instruments, and not because of the involvement of human consciousness and choice.

The main drawback of Cartwright’s model is that it denies the intelligibility of speaking of a cosmic wavefunction embracing all of reality, an approach that has become popular in recent years. Alexander Pruss’s traveling-forms model (Pruss 2014) offers an interpretation of quantum mechanics that is both friendly to hylomorphism and consistent with a cosmic wavefunction. On Pruss’s picture, there is a single quantum wavefunction which describes the state of the whole of microphysical reality and which evolves according to a unified, deterministic law (based on Schrödinger’s equation). However, this quantum realm is not the whole of reality, nor does the macroscopic world supervene upon it.

This quantum wavefunction can be taken as ascribing potential positions to each of the world's quantum particles. Some of the potential positions of some particles are strongly correlated with those of other particles, as a result of the process known as *decoherence*. This decoherence can be thought of as delimiting a very large set of alternative consistent histories of the world's particles. On Pruss's view, just one of these histories has a metaphysically privileged status, forming the basis for the real composition of material bodies, including living organisms. Even though this history is not *microphysically* privileged, acting simply on a par with all other consistent histories in the uniform evolution of the quantum world, it is *ontologically* distinguished by the fact that it, and it alone, corresponds to a world of real composite objects. Pruss in effect uses facts about the "special question of composition" (to use Peter van Inwagen's (1995) phrase) to single out one micro-history as the material basis for a world of macroscopic objects.

Although Pruss's world is microscopically deterministic, the macroscopic world is dynamically indeterministic, since the consistent history that underlies that macroscopic world at one time can later "branch" into several, disjoint histories. The substantial forms of macroscopic objects travel together down just one of those branches, in a way that is not determined at the quantum level, and which may be indeterministic at the macro level as well, although macroscopic agency (including acts of free will) may contribute to determining the direction of "travel."

In neither model (Cartwright's or Pruss's) is the relation between macroscopic actions and microscopic reality one of efficient causation. In both cases, microscopic bodies acquire approximate positions and trajectories by way of metaphysical grounding in irreducible and fundamental macroscopic facts. Such a relation of grounding could even be used to make sense of Bohm's interpretation of quantum mechanics, in which the microscopic world is both complete and deterministic (but radically nonlocal in its interactions). We could take the inseparably and radically holistic four-dimensional world of Bohm's mechanics to be a metaphysically dependent projection of an underlying fundamental reality that consists of macroscopic Aristotelian substances that interact locally and indeterministically. This would be analogous to the way that Kant saved human freedom by taking the deterministic world of Newtonian mechanics to describe a phenomenal realm ultimately grounded in a noumenal realm of freedom. Once again, the relation between macroscopic substances and their microscopic parts would be one of formal and not efficient causation.

## 25.6 Conclusion

Emergent individualists concede too much to the microphysicalist. First, they concede that all of reality is to be explained ultimately in terms of the causal activity of the mereologically fundamental (simple) particles and fields. The generation and persistence of non-microphysical entities (the "emergent" entities) are always to be explained in terms of the causal powers of the microscopic entities. The microscopic domain is causally responsible for the existence and persistence of all exceptions to its causal completeness. The picture is one of a world that was originally exclusively microscopic in character (a world of mereological or compositional nihilism) from which composite substances are generated, without appeal to any agency except that of the simple particles. We have no reason to embrace such a picture, in light of quantum holism. The cosmos was there from the beginning, and it never consisted

of isolated and unrelated particles. Macroscopic substances have been there from the very beginning.

Second, emergent individualists (like Cartesian dualists and physicalists) accept a Democritean assumption about the metaphysical nature of the microphysical domain: namely, that the microphysical facts are metaphysically fundamental and ungrounded. This implies that if nonmicroscopic entities (whether macroscopic or immaterial) are to make a difference they must do so by acting upon microparticles through the exertion of force. In contrast, hylomorphists deny that there is a metaphysically independent microphysical domain in the first place, opening the possibility of formal causation from wholes to parts. Modern quantum theory has altered the imaginative landscape in such a way as to revive the plausibility of the hylomorphic story.

Can these two elements of emergent individualism come apart? It would seem to be impossible to reject the second assumption (the metaphysical fundamentality of the microphysical) while maintaining the first (the ultimate causal sufficiency of the microphysical). It would be impossible for the microphysical to be ultimately responsible for causing the macroscopic domains if microphysical facts are partly grounded by those domains.

What about the other way around? What if we were to affirm the metaphysical fundamentality of the microscopic realm while denying that there is an essential synchronic causal dependency of the mind on the body? There's a serious problem for this combination of views: a version of Jaegwon Kim's pairing problem. If both the microscopic entities and the emergent selves are metaphysically fundamental, then the only relations tying them together are relations of causal dependency. Why is this mind tied to these microparticles? Clearly, the microparticles are not continuously causally dependent on the mind, so it seems that we must suppose that the mind is continuously causally dependent on facts about these particular particles.

The Thomistic hylomorphist, in contrast, has a different and ultimately more satisfying solution to the pairing problem: the microscopic parts are metaphysically dependent on the whole, and the soul is the ground of that metaphysical unity.<sup>7</sup>

## Notes

1. There are other ways of dividing up the logical space, which would generate other positions, including neutral monism, dual-aspect theory, or panpsychism. For present purposes, each of these could be seen either as a variant of one of the four positions (for example, dual-aspect theory seems to be a version of nonreductive materialism) or as introducing orthogonal issues (for example, the question of how many things have minds – a question to which panpsychism provides an answer).
2. It is obvious, I think, that no one will want to defend the position according to which it is possible for a human being to be immaterial at the first moment of its existence but impossible to be immaterial at later moments. So, three categories of anti-physicalism seem sufficient.
3. *Metaphysics* Delta, 1019a.
4. Cartesian dualists and emergentists might argue that quantum collapse phenomena provide an avenue for real downward causation without new fundamental forces or violations of conservation. In fact, I will defend such a position myself, but I argue that such downward influence is best understood as involving, at least in part, the Aristotelian notion of formal causation, rather than being understood in terms of standard event–event causation.
5. Are the souls of human beings who have not yet reached the age of reason not naturally immortal? If so, it would still be possible for God to enable (supernaturally) any such human being to begin to exercise such intellectual powers prematurely and thereby to survive death. Alternatively, it might be that once the human being exists,

- with the natural capacity for abstract thought, the human soul can persist as the enduring ground for that capacity.
6. God can enhance the intellectual activities of disembodied souls by providing them (miraculously) with the required phantasms. However, there would be no subsisting soul so to enhance if there were not something (pure contemplation) that the disembodied soul could do naturally.
  7. I would like to acknowledge the support during the 2014–2015 academic year of the James Madison Program in American Ideals and Institutions at Princeton University (for a visiting fellowship) and the University of Texas at Austin (for a faculty research grant).

## References

- Alexander, Samuel. 1920. *Space, Time, and Deity: The Gifford Lectures 1916–1918*. London: Macmillan.
- Cartwright, Nancy. 1999. *The Dappled World: A Study of the Boundaries of Science*. Cambridge: Cambridge University Press.
- Fine, Kit. 1994. "Ontological Dependence." *Proceedings of the Aristotelian Society*, 95: 269–290.
- Fine, Kit. 2012. "Guide to Ground." In *Metaphysical Grounding: Understanding the Structure of Reality*, edited by Fabrice Correia and Benjamin Schnieder, 37–80. Cambridge: Cambridge University Press.
- Gillett, Carl. 2002. "The Dimensions of Realization: A Critique of the Standard View." *Analysis*, 62: 316–323.
- Gillett, Carl. 2003. "Strong Emergence as a Defense of Non-Reductive Physicalism: A Physicalist Metaphysics for 'Downward' Determination." *Principia*, 6: 83–114.
- Gillett, Carl. 2006. "Samuel Alexander's Emergentism: Or, Higher Causation for Physicalists." *Synthese*, 153: 261–296.
- Koons, Robert C. 2014. "Staunch vs. Faint-Hearted Hylomorphism: Toward an Aristotelian Account of Composition." *Res Philosophica*, 91: 1–27.
- O'Connor, Timothy, and John Ross Churchill. 2010. "Nonreductive Physicalism or Emergent Dualism: The Argument from Mental Causation." In *The Waning of Materialism*, edited by Robert C. Koons and George Bealer, 261–280. Oxford: Oxford University Press.
- O'Connor, Timothy, and Jonathan D. Jacobs. 2003. "Emergent Individuals." *Philosophical Quarterly*, 53(213): 540–555.
- Pruss, Alexander. 2014. "The Traveling Minds Interpretation of Indeterministic Theories." Alexander Pruss's Blog, November 5. Accessed June 22, 2017. <http://alexanderpruss.blogspot.com/2014/11/the-traveling-minds-interpretation-of.html>.
- Rosen, Gideon. 2010. "Metaphysical Dependence: Grounding and Reduction." In *Modality: Metaphysics, Logic and Epistemology*, edited by Bob Hale and Aviv Hoffman, 109–136. New York: Oxford University Press.
- Schaffer, Jonathan. 2009. "On What Grounds What." In *Metametaphysics: New Essays on the Foundations of Ontology*, edited by David Manley, David J. Chalmers, and Ryan Wasserman, 347–384. Oxford: Oxford University Press.
- van Inwagen, Peter. 1981. "The Doctrine of Arbitrary Undetached Parts." *Pacific Philosophical Quarterly*, 62: 123–137.
- van Inwagen, Peter. 1995. *Material Beings*. Ithaca, NY: Cornell University Press.

## Further reading

- Jacobs, Jonathan D., and Timothy O'Connor. 2010. "Emergent Individuals and the Resurrection." *European Journal for Philosophy of Religion*, 2(2): 69–88.
- O'Connor, Timothy. 1994. "Emergent Properties." *American Philosophical Quarterly*, 31(2): 91–104.

- O'Connor, Timothy. 2000. "Causality, Mind, and Free Will." *Noûs*, 34(s14): 105–117.
- O'Connor, Timothy. 2000. *Persons and Causes: The Metaphysics of Free Will*. New York: Oxford University Press.
- O'Connor, Timothy. 2009. "Conscious Willing and the Emerging Sciences of Brain and Behavior." In *Downward Causation and the Neurobiology of Free Will*, edited by Nancey Murphy, George Ellis, and Timothy O'Connor, 173–186. Berlin: Springer.
- O'Connor, Timothy, and Georg Theiner. 2010. "Emergence and the Metaphysics of Group Cognition." In *Emergence in Science and Philosophy*, edited by Antonella Corradini and Timothy O'Connor, 6–78. New York: Routledge.
- O'Connor, Timothy, and Hong Yu Wong. 2002. "Emergent Properties." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. Accessed June 22, 2017. <http://plato.stanford.edu/archives/win2002/entries/properties-emergent/>.
- O'Connor, Timothy, and Hong Yu Wong. 2005. "The Metaphysics of Emergence." *Noûs*, 39(4): 658–678.

# Why Reject Christian Physicalism?

ANGUS J. L. MENUGE

## 26.1 Introduction

According to Christian physicalism (CP), a human person does not have an immaterial soul, but is identical to (van Inwagen 2007; Merricks 2001) or constituted by (Baker 2000; Corcoran 2006) a physical object. On this view, what appear to be the distinguishing features of minds either reduce to or emerge from (Murphy 2006) physical properties of the brain. CP is appealing to many because it seems to reconcile Christian anthropology with a modern scientific worldview and to avoid the classic interaction problem for substance dualism. Yet many Christian thinkers are concerned that CP is incompatible with important Christian doctrines, such as the incarnation (Van Horn 2010) and resurrection (Loose 2012). In this chapter, I will focus on several reasons to think CP does not adequately account for our stewardship obligations.

Stewardship requires a first-person perspective so that the steward can distinguish himself from the rest of creation and grasp moral obligations that apply consistently to his present and future actions. Despite claims to the contrary, I argue CP does not give an adequate account of our first-person perspective (Section 26.2). To be a steward, one must also carry out plans to take care of the world, and this assumes that we have knowledge of the natural world, can reason out solutions to environmental problems and can implement those solutions by our actions. But I argue that CP makes it very difficult to see how we can know the world (Section 26.3), or reason to conclusions about how best to care for it (Section 26.4) or act on our reasoning (Section 26.5). Finally, I offer some reasons to doubt that CP really qualifies as a physicalist view (Section 26.6).

If correct, my critique cuts deeply against CP for two main reasons. First, since stewardship is a scriptural mandate, an adequate Christian anthropology must account for our ability to carry out that mandate. Second, the capacities required to be stewards of the natural world are also presupposed by scientific investigation, so if CP fails to account

for these capacities, this undermines CP's claim to fit well with a modern scientific worldview.

## 26.2 The First-Person Perspective

Both Lynne Baker and Kevin Corcoran defend versions of the view that a human person is constituted by a living human body. They claim that a person is spatially coincident with its body, sharing all of its matter, but that the person is not identical with that body because the person and the body have different persistence conditions. In order to be a person, one must not only have a living human body, but also possess certain psychological capacities, including intentionality (Corcoran 2006, 67), and a first-person perspective (Baker 2001, 2013; Corcoran 2006, 68).<sup>1</sup> By a first-person perspective is meant the unique way in which a self is aware of itself subjectively (that is, *as itself*, e.g., as *me* or *I*) rather than in an impersonal manner (that is, *as a person*, e.g., as *the person typing this*). There is a difference between my thought that *I* will escape to Santa Fe tomorrow and my thought that the person typing this will do so. In the latter case, my wife could have a thought with the same content; in the former case she could not, since she can think subjectively of herself but not of myself.

Now in order to possess a (nonillusory) first-person perspective, there must first be a self – a single, unified, persisting consciousness – to become aware of. If there are multiple, independent streams of consciousness, then the referent of “I” is problematic even *at* a time, and if there are many independent self-stages *over* time, we have the problem that the “I” who escapes to Santa Fe tomorrow may not be the same as the “I” who thinks he will do so today. It seems undeniable that we normally do think of ourselves as single, unified, persistent conscious beings (even if the consciousness is intermittent, due to sleep or anesthesia), and without this assumption, we have no foundation for the first-person perspective necessary to clearly distinguish ourselves from the world and to recognize our ongoing obligation to care for it.

But does a physicalist anthropology provide an adequate account of the emergence of such consciousness? Consider the way in which thoughts and experiences are integrated into the mental life of one subject. Suppose Paul simultaneously sees the orange glow of a sunset, hears birds roosting, smells the aroma of coffee, and thinks about Plato's argument for the immortality of the soul. Part of what is interesting is that such diverse mental states all belong to one person. This is already a serious puzzle for physicalism, as the brain does not appear to provide a unified subject. In fact, neuroscience reveals a brain which processes widely distributed informational signals in parallel. Considering only the physical facts, it seems that there is no one entity that exists over and above the parallel streams of information processing in the brain. For example, evolutionary psychologist Steven Pinker asserts:

There's considerable evidence that the unified self is a fiction – that the mind is a congeries of parts acting asynchronously, and that it is only an illusion that there's a president in the Oval Office of the brain who oversees the activity of everything. (Pinker 1999, 14)

Likewise, Daniel Dennett compares cognition to multiple, partial narratives assembled by different sources at different times, like different trains carrying different informational freight on different railway lines. He insists that “there is no one place in the brain through



which all these causal trains must pass in order to deposit their content" (Dennett 1991, 135).

But more surprising, and a deeper problem for physicalism, is that Paul's mental states *cannot* be shared with anyone else: they are inseparable from his consciousness and intrinsically tied to it. Of course, it may be that Paul causes Jennifer to have the same attitude toward the same content as he does (so that, for example, both believe that the Eiffel tower is in Paris), but there is no physical process by means of which Paul can transfer one of *his* mental states to her: however alike, Jennifer's thought that the Eiffel tower is in Paris is different from Paul's simply because it is a mode of her consciousness and not his. However, an individual's brain is an aggregate of separable parts: these parts can exist without the whole, and can exist inside another whole as well (Moreland 2009, ch. 5). So, it is possible for a part of Paul's brain to exist outside of his brain, and even for it to exist in Jennifer's brain: there is a physical process by means of which Paul can transfer a neuron from his brain to Jennifer's brain.

Due to this basic contrast between physical aggregates and mental states, physicalism has a very hard time explaining why an agent's mental states are inseparable from him. If physical aggregates consist of separable parts, how can their mere recombination in different external relationships explain the occurrence of mental states which are intrinsically tied to the mental life of one subject?

If it is hard to see how consciousness could be broken down into parts like physical ones, perhaps that is because consciousness cannot be composed of such parts in the first place. That is what David Barnett maintains in what I will call the "union argument" (Barnett 2010). Barnett's argument depends on a simple "datum" which appears to lack counterexamples:

- (U) The union of two conscious beings is not conscious.

Thus, no matter how we physically configure Paul and Jennifer, and no matter how that alters each person's consciousness, we do not find that a third consciousness *instantaneously* emerges. The hard part is to explain why (U) is true. Barnett argues that if physicalism is correct, (U) must be true only because of contingent physical circumstances. For on the physicalist account, a unified consciousness must be something produced when a physical system is in the right state. So, a physicalist could maintain that (U) is true because the pairs of people: (1) lack a sufficient *number* of parts; (2) lack parts capable of standing in appropriate *relations*; (3) lack parts of the right *nature*; (4) lack the right *structure*; or (5) lack some combination of (1)–(4).

However, we can add billions of people and still no new consciousness emerges, so (1) is not the problem. And, with the help of another individual, David, and some science fiction, we could shrink Paul and Jennifer down to the size of David's left and right cerebral hemispheres, train Paul to function just like a left hemisphere and Jennifer to function just like a right hemisphere, and then replace the hemispheres of David's brain with Paul and Jennifer. Whether David's consciousness survived or not, it is implausible that a *new* consciousness arises from the mere union of Paul and Jennifer (that is, a consciousness that is not Paul's, Jennifer's, or David's), so (2) does not seem to be the problem either. Nor does the nature of the parts seem relevant, since consciousness does not seem to arise from pairs of anything, conscious or not: dogs, elephants, galaxies, circus trains, and so on. So (3) is not the problem. And finally, if having the right structure were all that was needed then so long as the same connections are maintained when Paul and Jennifer are inserted as David's left

and right hemispheres, the resulting structure should generate a new consciousness. But there is no good reason to think that this is true. And, since the brain is a structure of separable parts, surely one would predict that if a conscious subject did emerge, its experiences and thoughts would be separable parts, which would not explain the unity of consciousness.

Still, what about combining (1)–(4): would that account for the unity of consciousness? Barnett's ingenious response adapts a famous thought experiment due to Ned Block (1978) to consider how one might attempt to close the gap between a pair of people and a human body in stages. We just need billions of people (number) who emulate the interconnections of neurons (relation) in the same configuration as the brain (structure). The nature of the parts cannot be the crucial issue either because we can interchange neurons and people with the same function. So, if it is absurd to think that the structure of people gives rise to a new consciousness (as it were, a consciousness of the crowd), it must be equally absurd to suppose that the structure of neurons does.

It seems to me that the union argument provides a good response to one of Kevin Corcoran's concerns about dualism. Corcoran frankly admits that no version of physicalism has been able to explain consciousness. However, his disarming *tu quoque* is that this provides no grounds for preferring dualism:

Is it really any easier to see how an immaterial soul could be conscious than to see how a material being could be? It does not seem so . . . Consciousness is a mystery for all of us. (Corcoran 2006, 63)

However, the union argument does make it easier to see how an immaterial soul could be conscious. Immaterial souls are simples: they do not consist of spatial parts, and cannot be decomposed into spatial parts or separable parts of any other kind. A soul is therefore precisely the kind of being we would expect to be capable of a unified consciousness with thoughts as inseparable parts (or modes) of that consciousness. By contrast, bodies are aggregates of separable parts, and the union argument shows us that combining such parts does not suffice to account for a new conscious being. And how could it, if the parts of such a being are intrinsically inseparable?

This argument can be buttressed by noticing that brains are vague entities, in that, at any given time, it is unclear which physical entities do and do not belong within them. As Dean Zimmerman points out, when we consider plausible physical candidates for the referent of "I," they all "appear surprisingly like clouds on close inspection: it is not clear where they begin and end, in space or time. Many particles are in the process of being assimilated or cast off; they are neither clearly 'in,' nor clearly 'out'" (Zimmerman 2011, 187). Just as there are many sets of particles with an equal claim to be a particular cloud, there are many sets of particles with an equal claim to be a particular brain.

Igor Gasparov persuasively argues that the vagueness of brains inevitably affects any mental subject that allegedly emerges from them (Gasparov 2015, 434). The explicit target of Gasparov's argument is Dean Zimmerman's own emergent dualism, but his point also makes trouble for the nonreductive materialism of Baker, Corcoran, and Murphy, since each of them requires a self to emerge. Gasparov observes that if there are many sets of particles  $P_1, P_2, P_3, \dots, P_n$ , and each of the  $P_i$  ( $1 \leq i \leq n$ ) has an equal claim to be the brain, then if any one of them generates a mental subject, then presumably they all should do so. But it is also implausible that each of the  $P_i$  generates the *same* mental subject: this

would entail an extraordinary degree of overdetermination and would also make it likely there is no informative explanation of why a single, persistent particular self emerges because there is no stable feature that all of the  $P_i$  have in common. This generates a synchronic problem: why is there just one mental subject at a time? But there is also a related diachronic problem, since over time there will be many different sets of particles  $P^*_1, P^*_2, P^*_3, \dots, P^*_n$  and each of the  $P^*_i$  might equally well generate an emergent self, so there is no clear answer to when such a self would begin (Gasparov 2015, 437–438). If all of the  $P^*_i$  are equally sufficient to generate a self, it seems unlikely that they would all generate the same self since this would again entail massive overdetermination and risks making emergence nonexplanatory because the  $P^*_i$  share no significant common feature in virtue of which a unified, persistent self is to be expected. More likely would be a succession of instantaneous, nonidentical selves,  $S_1, S_2, S_3, \dots, S_n$ , but this does not explain how a self can persist over time. In sum, how does a unified, persistent self arise and persist, given a multiplicity of sets of particles at and over time, each with an equal claim to make up the brain?

One way defenders of CP may respond to such worries is by denying that the human bodies that constitute persons can be understood as mere physical aggregates of separable parts. The same living organism plausibly persists despite massive changes in those parts. So if a living human body gains its unity from a source independent of its parts, it is not a mere aggregate, and so the vagueness of that aggregate may be irrelevant. Then the unity of a living human body might account for the unity of the person it constitutes, making less surprising that person's unified, persistent consciousness.

In this vein, Corcoran follows John Locke in maintaining that the identity of living things does not depend on the particular material parts they include, but only on there being some parts or others united into one life. Corcoran tells us that a life is a "biological event . . . that is remarkably stable, well individuated, self-directing, self-maintaining and homeodynamic" (Corcoran 2001, 206). And he offers a striking simile of the human body with a storm, a simile which arguably accommodates Zimmerman's and Gasparov's concerns about vagueness, since storms do not have well-defined boundaries and we can hardly say which elements are definitely inside and outside of a storm:

Human bodies are like storms. A tornado, for example, picks up new stuff and throws off old stuff as it moves through space. Human bodies . . . are storms of atoms moving through space and time. They take on new stuff . . . and throw off old stuff as they go. (Corcoran 2006, 72)

Corcoran's account is aimed at explaining persistence over time (the diachronic problem), but presumably sharing in one common life also explains the identity of a body at a time (the synchronic problem). If each body is distinguished from all others not by its parts, which may be in flux and subject to vagueness, but by its unique life, perhaps the unity of the person and that person's consciousness derive (under appropriate conditions) from the unity of the body.

While this seems to be the most promising option for a (nonreductive) defender of CP, it faces several objections. First, the life that defines a body's identity does not seem to be defined in physicalist terms. Talk of a system that is "self-directing," "self-maintaining," and "homeodynamic" (capable of dynamic self-organization as an organism adapts to change) sounds irreducibly teleological, and it is not clear how such talk can be anchored in a truly physicalist ontology. And this suggests a dilemma. If the teleology is simply a brute part of

the physical world, this no longer sounds like physicalism, which at the most fundamental (nonemergent) level standardly recognizes only undirected efficient causation between particles. But if it is claimed that such teleology “emerges” from suitable complex physical systems, this is almost as puzzling as the claim that a unified, persistent consciousness emerges. Surely the arguments of Zimmerman and Gasparov can be adapted to show it is unlikely that one and only one life emerges and persists from the many different sets of particles with a claim to belong to a human body.

Second, the phrases “self-directing,” “self-maintaining,” and so on seem question-begging because in order for a system to be *self-directing*, and so on, there must be something that makes that system a single entity at and over time in the first place. To say that the identity of something depends on its *self-directing* power assumes that there is *one* thing which has such a power, but this is what needs to be justified. In this regard, Barnett points out that we must beware of a sleight of hand which occurs when physicalists talk of “one” body (Barnett 2010, 166–167). If consciousness is a property of simples, then it will be conceivable to us that anything *presented to our minds as a simple* could be conscious. We can easily present a body to our minds as a simple, as when we refer to it as “Paul” or “the chap over there.” But seeing a body *as* a simple does not make it so. In purely physical terms, bodies appear to be no more than structured aggregates of parts (albeit vague ones). As soon as we think of bodies in this way (i.e., as composites), we can no longer see how a unified consciousness can arise from them.

A third problem is that, to the extent that Corcoran’s account is plausible, it appears to rely on a dualism of kind between living and nonliving systems. This sounds vitalist, and a hard-nosed physicalist would find vitalistic dualism just as objectionable as mind-body dualism. Here it is worth noting that for both Aristotle and Aquinas, a “nutritive” soul is required to explain why something is alive; animals are distinguished from plants by a “sensitive” soul that includes the capacities of a nutritive soul but adds the capacity for sensations (Goetz and Taliaferro 2011, 19–20, 38–49). Arguably, Corcoran’s proposal removes the need to recognize souls in psychology only by appeal to something very like a soul in biology.

And it seems a version of the union argument can be run all over again, this time targeting the idea of a single living organism. It does not appear that the union of two or more living systems instantaneously yields a new living organism. Thus, if a patient receives one or more donor organs, they are (at best) assimilated into that patient’s body, but they do not give rise to a new living body (nor do the donors become new living bodies). And, it is conceivable that a living organism might go through an organic “ship of Theseus” transition: starting with a body at time  $t$ , there could be an extended series of part replacements, such that when the series is complete at time  $t + k$ , the resulting body has no part in common with the body at  $t$ . Even supposing that the body at  $t$  and the body at  $t + k$  each explain the emergence of a life, physicalism gives no reason to expect it to be the *same* life in both cases. Yet it seems plausible that such a transition could allow us simply to extend the life of the body at  $t$ . After all, although our part-replacement scenario is science fiction, something very like this happens to human bodies naturally as their constituent matter is renewed.

Corcoran might respond by appealing to the idea of immanent causation. He tells us that “In immanent causation, a state  $x$  of thing  $A$  brings about a consequent state  $y$  *in A itself*” (Corcoran 2006, 72). So he might say that there are immanent causal connections between the body at  $t$  and the body at  $t + k$ . However, appeal to immanent causation begs the

question by assuming that “the thing A” is a well-defined, persistent entity and this is what physicalism must justify.

On physicalism, it seems that the life of an organism is not a plausible emergent feature and so does not help to explain the emergence of a unified, persistent consciousness. At best, it seems that psychological dualism has been eliminated only by appeal to biological dualism. And if one cannot locate the unity of a living organism within a physicalist ontology, why say that talk of such unity is physicalist talk rather than dualist talk?

I conclude that nonreductive versions of CP do not succeed in providing a credible physicalist account of the emergence of a unified, persistent consciousness. And it is well-known that reductive forms of physicalism (including versions of CP that simply identify persons with material objects) have an even harder time providing such an account. As David Chalmers (1996) has argued, the “hard problem of consciousness” is that, given all of the physical facts about the brain, consciousness is simply a new fact, in no way entailed, predicted or even suggested by those physical facts. But without a unified, persistent consciousness, there is no self to become aware of, and hence no basis for a first-person perspective.

Yet such a perspective is essential for a steward to distinguish himself from his environment and to recognize obligations that apply to his ongoing actions of caring for that environment. This is strike one against CP as an account of creatures called to care for the world.

## 26.3 Knowledge of the World

Implicit within the previous argument are reasons to doubt that, on physicalism, we can have self-knowledge, mainly because physicalism makes it hard to see how selves can exist. But even if CP can solve this problem, a further difficulty is that a physicalist anthropology appears incompatible with the knowledge of the natural world required to be its stewards. There is a general problem – which makes it difficult to see how physicalism can account for any knowledge of external reality – and more specific problems for our ability to know reality at various levels. There is also the problem of how we can know God, and hence know that stewardship is His command.

### 26.3.1 *The general problem*

As Scott Smith (2012) has argued, in order to know something X, one must have a valid concept of X and an ability to verify whether X is actually present. Thus to know that there is an apple on the counter, I must have a valid apple concept and an ability to verify that an apple is present, and I do this by comparing an object given in experience with my apple concept to see if they match. But as simple as this sounds, it is deeply problematic on a physicalist ontology.

The root of the problem is that, on physicalism, there is no intrinsic intentionality. For physicalism, intentionality – the fact that thoughts and experiences are about, or of, certain objects beyond themselves – is not a fundamental feature of anything physical, including brain states, because intentionality is not recognized by any physical science. So, intentionality is at best an emergent capacity. Intentionality arises when we acquire the ability to conceptualize our brain states in a certain way, allowing us to interpret those states as being

about something beyond themselves. Now this applies to both our thoughts and our experiences. Neither a thought about an apple nor the experience of an apple is intrinsically of its object (an apple). But this means we have no nonconceptual access to objects (apples) and this makes it difficult to see how we can verify that an external object is present.

In fact, there are two problems. The most obvious is that we have no good reason to think that we actually perceive external objects like apples. All we have is a brain state conceptualized to be an experience of an apple, but since we have no *nonconceptual* access to apples, it is entirely possible that what we are actually experiencing is just the state of our own brain. And we also cannot confidently verify our thought that an apple is present by comparing features of the apple to our concept of the apple, because there is no reason to think that an apple has been given in our experience. Of course, it might just be that our thoughts and experiences carve nature at the joints, but this would seem to be an unlikely and unexplained coincidence.

But matters are made worse by a second problem: physicalism does not give a credible account of how concepts are formed in the first place. For me to know an apple is present, I must not only be able to experience an apple as it is, I must also have a valid apple concept, one that accurately represents apples as having the properties that apples have. But such concepts are not innate but develop gradually. For example, one is introduced to many apples at stores and at home, in picture books, and so on, and one begins to form a rudimentary concept of an apple. This proto-concept is further refined by noticing the subtle differences between apples and certain near-misses easily mistaken for apples, until one has a fairly stable (if fallible) apple concept that reliably (if approximately) represents the features that (most) apples have and (most) non-apples do not have in normal conditions. However, this process of concept formation presupposes that one has non-conceptual access to apples, by means of which the apple concept can be refined and corrected by apples themselves. Our concept cannot become an even approximately valid concept of apples unless it is rooted in experiences that present apples as they are. But physicalism seems bound to say that all experiences of an object are already conceptualizations of brain states. If those conceptualizations are in error, there is no way to correct them by experiencing objects as they are.

For these reasons, a physicalist ontology would make it likely that we are locked in to the prison-house of our brain, unable to know the world as it is. It is fair to say that the scriptural mandate to be stewards of creation is not a call to care for our brain states. Rather, stewardship requires us to know external reality as it is, and arguably, this requires immaterial beings like souls whose thoughts and experiences exhibit intrinsic intentionality.

### 26.3.2 *Levels of knowledge*

Our knowledge of the natural world comes at a number of levels. At the highest level, we know that there *is* a natural world – a cosmos or universe – which, *inter alia*, makes it possible to consider whether that world exists of itself or has an external cause. At the next level, we are aware of natural kinds, both living and nonliving, vital to our ability to steward natural resources, and to our concerns about biodiversity. Then there is knowledge of individuals – living and nonliving things, like particular lumps of iron or gold and particular cats or dogs, which is equally vital to our stewardship call because it is individuals that our actions directly affect.

If we set aside the general worry about physicalist epistemology explained above, there is a further problem for our knowledge of the cosmos. To know the cosmos as a cosmos, one must first acquire a valid concept of the cosmos. But on physicalism, concepts emerge as the result of causal interactions with the brain, and this means that those concepts are limited to what can be explained by the individual learning history of a human organism or natural selection operating on the human species. But the totality of these causal interactions occur in only small subregions of space and time which collectively do not exhaust the whole of space and time. It is not true that brains of any human organism, or of the species as a whole, have causally interacted with the entire cosmos. If so, one would predict that we only have local and parochial concepts of the spatiotemporal regions of the universe with which we have interacted, but that we would not have a valid concept of the cosmos. More generally, the problem for physicalism is that we do have (apparently valid) concepts that are not limited to the parochial spatiotemporal physical interactions of human beings with their environment.

We also appear capable of knowing that the world is made up of natural kinds of living and nonliving things, like wallabies and water. But in order to know a kind, there must be such a thing and we must be able to access it. On physicalist grounds, neither of these grounds for knowledge is obviously met. First, on physicalist assumptions, it is not clear that there really are stable kinds. This is because physicalists deny nonphysical essences, preferring to say that kinds naturally emerge when particles are regularly configured in certain ways. However, the boundaries between “kinds” thereby become fuzzy and arguably arbitrary. Can we say that gold is a stable kind if transformations to the elementary particles from which gold emerges would produce a different “kind” of substance? And, on a Darwinian view, species seem to be labels for creatures in a never-ending process of becoming other “kinds” of creatures. Darwin himself saw “the term species as one arbitrarily given, for the sake of convenience, to a set of individuals closely resembling each other” (Darwin 1993, 78–79). On this view, there is no such thing as a creature’s essence, shared by all and only creatures of that kind, a conclusion explicitly drawn by the great evolutionary biologist Ernst Mayr: “the essentialist philosophies of Aristotle and Plato are incompatible with evolutionary thinking” (Mayr 1970, 4). How can we be called to care for God’s natural resources and creatures if they do not belong to well-defined kinds? Arguably, these problems can be solved only if kinds are recognized as nonemergent, nonphysical essences, which abandons physicalism.

Our belief in organic or inorganic individuals is a belief in well-defined, persistent objects. Given that all of the candidate individuals we experience are subject to change, this makes sense only if we are able to access the modal properties defining an individual’s identity and persistence conditions: these modal properties tell us what must be the case, and what may or may not be the case, if the individual is to exist or persist. So, for example, we think that in order for a piece of gold to exist or persist, it must retain atomic number 79, but it may or may not be tarnished; and in order for a cat *Felix* to exist or persist, *Felix* must have cat DNA, but may or may not have fur.

Michael Rea has argued that on physicalist assumptions, the problem is that the modal concepts which specify the identity and persistence conditions of individuals appear to be causally inert, and this makes it impossible for us to know individuals (Rea 2000; 2002, ch. 4). To see the problem, suppose that *Felix* the cat appears before me. I can know that an individual cat is present only if I have some access to the cat’s identity and persistence conditions. Yet at the purely physical level, I can only interact with the series of clouds of particles that compose *Felix* over time. Whether any of these clouds composes a distinct individual at a time can make no difference to the causal powers of that cloud, so no cloud

enables me to know Felix as an individual at a time. Further, it also makes no difference to the causal powers of the entire series of clouds of particles whether a cloud later in the series composes the same or a different individual than a cloud earlier in the series. So I have no grounds for distinguishing a persistent cat from a series of cat-stages and therefore cannot know, for example, that Felix is the same cat I fed this morning.

The upshot is that, on physicalism, it is inexplicable how we can form a valid concept of a well-defined persisting individual and thereby know that such an individual is present. True, it is logically possible that we do have such concepts, but as Robert Koons (2010) points out, this still would not enable knowledge: we cannot know that  $p$  if the fact that  $p$  has nothing to do with why we believe that  $p$ , and so we cannot know that an individual  $X$  is present if the fact that  $X$  is present has nothing to do with why we believe it is.

Admittedly, we saw that Corcoran would want to ground the identity and persistence conditions of a living individual in a common life over and above the clouds of particles composing it. But not only is it doubtful that such a life can be located in a physicalist ontology, it is also unclear how such an entity could causally interact with brains. Physics and chemistry allow that particles can affect brains, but no physical science explains how a “life” could do so.

### 26.3.3 Knowledge of God

Christians claim to know a lot about God, both from His general revelation in nature and His special revelation in scripture. And in particular, they claim to know that God commands us to take care of the world. But a physicalist ontology makes it difficult to see how we could have certain “transcendent” concepts necessary to think about God and hence know that He exists, what He is like, and what He commands. Among others, these transcendent concepts include those of an infinite, necessary and perfect being. However, on physicalist assumptions, our valid concepts are limited to those which can be explained by a finite number of interactions between contingent, imperfect brains and a contingent, imperfect physical world. But then it seems that we could have no basis for forming valid transcendent concepts. Even if these concepts did emerge, they would not be the result of reality exemplifying the corresponding qualities (infinity, necessary existence, perfection) and so they would not suffice for knowledge. Indeed, Thomas Hobbes is quite explicit in arguing along these lines.

Whatever we imagine is *finite*. Therefore there is no idea or conception of any thing we can call *infinite*. No man can have in his mind an image of infinite magnitude, nor conceive of infinite swiftness, infinite time, infinite force or infinite power. When we say any thing is infinite, having no conception of the thing because of our own inability, we only mean that we are not able to conceive the ends and the bounds of the thing named. And so the name of God is used that we may honor him and not to make us conceive him, for he is incomprehensible and his greatness and power are inconceivable. (Hobbes 2008, 13)

If what we can “imagine” is limited to the results of the physical interactions of brains with the physical world, similar arguments would show that we have no concepts of necessary existence or perfection.

Of course, unlike standard physicalism, CP allows that God exists and created and sustains the world. So it seems God can causally influence His creatures, either directly or by means of that world. Could that explain our possession of the transcendent concepts needed to think about God? It is hard to see how. The bottleneck is that, on physicalism, our



concepts are limited to those which can emerge from a finite, contingent and imperfect brain. Even if an infinite, necessary and perfect God interacts with that brain (either directly or through nature), the brain's limited bandwidth makes it hard to see how we could acquire the information required to have transcendent concepts.<sup>2</sup> It is also difficult to see how, on physicalism, this knowledge could be implanted simply because (as we saw) there is no well-defined persistent subject in whom such knowledge could be implanted. It seems likely that any account of the knowledge of God would require resources that are not a good fit with physicalist anthropology.

Stewardship requires knowledge of reality as it is – of the cosmos, natural kinds, and individuals, and also of God. But all of this is deeply problematic if we are, at base, purely physical beings. In this context, appeal to emergence is explanatory only if one can specify what it is about the physical base properties that makes it probable that such knowledge would emerge. Without this, no reason has been given to prefer a physicalist over a dualist anthropology. This is strike two against CP as an account of creatures called to care for the world.

## 26.4 Reasoning

Stewardship requires both theoretical reason (for example, to conclusions about the nature of environmental problems and effective solutions) and practical reason (for example, from beliefs about the environment and the desire to care for it to a decision to curb emissions or reduce toxic waste). However, an ontological analysis reveals that an agent must satisfy a number of conditions in order for an act of reasoning to occur. Once those conditions are clearly understood, it makes it very difficult to see how reasoning is possible if we assume a physicalist anthropology.

Consider a simple act of conscious theoretical reasoning. For example, suppose that Karl consciously draws a conclusion from two beliefs by *modus tollens* (if A then B, not-B, therefore not-A). For this to make sense, we must suppose that there is some one entity that consciously believes both that if A then B and that not-B. It is not sufficient for the information that if A then B and not-B merely to reside in Karl's brain. The brain is a distributed system, a structured aggregate of parts, and subject to vagueness, and as we have seen, from a physicalist perspective it is hard to see how this system accounts for the existence of a single, unified mental subject. But a soul, being substantively simple, is not composed of separable parts and is the kind of entity that can unify two beliefs, so that we can say that one subject, Karl, believes both of them. This is essential to understanding Karl's rationality, for it is only if Karl is one subject, who believes both that if A then B, and that not-B, that it can be rational for Karl to draw the conclusion that not-A. If one part of Karl's brain contains the information that if A then B and another contains the information that not-B, then neither of those parts can have a reason to conclude that not-A. But in fact, even if the information happened to be held in the *same* place in the brain, there is no reason to think that that place, being itself a vague aggregate of parts, is capable of conscious unity, so there is still no reason to say that there is a unified subject that owns all of the beliefs.

And regardless of whether the information is distributed or localized, there is no good reason to think that the location or locations in the brain exhibit subjectivity either because of the hard problem of consciousness. Merely invoking emergence does not solve this problem: physicalists must say what it is about the physical base properties of brains that

necessitates consciousness (or makes it probable). This has not been done and the union argument indicates that physicalists lack the resources to do it.

However, even supposing physicalists could solve these problems, they would not be out of the woods. For not only does such an act of reasoning require conscious unity at a time, it also requires conscious unity over time. Suppose after attending to his beliefs, Karl draws the logical conclusion that not-A. We can only credit Karl with reasoning to this conclusion if there is some one entity that persists over the time it takes to draw the conclusion. It must be that there is one thing which believes if A then B and not-B and which concludes that not-A. Otherwise it will be like Jack believing if A then B and not-B and Jennifer believing not-A. Neither Jack nor Jennifer can be said to reason to the conclusion, because Jack does not survive long enough to draw it, and although the correct conclusion occurs to Jennifer, she cannot be said to have drawn that conclusion from her prior beliefs.

But as we have seen, physicalism has an equally hard time explaining the persistence of a mental subject over time. For over time, the brain is in constant flux and there is a sequence of clouds of particles  $P_1, P_2, P_3, \dots, P_n$ , and each of the  $P_i$  ( $1 \leq i \leq n$ ) has an equally good claim to serve as the emergent base of a mental subject. If that is the case, supposing that any one of the  $P_i$  gives rise to a mental subject, they all should do so, and this makes it most probable that what emerges is a sequence of different self-stages,  $S_1, S_2, S_3, \dots, S_n$ , not a single persistent self. If that is the case, sustained acts of reasoning do not make sense, because the self-stage that considers the premises of an argument will likely not be the same as the self-stage which considers the conclusion. For example, if  $S_5$  considers if A then B and not-B, but  $S_6$  considers not-A, then  $S_5$  did not survive long enough to draw the conclusion, and although the correct conclusion occurred to  $S_6$ ,  $S_6$  did not reason to that conclusion either, since  $S_6$  was never aware of the premises.

What sort of being could persist through an act of reasoning? One thing that could do it is a soul. Not being composed of separable parts, a soul cannot be destroyed by removing or disconnecting its parts. And a soul is a substance, so it can endure all sorts of accidental change in its mental contents and relations to physical objects. So a soul is the ideal thing for persisting over time and through change: a series of internal changes in its occurrent thoughts or of external changes to the organism to which the soul belongs (e.g., metabolic changes which occur while someone is thinking) are compatible with it being the same soul from beginning to end of the series. The soul also has the right kind of unity over time to explain how an agent can be credited with drawing a conclusion. Although a soul's thoughts occur at different times, each of those thoughts belongs to the same soul (all are Karl's thoughts and no one else's), so that one can say it is the very same soul which believes that if A then B and not-B and which draws the conclusion that not-A.

Again, Corcoran would presumably seek to avoid this conclusion by appeal to the common life uniting the clouds of particles in a brain at and over time. But as we saw, it seems no more credible to claim that a unified persistent life emerges from a base consisting of a sequence of clouds of particles than to claim that consciousness emerges from this base. If so, reasoning makes no sense, and this is strike three against CP.

## 26.5 Mental Causation

In baseball the saying is "three strikes and you are out." But there is another reason why I think CP should remain on the bench: it has great difficulty in accounting for actions – like batting – since it does not plausibly account for mental causation (see Menuge 2013). This is because, as

Jaegwon Kim (1998, 2007, 2011) has argued, if mental properties emerge from a physical base they appear to be excluded by that base from having any distinctive causal power.

For suppose that emergent mental state  $M$  causes emergent mental state  $M^*$ . For example, a conscious self affirms some reasons and not others for an environmental policy ( $M$ ) and this produces a rational decision ( $M^*$ ). According to emergentism,  $M^*$  is synchronically necessitated by its physical base state  $P^*$ . It is also reasonable to assume that if  $M^*$  is multiply realized, no alternative base state is actually present. So, in the closest worlds in which  $P^*$  is absent,  $M^*$  would not happen. But if so, the only way  $M$  could cause  $M^*$  is by causing  $P^*$ . Thus, if emergentism is true, mental-to-mental causation requires downward (mental-to-physical) causation. However, as an emergent state,  $M$  must also have a physical base  $P$ . But then we have a problem:

If causation is understood as nomological (law-based) sufficiency,  $P$ , as  $M$ 's emergence base, is nomologically sufficient for it, and  $M$ , as  $P^*$ 's cause, is nomologically sufficient for  $P^*$ . It follows that  $P$  is nomologically sufficient for  $P^*$  and hence qualifies as its cause. The same conclusion follows if causation is understood in terms of counterfactuals. (Kim 2006, 558)

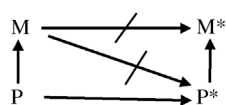
Yet if  $P$  qualifies as the cause of  $P^*$ , and hence  $M^*$ , unless we allow an implausible systematic overdetermination,  $M$  is excluded from any causal role. The situation is summarized in Figure 26.1.

Now some nonreductive physicalist argue that mental properties are not made redundant by their realizing brain states because the mental properties account for abstract patterns we would miss by attending to the brain states themselves. Thus there are many physically different ways to greet someone (saying "Hi," waving a hand, sending a note, sky-writing a welcome, etc.) and while a variety of different specific neurological causes would best explain each of the sequences of movements, the common action of greeting is best explained by a common mental intention to greet. This is why Terence Horgan, a leading exponent of nonreductive materialism, says that there is "no contest" between mental and physical causation: they coexist at different levels (Horgan 1997). A similar position is defended by John Gibbons (2006).

However, I think Kim's exclusion argument prevails. Kim's argument begins by noting that any serious physicalist must accept the principle of the causal closure of the physical domain (PCC):

Pick any physical event . . . and trace its causal ancestry or posterity as far as you would like . . . this will never take you outside of the physical domain. Thus, no causal chain involving a physical event ever crosses the boundary of the physical into the nonphysical. (Kim 2011, 214)

To deny PCC is to allow the causal contribution of nonphysical entities, such as immaterial souls, and no physicalist can do that. Yet if one grants PCC, it seems impossible to maintain that mental properties make any causal contribution.



**KEY:**

Vertical arrows signify *emergent determination*.

Horizontal arrows signify *causation*.

Crossed arrows are excluded causal pathways.

**$M$  is excluded by  $P$ , which causes  $M^*$  by causing  $P^*$ .**

**Figure 26.1** The Exclusion Problem for Emergent Physicalism.

To see why, suppose Jack intends to greet his friend and does so. Then mental state *M* (intending to greet) emerges from physical state *P* (a brain state), and *M* is succeeded by physical state *P*<sup>\*</sup> (some greeting behavior).<sup>3</sup> So it looks as if *M* causes *P*<sup>\*</sup>. However, given PCC, there must be some sufficient purely physical cause of *P*<sup>\*</sup>. The obvious candidate for this is the emergent base *P* of *M*, for two reasons. First, *P* is at the right level of specificity to explain the particular movement *P*<sup>\*</sup>, whereas *M* is compatible with a variety of different greeting movements. Second, consistently with PCC, *M* cannot contribute anything over and above the purely physical causes, so anything *M* does must reduce to what *P* does. But if *P* suffices to produce *P*<sup>\*</sup>, what work is there left for *M* to do? The idea of systematic causal overdetermination (that *P*<sup>\*</sup> is produced by both *P* and *M*) seems implausible, so it is natural to conclude that *M* has no causal power at all.

Can this argument be mitigated by multiple realization and talk of levels, as Horgan and Gibbons believe? It is true that there may be multiple physical bases *P*<sub>1</sub>, *P*<sub>2</sub>, *P*<sub>3</sub>, . . . *P*<sub>*n*</sub> which realize *M*, and that *P*<sup>\*</sup> is only one of many realizers *P*<sup>\*</sup><sub>1</sub>, *P*<sup>\*</sup><sub>2</sub>, *P*<sup>\*</sup><sub>3</sub>, . . . *P*<sup>\*</sup><sub>*n*</sub> of the action of greeting. This certainly makes it convenient, in capturing abstract patterns, to continue to speak of *M*. But it is no reason to say that *M* has independent causal power. On pain of denying PCC, the physicalist must concede that on any given occasion, it is always one of the *P*<sub>*i*</sub> which is causally responsible for one of the *P*<sup>\*</sup><sub>*i*</sub>, and the causal action happens at the level of particular states (property instances), not types of state (the properties themselves). If so, it follows that the causal contribution of *M* is exhausted by the disjunction of its bases *P*<sub>1</sub> ∨ *P*<sub>2</sub> ∨ *P*<sub>3</sub> ∨ . . . *P*<sub>*n*</sub>. So although *M* remains useful (perhaps even practically indispensable) shorthand for this disjunction, *M* makes no distinctive causal contribution.

Kim's argument shows that emergent physicalism is a house divided against itself, because the whole point of emergent properties is that they bring new causal powers into the world, yet these powers are excluded by physicalist assumptions.<sup>4</sup> It seems that if one wants to defend what makes the idea of emergence interesting (the appearance of novel causal powers), one should be an emergent dualist, like William Hasker (1999), although that view faces its own problems, as Gasparov (2015) notes. Kim's own solution to this problem is to assert that mental properties just are physical properties, yet he admits this makes paradoxical his claim to affirm mental causation:

If we save mental causation by reducing mentality to mere patterns of electrochemical activity in the brain, have we really saved mentality as something special and distinctive? (Kim 2011, 220)

It seems that both reductive and nonreductive versions of physicalism find it difficult to account for *distinctively* mental causation, and this may be a reason to reconsider the soul as a mental entity with its own causal powers.

## 26.6 But is It Physicalism?

Beyond its apparent biological dualism, another reason for doubting that CP is really physicalist is very simple. God manages to do all those things which the philosophy of mind finds metaphysically "hard" without a body: He is conscious, He reasons, and He persists as the same person (eternally, in fact). So, although bodies are God's idea, and Christians emphasize the resurrection of the body, no Christian can maintain that bodies are *required* for any of these hard things. Further, if we say it just seems inconceivable that human beings

can do these hard things without bodies, then the atheist seems entitled to say, “Quite so, and that is why there can be no such being as God.” To maintain that God exists but that human beings require bodies for the hard things makes God seem a strange add-on, the denizen of a “physicalism plus” position. After all, in one sense, believing in God *already* makes one a dualist, because one must acknowledge at least one wholly immaterial being capable of thought (not to mention angels, which can assume bodies but which are not essentially constituted by them).

So it seems Christian physicalism embraces a suspiciously selective form of dualism: God is oddly exempt from all the arguments insisting that embodiment is essential to do hard things. And it is arguable that Christian physicalism unwittingly provides comfort to atheism by making of God an awkward special case. If souls are redundant in explaining intelligent human thought and action, one may wonder why a pure spiritual being like God is needed to explain anything about our world that might seem to point to its being an intelligent creation. And here the alternative to physicalism is appealing. Dualism sees God not as an awkward special case but as the paradigm case of an intelligent personal being. If God’s maximal intelligence and agency is compatible with His having a soul, it seems quite natural to suggest that beings made in His image have souls as well.

## 26.7 Conclusion

If CP is properly confined to the resources actually available to a physicalist anthropology it seems unable to account for the capacities of stewards, including a first-person perspective, knowledge of the natural world, reasoning, and the ability to act. To the extent that CP does seem able to account for these capacities, it relies heavily on a nonexplanatory notion of emergence and apparently nonphysical notions of teleology and life. There is good reason for Christians not to embrace CP and to reconsider the case for an immaterial soul.

## Notes

1. Baker distinguishes between a “robust” and a “rudimentary” first-person perspective, where the former, but not the latter, requires self-referential language use. But even the pre-linguistic rudimentary form is sufficient for a person to exist.
2. My own view is that the transcendent concepts required to think of God are innate, instilled directly by God. This in no way denies that we may come to know of God via His creation. When St Paul says in Romans 1 that we can know of God’s existence and attributes from His creation, he does not say that our *concepts* of God arose from that creation. Rather, it seems plausible that on the basis of transcendent concepts we *already* possess, we are able to conclude that nature is the work of a transcendent being.
3. A more precise account would cite mental and physical property instances rather than the vague term “states.”
4. While abstract objects may be an exception, it is a fairly uncontroversial to state the identity conditions of temporal entities in terms of their causal powers.

## References

- Baker, Lynne Rudder. 2000. *Persons and Bodies: A Constitution View*. Cambridge: Cambridge University Press.
- Baker, Lynne Rudder. 2001. “Materialism with a Human Face.” In *Soul, Body and Survival: Essays on the Metaphysics of Human Persons*, edited by Kevin Corcoran, 159–180. Ithaca, NY: Cornell University Press.

- Baker, Lynne Rudder. 2013. *Naturalism and the First-Person Perspective*. New York: Oxford University Press.
- Barnett, David. 2010. "You Are Simple." In *The Waning of Materialism*, edited by Robert C. Koons and George Bealer, 161–174. New York: Oxford University Press.
- Block, Ned. 1978. "Troubles with Functionalism." *Minnesota Studies in the Philosophy of Science*, edited by C. Wade Savage, vol. 9, 261–325. Minneapolis: University of Minnesota Press.
- Chalmers, David. 1996. *The Conscious Mind*. New York: Oxford University Press.
- Corcoran, Kevin. 2001. "Physical Persons and Postmortem Survival without Temporal Gaps." In *Soul, Body and Survival: Essays on the Metaphysics of Human Persons*, edited by Kevin Corcoran, 201–217. Ithaca, NY: Cornell University Press.
- Corcoran, Kevin. 2006. *Rethinking Human Nature: A Christian Materialist Alternative to the Soul*. Grand Rapids, MI: Baker Academic.
- Darwin, Charles. 1993. *On the Origin of Species*. New York: The Modern Library.
- Dennett, Daniel. 1991. *Consciousness Explained*. Boston, MA: Little, Brown.
- Gasparov, Igor. 2015. "Emergent Dualism and the Challenge of Vagueness." *Faith and Philosophy*, 32(4): 432–438.
- Gibbons, John. 2006. "Mental Causation without Downward Causation." *The Philosophical Review*, 115(1): 79–103.
- Goetz, Stewart, and Charles Taliaferro. 2011. *A Brief History of the Soul*. Malden, MA: Wiley Blackwell.
- Hasker, William. 1999. *The Emergent Self*. Ithaca, NY: Cornell University Press.
- Hobbes, Thomas. 2008. *Leviathan*, edited by Marshall Missner. New York: Pearson Longman.
- Horgan, Terence. 1997. "Kim on Mental Causation and Exclusion." *Philosophical Perspectives*, 11: 165–184.
- Kim, Jaegwon. 1998. *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation*. Cambridge, MA: MIT Press.
- Kim, Jaegwon. 2006. "Emergence: Core Ideas and Issues." *Synthese*, 151(3): 547–559.
- Kim, Jaegwon. 2007. "Causation and Mental Causation." In *Contemporary Debates in Philosophy of Mind*, edited by Brian P. McLaughlin and Jonathan Cohen, 227–242. Malden, MA: Blackwell.
- Kim, Jaegwon. 2011. *Philosophy of Mind*, 3rd edn. Boulder, CO: Westview Press.
- Koons, Robert C. 2010. "Epistemic Objections to Materialism." In *The Waning of Materialism*, edited by Robert C. Koons and George Bealer, 281–306. New York: Oxford University Press.
- Loose, Jonathan. 2012. "Constitution and the Falling Elevator." *Philosophia Christi*, 14(2): 439–449.
- Mayr, Ernst. 1970. *Population, Species, and Evolution*. Cambridge, MA: Harvard University Press.
- Menuge, Angus. 2013. "Neuroscience, Rationality and Free Will: A Critique of John Searle's Libertarian Naturalism." *Philosophia Christi*, 15(1): 81–96.
- Merricks, Trenton. 2001. "How to Live Forever Without Saving Your Soul: Physicalism and Immortality." In *Soul, Body and Survival: Essays on the Metaphysics of Human Persons*, edited by Kevin Corcoran, 183–200. Ithaca, NY: Cornell University Press.
- Moreland, J. P. 2009. *The Recalcitrant Imago Dei*. London: SCM Press.
- Murphy, Nancey. 2006. *Bodies and Souls, or Spirited Bodies?* New York: Cambridge University Press.
- Pinker, Steven. 1999. "Is Science Killing the Soul?" *Edge*, 53, April 8. Accessed July 25, 2016. <http://edge.org/documents/archive/edge53.html>.
- Rea, Michael. 2000. "Naturalism and Material Objects." In *Naturalism: A Critical Analysis*, edited by William Lane Craig and J. P. Moreland, 110–132. New York: Routledge.
- Rea, Michael. 2002. *World without Design: The Ontological Consequences of Naturalism*. New York: Oxford University Press.
- Smith, R. Scott. 2012. *Naturalism and Our Knowledge of Reality: Testing Religious Truth-Claims*. Farnham, UK: Ashgate.

- Van Horn, Luke. 2010. "Merricks's Soulless Savior." *Faith and Philosophy*, 27(3): 330–341.
- van Inwagen, Peter. 2007. "A Materialist Ontology of the Human Person." In *Persons: Human and Divine*, edited by Peter van Inwagen and Dean Zimmerman, 199–215. New York: Oxford University Press.
- Zimmerman, Dean. 2011. "From Experience to Experiencer." In *The Soul Hypothesis*, edited by Mark C. Baker and Stewart Goetz, 168–196. New York: Continuum.

## PART III

# Substance Dualism, Theology, and the Bible



# Debating Biblical Anthropology

# Biblical Anthropology is Holistic and Dualistic

JOHN W. COOPER

## 27.1 Overview

From earliest times, the vast majority of Christian churches and scholars have professed a doctrine of humanity derived from Scripture that is generically holistic and dualistic. It is holistic in emphasizing the unity and integrity of human nature as created, redeemed in Christ, called to live in this world, and resurrected in the world to come. It is dualistic in affirming temporary disembodied personal existence between death and bodily resurrection. I defend this view of biblical anthropology.

But recently some Christian scholars have claimed that biblical anthropology is monistic and that body-soul dualism is a long-standing mistake resulting from the influence of Plato on the church fathers. Most Christian monists affirm nonreductive physicalism or psychosomatic monism, which entail that persons cannot exist without bodies and that any mode of postmortem existence must be embodied.

To engage the best of the current debate about biblical anthropology, I consider the main contributions of Joel Green (1998, 2002, 2005, 2008) and N. T. Wright (1992, 2003a, 2003b, 2011) – eminent Scripture scholars who address the philosophical issues. All three of us affirm *biblical holism* – the view that humans are integral, multidimensional beings. We all reject gnostic, Platonic, spiritualist, and materialist anthropologies which posit ontological and axiological divisions or reductions in human nature. Holism is not at issue.

But we do disagree about the relations among holism, monism, and dualism. Green and Wright reject dualism as inconsistent with holism. Green holds that both Scripture and science support the sort of monism which entails that persons are essentially embodied (Green 2008, 31), and that “there is no part of us, no aspect of our personhood, that survives death . . . Life after death requires embodiment” (179). Wright strongly affirms the two-stage biblical eschatology on which historic Christian dualism is based (Wright 1992, 2003a, 2003b). But he rejects both dualism and materialism in favor of ontological holism

(Wright 2011). I counter that biblical anthropology is holistic dualism or dualistic holism (Cooper 2000 [1989], 2001, 2007, 2009a, 2009b, 2015).

In Section 27.2 I show that key arguments against dualism are compromised by problematic hermeneutics, conceptual confusions, and faulty reasoning. Section 27.3 makes the case for dualism from Scripture's account of God's creation of humans and his sustaining us in relationship with him between death and bodily resurrection. I also show that monism cannot account for the texts which imply dualism, and that dualism is compatible with holism.

## 27.2 Key Terms and Interpretive Principles

Before considering Scripture, it is necessary to clarify key terms and issues that affect the outcome of the debate.

*Biblical anthropology* is the comprehensive understanding of human nature implied by all relevant data of Scripture and its context, synthesized into a coherent whole. Accounts of human nature that do not deal adequately with all the relevant texts, their relations, and background data are not fully *biblical* anthropologies. Green's anthropology is less than fully biblical.

I agree with Green and Wright that the anthropology of Scripture ought to be interpreted in terms of the texts' linguistic, literary, historical, and religious worldview meanings, and that words such as *body*, *soul*, *spirit*, *mind*, *flesh*, and *heart* have multiple connotations that developed from the Old to the New Testament. We agree that the biblical writers did not intend to teach a systematic anthropology, and that dualist theologians have sometimes read *body*, *soul*, and *spirit* as metaphysical terms. But the case for biblical dualism does not depend on such mistakes.

Regarding hermeneutics, I generally appreciate Wright's approach to Scripture. But I disagree with Green's, and this difference affects our conclusions about biblical anthropology.

Green's "neuro-hermeneutics" makes the historicist-postmodern assumption that "the 'truth' about human nature" is not "decisively determined by Scripture," but "what the Scriptures teach" is always found in dialectical relation to the presumptions brought by the interpreter" (Green 2008, 28). This approach implies that the truth-content of Scripture itself – not just our interpretations – is co-constituted by and changes with our presumptions. Green also argues that our presumptions for interpreting Scripture and constructing theology ought to *include* science, not just consider it, because we cannot avoid doing so and because the theology of the Bible writers and past interpreters includes the science of their times.<sup>1</sup> Thus he makes science one of the lenses through which he determines biblical doctrine. Even more problematic, his neuro-hermeneutics actually applies philosophical monism, not just science, to interpret Scripture and construct theology: Throughout his work he reiterates that science warrants the sort of monism which entails that human persons are essentially embodied, and he uses this assumption to determine biblical teaching. His conclusion that Scripture and science both support monist anthropology follows from his hermeneutical and philosophical assumptions, which beg the question of the debate.

I follow the classical Christian approach to the "two books" of revelation. It regards Scripture itself as expressing recognizable enduring truths, and it reads "the book of nature" with the lenses of Scripture to understand how creation reveals God and his will. It evaluates

readers' presumptions accordingly. It takes note of the human writers' beliefs about the natural order and determines whether they are enduring doctrine or dated perspectives. It does not confuse natural revelation with science or philosophy, or read them into the teaching of Scripture. This method leads to biblical dualism without privileging it in its hermeneutics.

*Dualism* is primarily a generic nonphilosophical worldview belief about human nature widely embraced by people who affirm an afterlife. *Biblical dualism* is the Christian belief that although God created and redeems humans as embodied persons, he sustains us disembodied between death and bodily resurrection. Thus it is also holistic. It views the person-body dichotomy as an abnormal and diminished condition resulting from sin and death. Therefore, holistic dualism does not imply the intrinsic tensions, separations, or reductions characteristic of such "radical" dualisms as Gnosticism, Platonism, and caricatured Cartesianism. Scripture's holistic dualism was formulated in various ways by the church's great theologians, and it remains normative in the liturgies and doctrinal standards of virtually all ecclesiastical traditions.

Philosophical dualism elaborates worldview dualism. It holds that a human being is an integral whole constituted by a material and an immaterial principle or ingredient and consisting of a physical organism integrated with a nonphysical person (synonyms: *soul, spirit, self, ego, subject, agent, mind*). It has several versions, including two substances (Augustine, Descartes, Swinburne), a subsistent and a nonsubsistent constituent (Aquinas, Stump, Moreland), and recently, a subsistent person emergent from a physical organism (Hasker, O'Connor) (See Goetz and Taliaferro 2011).

Current Christian dualists strive to avoid or resolve the exegetical, theological, and philosophical problems of historic dualisms. Current versions of dualism ought to be evaluated on their own terms and not anachronistically blamed for the (alleged) sins of their ancestors.

Green and Wright both attempt to discredit dualism with anachronistic criticisms, arbitrary definitions, caricatures, and fallacious arguments. A few examples must suffice.

In *Body, Soul, and Human Life* Green initially distinguishes "radical dualism," which devalues the body and makes the soul a separate and independent entity, from "wholistic dualism," which emphasizes the functional unity and integrity of human life and, he acknowledges, is widely held by current dualists (Green 2008, 10, 31). But throughout the rest of this book and his other publications, he ignores this distinction and simply criticizes "dualism" or "body-soul dualism" as though it were radical dualism. He faults "dualism" for producing otherworldly Christians interested in the soul and heaven but not the body and earth, and he even blames a caricature of Descartes's substance dualism for the isolated egoism that pervades modern society (Green 2008, 48–51, 70, 107). Many other instances of Green's equivocal and fallacious anti-dualist rhetoric can be cited. He offers no critique of "wholistic dualism" except his unsupported allegation that it leads to radical dualism.

Wright's view is puzzling. In *Resurrection* he attributes non-Platonic body-soul dualism to the Pharisees and Rabbis and identifies New Testament eschatology as a species of the Pharisees' teaching (Wright 2003a, 200, 477). But he rejects dualism of any sort in a subsequent address to Christian philosophers: "we do not need what has been called 'dualism' to help us over the awkward gap between bodily death and bodily resurrection" (Wright 2011). His reasons include the multiple meanings and problematic connotations of the term *dualism*; its complicity in spiritualism, otherworldliness, and consequent secularization of modern life, as well as other unbiblical separations, tensions, and reductions in

human existence; the incompatibility of Platonic, Cartesian, and Kantian anthropology with the biblical view, and dualism's being "too thin and flat" to account for the rich diversity of human life. He rejects dualism categorically.

But Wright does not consider the evidence of past or present dualistic theologians and philosophers who avoid these problems and present holistic accounts of human nature.<sup>2</sup> In addition, dualism need not conflict with holism or reduce the rich diversity of existence to just two categories. The unity of an entity's diverse parts and functions and the number of its metaphysical constituents are distinct issues. The mere idea that human beings are material and immaterial does not imply or promote any of the defects that Wright alleges. He deserves the criticisms he has received (Rickabaugh 2012; Goetz 2012).

*Monism* is the generic view that humans consist of one basic part or ingredient that constitutes all other parts and aspects so that they cannot exist without it. It comes in several religious and philosophical forms and is far more diverse than dualism. In major Asian religions, spirit or consciousness is basic, embodiment is derivative, and existence beyond death is therefore possible. In Greek naturalism and materialism, the body or matter is basic to mind and spirit and an afterlife is impossible. (However, some primal peoples who do not affirm an afterlife are animistic dualists who believe that souls are mortal.) *Metaphysical monism* comes in four varieties, each with a different basic principle: immaterialism (e.g., idealism, spiritualism, personalism), materialism (e.g., physicalism, emergentism), psychophysicalism (the principle is both), and neutral monism (the principle is neither but generates both).

Green and most current biblical monists endorse emergent physicalism, nonreductive physicalism, neutral monism, and/or psychophysical monism. I refer to them collectively as *bodily monists* because they all hold that embodiment is metaphysically essential and disembodied existence is impossible. They must derive both claims from Scripture as a whole to warrant biblical monism. Green endorses several kinds of bodily monism (Green 2008, 31).

*Holism* (from Greek for *whole*) is the view that humans are single beings whose diverse parts, dimensions, and operations are integrated, interactive, and interdependent within a basic unity. A human is an existential-functional whole, not the conjunction of independent, externally related parts – like an organism rather than a machine. Holism does not claim that humans are simple wholes without distinct parts and functions, as holist rhetoric sometimes suggests. More important, holism does not imply monism because it does not claim that humans consist of one metaphysical constituent or preclude a part surviving dissolution of the whole.

Thus holism should not be confused with monism or counted as evidence for it – mistakes commonly made by Green and many biblical monists. Once the difference between holism and monism is clear, the alleged biblical evidence for monism evaporates. All of Green's sound biblical, scientific, and common-sense evidence is consistent with and often better explained by holistic dualism.<sup>3</sup>

Green, Wright, and I do not dispute that biblical anthropology is holistic. Whereas Green mistakenly assumes that holism implies monism, Wright offers *ontological* holism as the biblical and philosophical alternative to dualism and materialism alike. He argues that Scripture uses terms such as *body* and *soul* to connote "the whole human being seen from one angle," "many aspects, one single reality" (Wright 2011, sec. 2). He identifies this anthropology as holistic, labels it *ontological differentiated unity*, and commends it to philosophers. However, he also affirms that death divides and separates. "Paul does indeed

envisage the possibility of a bodiless intermediate state in which one will be ‘naked’ (2 Cor. 5:3)” (Wright 2011, sec. 1) Thus Wright thinks that ontological holism is compatible with a dichotomy of persons and bodies at death. But if dichotomy entails dualism, then he concedes my claim that biblical anthropology is holistic and dualistic. Holism only conflicts with Wright’s pejorative definitions of dualism. Most current dualists affirm with Wright that humans are ontological unities and that disembodied existence is possible only by an act of God (for example, Goetz and Taliaferro 2011 promote “integral dualism”). Thus the dispute seems largely a problem of Wright’s terminology.

## 27.3 The Anthropology of the Bible

To make the biblical case for dualistic holism, I consider God’s creation of the man in Genesis 2:7, other Old Testament anthropological terms, and its references to the dead. I then survey visions of the afterlife in Second-Temple Judaism and conclude with the New Testament’s anthropological terms and texts about death and the life to come. Along the way I show that monism does not make its case.

### 27.3.1 The Old Testament

#### 27.3.1.1 Genesis 2:7

Genesis 2:7: “then the Lord God formed man from the dust of the ground and breathed into his nostrils the breath of life [*neshamah*]; and the man became a living being [*nephesh hayah*]” (NRSV) or “living soul.” (KJV)

I agree with Green that the terms in this text are not philosophical and that traditional dualists often mistakenly interpreted them as a substantial soul and material body.

But Green’s case for bodily monism fails. First, his main argument is a false dilemma: since Genesis 1 and 2 do not teach substance dualism, they imply bodily monism (Green 2008, 61–65). But the text is consistent with several other philosophical options. The *non sequitur* is glaring.

Second, Green translates God’s breathing *neshamah* into the man merely as a metaphor for God making the dust live: “the human being became fully alive.” (p. 64) He treats the earthly body as real but *neshamah* as a metaphor even though it is a permanent component of human nature (like its synonym *ruach*) in the Old Testament. Thus he reads bodily monism into Genesis 2:7, which is an anachronism as well as a genre mistake. The text actually reflects ancient Near Eastern animism, which holds that life is a power not contained in the dust of the earth. Animism regards the wind, vital powers, human capacities, paranormal events, and spiritual manifestations as the effects of real unseen powers and beings – gods, spirits, and ghosts. The Old Testament demythologizes animistic beliefs in terms of the true God, but it does not de-reify *neshamah* and *ruach* or reduce them to metaphors for bodily monism.<sup>4</sup> In Genesis 2 *neshamah* and dirt are two basic ingredients from which God makes a *nephesh chayah*.

This text therefore expresses generic dualism – two ingredients in one being. If *neshamah* is an individuated power source, as in some kinds of animism, then Genesis 2:7 implies a dualism of a body formed of earth and an intrinsic, God-given source of life,

consciousness, and agency. This account might imply less than substance dualism, but it certainly rules out monism (one ingredient).

Green's appeal to Aristotle to align Genesis 2:7 with bodily monism is likewise misconceived.<sup>5</sup> Aristotle is neither a two-substance dualist, like Plato, nor a bodily monist. His anthropology is *hylomorphic* – dual constituent holism. Natural things are *primary substances* constituted by two metaphysical principles – *form* and *matter*. Form is a metaphysically distinct and active principle, not a function of matter. In humans it is the rational animal soul which not only structures and activates matter to form a living sensing organism, but also has rational and spiritual capacities not dependent on psychophysical processes. The rational soul continues in the universal intellect after death even though it is no longer individuated.<sup>6</sup> Aristotle's anthropology is similar to Genesis 2:7 in some ways, but it is essentially different than bodily monism.

### 27.3.1.2 Anthropological terminology

Genesis 2:7 fits with other Old Testament anthropological terms. All scholars agree that words such as *nephesh*, *ruach*, *neshamah*, *leb* (heart), and *basar* (flesh) have multiple meanings that sometimes overlap and sometimes contrast with each other. *Nephesh* and *leb*, for example, have life functions, refer to body parts (throat, heart), engage in intentional acts, and refer to whole persons (“my *nephesh*, *ruach*, or *leb* cries out” means “I cry out”). Terms are sometimes literal and sometimes figurative, as when a part stands for a person (*synecdoche*). Collectively, these terms generate the impression of humans as integral psychophysical unities with diverse parts and powers. Thus Old Testament anthropology is generically holistic, although it does not clearly map the parts and their relations. But it is not monistic because it does not suggest that humans consist of one basic ingredient or that death is the end of existence.

Whether Old Testament anthropology is dualistic depends on whether humans “come apart” at death and subsist without bodies (not whether *nephesh* and *ruach* refer to the dead). If so, then terms that refer to this condition have dualistic (as well as holistic) meanings.

### 27.3.1.3 The afterlife in the Old Testament

The decisive issue is whether persons subsist after death. By *person* I mean the proper referent of someone's name, whatever his/her metaphysical nature. A ghostly replica or memory or impersonal part of Samuel is not Samuel. But an existentially functionally reduced Samuel is still Samuel after death, just as a comatose Samuel is still Samuel in life. Samuel is more than his body. By *subsistence* I do not necessarily mean more than sheer inactive unconscious existence to avoid objections that *existence* and *survival* are too generous to describe the dead.<sup>7</sup> Mere subsistence involves a dichotomy of the living person, which entails dualism. The Old Testament affirms at least subsistence, and this view is expanded in Judaism and the New Testament (Wright 2003a, ch. 3; Bernstein 1996, ch. 5; Segal 2004, 142–145; Walton 2006, ch. 14).

Critics raise irrelevant issues to avoid dualism. Green objects that the Old Testament says little about the afterlife; that the word *soul* (*nephesh*) is not used for the dead; that commentators have imposed Platonic or “essentialist” meanings on *nephesh* and *ruach*; that *Sheol* (the place of the dead) is merely a poetic metaphor for the grave; that what follows death is so minimal that it amounts to nonexistence; that the dead are completely cut off

from God; and that because the Old Testament is monistic, it does not envision existence after death (Green 2008, 152–157). But most of these objections do not even challenge the claim that persons subsist after death, much less refute it, and none reflects the historical and textual evidence.

Consider the historical evidence.<sup>8</sup> Belief in an afterlife (perhaps only for the elite), usually in the underworld, was widespread around the Mediterranean from Greece to Egypt. Views of the dead ranged from ghosts – ethereal bodily beings – to powerful spirits and demi-gods who can affect the living. The Egyptians embalmed bodies to be reunited with souls after judgment – a two-stage eschatology. Some Greeks believed in disembodied immortality. But perhaps the Old Testament is quite different.

So consider the textual evidence. The Old Testament addresses the afterlife sufficiently to form a general idea. Sometimes a reference lacks clarity, as when Psalm 23 anticipates dwelling in the house of the Lord forever. Sometimes it is specific, as when Daniel 12:2 predicts the resurrection of all the dead to glory or judgment. The place of the dead is *Sheol* or *Abaddon*, the underworld and not just a metaphor for the grave. Isaiah 14's prophetic depiction of Nebuchadnezzar in *Sheol* among other dead kings is not pious fiction but eschatological realism. Psalm 139:8 affirms that God himself is in *Sheol* with the faithful dead (contra Green). The dead may be weak, inactive, and unable to praise God, but they subsist and retain their personal identity.

Consorting with the dead corroborates this belief. The clearest case is Saul asking a pagan medium to consult deceased Samuel (1 Samuel 28). Samuel appears to her and converses with Saul, prophesying that he and his sons would be killed and join him the next day. This text is historical narrative, not apocalyptic imagery. Samuel was buried at Ramah but appeared at Endor, far from his corpse. He appears as a ghost wearing a robe.

The Old Testament refers to the dead in several ways – primarily by their names (Samuel, David) and personal pronouns (I, you, those in *Sheol*), but also *rephaim* (*gigantes*, giants, in Septuagint), *elohim* (the medium called Samuel *a god*), and, yes, occasionally by *nephesh* (self, life, soul? Ps. 16:10, 49:15). There is no clear concept of what remains of a person. A plausible account is that a *nephesh chayah* loses its *ruach*, flesh, blood, and bones but remains the same person. Old Testament eschatology, minimal though it is, envisions subsistent residual humans. This fact is sufficient to corroborate dualism and disconfirm monism. Note that dualism would follow even if *nephesh* and *ruach* were never translated as *soul* and *spirit*, never referred to the dead, and were never interpreted “essentially.” Criticisms of traditional word studies are finally irrelevant.<sup>9</sup>

Coupling postmortem subsistence of the dead with the few references to future resurrection yields the outline of the two-stage eschatology of the Pharisees, Rabbis, and New Testament. Daniel 12:2 refers to a single universal resurrection to judgment or glory, and Ezekiel 37 envisions the resurrection of the faithful dead (not just future generations) by God's reconstructing their bodies and reanimating them with *ruach*, an echo of Genesis 2:7. Although Ezekiel and Daniel do not mention the dead in *Sheol*, Isaiah 26:19 presents both stages. The *rephaim* in *Sheol* will rise, and their corpses will live again.<sup>10</sup>

Wright summarizes three affirmations of Old Testament eschatology: the dead are asleep with the ancestors; they will be received by the Lord into some kind of continuing life; and (some of) the dead can hope for resurrection “*after any such 'life after death'*” (Wright 2003a, 124). These teachings together imply the outline of the Pharisees' and New Testament's two-stage eschatology with its correlative dualistic anthropology.



### 27.3.2 *The New Testament*

To show that New Testament anthropology is holistic and dualistic, I briefly consider the eschatologies of Judaism engaged by Jesus and Paul, New Testament anthropological terms, and New Testament teaching about the life to come.

#### 27.3.2.1 Anthropology and eschatology in Judaism

It is common in the monism–dualism debate to summarize the views of the afterlife in Second Temple Judaism as background to New Testament eschatology. They fall into three main types – no afterlife, disembodied immortality, and temporary disembodiment between death and resurrection – represented by the Sadducees, Philo, and the Pharisees and Rabbis respectively. This information facilitates interpretation of the New Testament and evaluation of monist and dualist claims. I rely on N. T. Wright’s *The Resurrection of the Son of God*, the most detailed and comprehensive study to date (Wright 2003a, ch. 4).

One significant view – held by the Sadducees – reflects Ben Sirach and amounts to denial of an afterlife altogether, not merely denial of resurrection (Matt. 22:23; Mark 12:18–27; Acts 23:6–8; Wright 2003a, 131–140). The dead fade to nothing in *Sheol*, which is a metaphor for nonexistence in this account.

Disembodied immortality of the soul is another perspective (Wright 2003a, 140–146). It is expressed in 4 Maccabees, where “the martyrs go, immediately upon death, into the blissful immortality already enjoyed by Abraham, Isaac, and Jacob” (143). It is also found in Pseudo-Phocylides, Testament of Abraham, Ethiopic book of Enoch, Jubilees, Wisdom of Solomon, and most famously by Philo, who was trained in Platonic philosophy. Its anthropology is unambiguously dualistic. The soul is immortal and will be eternally disembodied. But Judaism affirms the goodness of God’s creation, so the body is not anti-spiritual, as in Platonism and Gnosticism, but transcended by the greater good of the soul’s presence to God. These texts consistently use the Hebrew and Greek terms for *soul* and *spirit* dualistically, as do the Septuagint and commentaries on it. This practice long predates Christian Platonists and medieval scholastics.

A third position was the popular alternative to the Sadducees and the Jewish Platonists. It emphasizes the created goodness of the body and expects bodily resurrection. But it also affirms that God sustains people between death and the final resurrection (pp. 146–200). According to Wright, “any Jew who believed in resurrection, from Daniel to the Pharisees and beyond, naturally believed also in an intermediate state in which some kind of personal identity was guaranteed between physical death and the physical re-embodiment of resurrection. This too is a form of ‘immortality.’” Details of the resurrection and intermediate state varied, but a generic “two-stage” eschatology “was widely believed by most Jews around the turn of the common era” (Wright 147). Wright documents it in the Septuagint, 2 Maccabees, 1 Enoch, Apocalypse of Moses, Sibylline Oracles, 4 Ezra, 2 Baruch, Wisdom of Solomon, Josephus, the Essenes, Pseudo-Philo’s Biblical Antiquities, and – most significant for the New Testament – the Pharisees and Rabbis.

A couple of observations are in order. First, simplistic alternatives between Jewish monism and Greek dualism are false. Much Judaism is dualistic but not Platonic or Gnostic – not divinizing the soul and degrading the body.<sup>11</sup> Even Philo affirms the body’s goodness. Thus most Jewish dualism is *holistic* – it affirms the natural goodness and resurrection of the body without denying the distinctness and separation of the soul. This sort of *holistic* dualism is affirmed by the New Testament and the best of the Christian

tradition. When monist scholars like Green (1998, 159–163) overlook Judaism’s two-stage eschatology and holistic-dualistic anthropology and substitute radical caricatures to discredit biblical dualism, they push against a lot of hard evidence.

Second, there is no basis in Second-Temple Judaism for the eschatologies of current bodily monists. They must either posit a bodily resurrection immediately at death (which implies replication, not personal identity) or a period of nonexistence between death and the general resurrection. Both views are inconsistent with the New Testament, as we shall see, and neither has precedence in Judaism. All Jewish accounts of bodily resurrection are communal and future. The only reference to immediate resurrection is a metaphor for the immortal soul’s ascent to God at death, which is Jewish Platonism. Temporary nonexistence is likewise without precedent. The Sadducees took Old Testament references to *Sheol* as permanent extinction. All Jewish affirmations of bodily resurrection assume interim existence. Lack of support in Judaism raises suspicion about bodily monism even before we consider the New Testament.

### 27.3.2.2 New Testament anthropological terms

The anthropological terms of the Old Testament were modified in most branches of Judaism to preserve holism and explicate dualism. For example, *nephesh* and *ruach* and their Greek translations, *psyche* and *pneuma*, refer both to living and dead humans – either permanently disembodied or awaiting resurrection – in texts and tomb inscriptions. The New Testament appropriates this vocabulary.

A few examples of New Testament holist and dualist uses must suffice. A well-known example of holism is 1 Thessalonians 5:23, Paul’s benediction: “you entirely . . . your spirit, soul, and body.” It is holistic because Paul blesses persons wholly by emphasizing significant parts. It is not monism, however, because it does not imply that the whole consists of one ingredient or that one part is basic. It is also consistent with trichotomism – three distinct components – and with dualism – body and soul/spirit – but it does not imply that persons consist of two or three components. It does not imply a metaphysical constitution at all. Another likely example of holism is the Song of Mary (Luke 1: 46–47): “My soul magnifies the Lord, and my spirit rejoices in God my Savior.” *Soul* and *spirit* may refer to distinct nonphysical parts of Mary but more likely are parallel terms for Mary’s whole being – Mary herself. Similarly, most uses of *heart*, *mind*, *will*, *flesh*, and *body* are holistic – referring to integral parts of living humans – without implying monism or dualism.

But there are also dualistic instances of *soul* and *spirit* in eschatological texts virtually identical to those in Judaism. Revelation 6:9–11 depicts the *souls* (*psuchai*) of martyrs pleading for God’s justice while awaiting the resurrection envisioned in Revelation 20. They are wearing white robes but are separate from their martyred bodies (Wright 2003a, 174, 471). Matthew 10:28 almost certainly expresses dualist Jewish eschatology. Jesus warns: “Do not fear those who kill the body but cannot kill the soul; rather fear him who can destroy both body and soul in hell.” If the text represented bodily monism, killing the body would *ipso facto* kill the soul.

*Spirit* (*pneuma*) is used dualistically in Hebrews 12:23, where the spirits of the righteous dead are with God and the angels in the heavenly Jerusalem awaiting resurrection (like the Maccabees alluded to in Hebrews 11:35). Another instance of *spirit* crucial for the monism–dualism debate is Paul’s affirmation of the Pharisees’ eschatology in Acts 23:6–8. Luke explains: “The Sadducees say that there is no resurrection, or angel, or spirit; but the

Pharisees acknowledge all three.” *Spirit* almost certainly refers to deceased humans awaiting resurrection (Wright 2003a, 131–134, 454).<sup>12</sup> Jesus himself *gave up his spirit* when he died according to all four Gospels. Matthew, Mark, and Luke use the verb *exepneusen*; Luke and John use the noun *pneuma* (Matt. 27:50; Mark 15:39; Luke 23:46; John 19:28). The verb does mean “breathed his last,” as monists argue, but also “gave up his spirit.” In the eschatology of Judaism and the Gospels, *spirit* is both the “breath of life” and the person who survives death.

In sum, there is no way to square the meaning of *soul* and *spirit* in these and other New Testament texts with anthropological monism. Green expounds at length on biblical word studies, but he avoids the clearly dualistic texts (Green 208, 54–60).

### 27.3.2.3 New Testament eschatology

Word studies have already considered some eschatological texts. Space allows only a representative summary of the whole New Testament, which according to Wright “speaks, if not with one voice, certainly with a cluster of voices singing in close harmony.” Their hymn is the two-stage eschatology of temporary disembodiment followed by bodily resurrection. New Testament eschatology represents “a united subbranch of Pharisaic Judaism” with two modifications: Jesus was raised first and believers will follow at his return; and the resurrection body is the earthly body transformed by God’s Spirit and not merely revived (Wright 2003a, 476, 477).<sup>13</sup>

To establish dualism I focus on a few key texts. I begin with Paul’s affirmation of Pharisee eschatology in Acts 23, link it to Luke’s account of Jesus’s death and resurrection, and then to Paul’s statements about resurrection and the interim in 1 and 2 Corinthians and Philippians 1. Green does not adequately explain these texts or consider their mutual implications.

Pharisee eschatology is well-documented, and its endorsement by Paul in Acts 23:6–8 is explicit and historically credible. This text is *prima facie* evidence that Paul affirms a two-stage eschatology, which requires person–body dualism. That fact should guide our approach to Paul’s letters and to Luke–Acts unless there is sufficient evidence otherwise. Green mentions Pharisee eschatology but it does not inform his exegesis.

Luke is familiar with Jewish two-stage eschatology (Wright 2003a, 435–439), as is evident in Jesus’s parable of the rich man and Lazarus in 16:19–31. Lazarus is with Abraham (in Paradise), but the rich man is in Hades. The key point is that the text is dualistic whether it depicts the final state, as Green argues, or the intermediate state, as most commentators hold.<sup>14</sup> Either way, the deceased men presumably left corpses on earth even though they appear in bodily form. If monism were true, they would have been transported bodily, like Enoch and Elijah. But the parable does not teach eschatology or anthropology, so I do not infer Luke’s doctrine from it.

A doctrinally decisive text is Luke 23:43, where Jesus assures the thief on the cross, “today you shall be with me in Paradise.” “Today” refers to the day of their crucifixion, not an *eschaton* beyond time, as Green claims (Green 2008, 163–165).<sup>15</sup> Jesus died on Friday and rose on Sunday, and he existed apart from his body during the interim. If personal existence depends on one’s organism, as bodily monism entails, then Jesus could not have existed, which would falsify his promise. And if Jesus’s resurrection occurred immediately at his death, as other monists claim, then Easter was meaningless, and when he died on Friday he either switched bodies, which entails dualism, or one body replaced the other body, which

implies two different persons. There is no alternative to dualism. Thus Luke's Gospel affirms the same Pharisaic eschatology and anthropology as his account of Paul in Acts 23. Green uses the diversity of Jewish eschatology to marginalize this reading of Luke 23:43, but he never considers Luke's approval of Paul or explains why it is irrelevant to interpreting Luke, Acts, and Paul's letters.

Paul's two-stage eschatology and (holistic) dualism are also evident from his letters, which presumably present a consistent eschatology unless they are shown not to.

There is no question that Paul affirms bodily resurrection at the future return of Christ (1 Thess. 4:16–17; 1 Cor. 15:52), so I merely note that he rejects immediate resurrection.

With respect to death, Paul contrasts living "at home in the body and away from the Lord" with being "away from the body and at home with the Lord" in 2 Corinthians 5:6–9. The only option here is being with or without his body, which entails dualism. There is no hint of a resurrection body in these verses. 2 Corinthians 12:2–4 also implies dualism. Paul recounts a spiritual experience of Paradise in heaven but does not know whether it was "in the body or out of the body." The mere possibility of disembodiment assumes dualism and precludes bodily monism. Philippians 1:20–24 poses the same option as 2 Corinthians 5 using flesh (*sarx*) instead of body (*soma*): "to live in the flesh," which is "to remain in the flesh," or "to depart [from the flesh] and be with Christ." These assertions are obviously dualist unless monism has a better explanation.<sup>16</sup>

Green and most monists read the tent/house/dwelling/clothing metaphors in the verses in 2 Corinthians 5:1–5 as referring to two kinds of embodiment – earthly and resurrection – without disembodied existence between them. But this exegesis is flawed. "Heavenly dwelling" in Paul's theology is not a heavenly spiritual resurrection body but "the place where the divinely intended future for the world is kept safe" (Wright 2003a, 365, 368). In addition, sound exegetical method interprets metaphors in terms of available clear assertions, not the reverse. The straightforward prose of verses 6–8 speaks of "the body," not "this body" in contrast to a resurrection body. Thus 1–5 should likewise be interpreted as contrasting "in the body" with "apart from the body." But monists first read two bodies into the metaphors of verses 1–5 and then struggle for an encore in verses 6–10. Green simply ignores verses 6–9, which makes it a lot easier to maintain monism (Green 1998, 169–172; 2008, 170–172, 174–178).

Paul's dualism is evident even though he does not use the terms *soul* or *spirit* for the dead. Instead he refers to people and uses personal pronouns just as he does for the living: "a man," "I," "we" are "in the body or apart from the body." His language might suggest person-body dualism instead of soul/spirit-body dualism, but that distinction is too technical. Both *soul* and *spirit* can mean "I," "self," or "person," in biblical semantics, so these are actually synonymous anthropologies. Green and Wright both argue that Paul's not using *soul* and *spirit* for the dead counts against dualism, but it does not. Biblical anthropology would imply dualism even if *soul* and *spirit* did not occur in Scripture. Another relevant Pauline duality is found in 2 Corinthians 4:16 – "the outer man (*anthropos*) who is being destroyed" and "the inner man who is being renewed." These phrases are correlates or synonyms of the mortal body and separable "I" which follow immediately in 5:1–10.

In sum, 2 Corinthians 5:1–10 and Philippians 1:20–24 taken alone are compatible with either permanent disembodiment or future resurrection – both of which are dualistic. But combined with 1 Corinthians 15 and 1 Thessalonians 4, they present a two-stage eschatology of being with the Lord apart from the body until the general bodily resurrection

at his return. Paul's anthropology is both holistic and dualistic, much like the Pharisees. He nowhere implies monism – that humans consist of one ingredient or substance – and his eschatology rules it out.

Green's claim that Paul is a monist is entirely unwarranted. He not only fails to provide a tenable case for monism but also avoids crucial data and standard methods that imply dualism. Wright's elaboration of Paul's eschatology is sound, but his claim that Paul is not a dualist stems from his narrow and negative definition of dualism, explained above. He affirms that persons exist apart from their bodies, which entails dualism as defined in this volume.

In closing I merely note some other texts that refer to or imply personal existence between death and resurrection. The Patriarchs are currently "alive" (Luke 20:38). Moses and Elijah participated in Jesus's transfiguration (Matt. 17:1–13; Mark 9:2–13; Luke 9:28–36). In John's Gospel, regeneration by the Holy Spirit transforms created life (*zoe*) into everlasting life (*zoe aionios*) already now, and *everlasting life* precludes temporary nonexistence (John 1:4, 3:3–16, 5:24–29, 11:25–26). Paul's assurance that death and the future cannot separate us from God's love (Rom. 8:38–39) rules out nonexistence because God's love for his people is covenantal and reciprocal, not temporarily unrequited. Most traditions understand the 144,000 in Revelation 7 to symbolize the church in heaven praising God while they await the final resurrection in Revelation 20. Bodily monism cannot account for these texts.

## 27.4 Conclusion

Biblical anthropology is demonstrably both holist and dualist. It is holist in teaching that God created, redeems, and will glorify humans as whole embodied persons. It is dualist in teaching that God created humans of two ingredients and that he sustains persons (souls, spirits) apart from their bodies between death and resurrection.

Wright corroborates the exegesis supporting these conclusions but defines and rejects dualism as incompatible with holism. Nevertheless, his understanding of holism is compatible with virtually all current Christian dualism.

Green fails to make the case that biblical anthropology is monistic and anti-dualistic. He does not show from Scripture that humans are constituted of one ingredient and that there is no intermediate state. He does not consider all the relevant texts and background data, and thus he does not present a full biblical anthropology. He confuses monism with holism, caricatures dualism, and presents sufficient irrelevant and fallacious arguments to undermine his case. However, his holistic exegesis, integral anthropology, and practical theology make valuable contributions that dualists can affirm.

## Notes

1. "Science already informs exegesis; it is only a question of which science or whose, good science or bad" (Green 2008, 21; see also 21–28); "The two, science and theology, interact in a more organic way than we often acknowledge, with the results that it is virtually impossible to extricate the one influence from the other. This is true of the 'science' presumed of the biblical writers and of the 'science' presumed of biblical interpreters and theologians from the second century onward" (Green 2005, 15).
2. I mention Bavinck and Kuyper in my Dutch Reformed tradition. They are integral dualists who strongly affirm the body and life in the world. Consider also John Wesley, Martin Luther King, Mother Theresa, and

- John Paul II. Many theologians, clergy, and mere Christians in all traditions are holistic dualists in belief and practice.
3. For example, Green points to the scholarly consensus that the Old Testament presents “a unified portrait of the human person” as evidence for biblical monism and scholarly endorsement of it (Green 2008, 8). He infers monism from the fact the whole person images God, not just the soul (pp. 61–65). He refers to the neural correlates and biological substrate of spiritual experience as evidence for monism (p. 108), although they are fully compatible with dualism. Tellingly, he appeals to neuroscientist Mario Beauregard, a dualist and author, with Denyse O’Leary, of *The Spiritual Brain: A Neuroscientist’s Case for the Existence of the Soul* (2007).
  4. Modern translations of *ruach* exclusively as *wind* or *breath* are naturalistic reductions of their broad Old Testament semantic range (Walton 2006, ch. 9).
  5. Green works from *De Anima* without attending to hylomorphism or the distinctness of the human soul (Green 2008, 55–56; 2005, 19).
  6. On Aristotle’s dualism see Robinson (1983) and Heinaman (1990).
  7. I do not affirm mere subsistence because of texts like Isaiah 14 that suggest some consciousness and activity.
  8. The Egyptians were trichotomists who believed in an interim state until resurrection of mummified bodies. The Hebrews knew Egyptian beliefs (see Wright 2003a, ch. 2; Taylor 2001).
  9. Green (2008, 54–57). Not all dualists make the errors Green alleges, but it makes no difference if they did.
  10. Wright (2003a, 117) concludes that Isaiah 26 affirms an interim existence and bodily resurrection of individuals and the nation.
  11. Wright elaborates: “the teachers [Pharisees] seem to have developed, in line with other writings that affirm ultimate resurrection, various ways of talking about an intermediate state, which, though they may sometimes use the language of the ‘soul,’ seem more or less innocent of any developed Platonic idea of the soul as an (or the) immortal element of all human beings” (Wright 2003a, 200).
  12. Wright elaborates on the Pharisees: “these great sages [Hillel and Johanan ben Zakkai] believed in eventual resurrection; here they seem to have been employing new concepts of a body/soul dualism to explain what happened between bodily death the final state of blessedness . . . The idea of a soul separable from the body . . . was widespread in the varied Judaism of the turn of the eras” (142); “the teachers seem to have developed, in line with other writings that affirm ultimate resurrection, various ways of talking about an intermediate state, which, though they may sometimes use the language of the ‘soul,’ seem more or less innocent of any developed Platonic idea of the soul as an (or the) immortal element of all human beings. Souls . . . are not the same kind of thing as the pre-existent beings of Plato’s *Phaedrus* and elsewhere . . . When the New Testament writers spoke of resurrection, both their own and that of Jesus, this is the grid of language-use within which they must have assumed their words made sense” (Wright 2003a, 200).
  13. Wright (2003a) surveys Paul in chs 5–8, the Gospels in ch. 9, and the other New Testament writings in ch. 10.
  14. Green (2008, 159–163) concedes a possible intermediate state but argues for immediate realization of the final state. Wright observes: “it is impossible to say whether it belongs with the ‘resurrection’ strand in Second-Temple Judaism, or with a ‘disembodied immortality’ strand” (2003a, 438). Both options are dualistic.
  15. Wright (2003a, 438nn.114–115) points out that Paradise is a temporary resting place as well as a permanent destination.
  16. On 2 Cor. 5:6–10 and Phil. 1:20–24: “this is as close as Paul ever comes to an account of the intermediate state between death and resurrection” (Wright 2003a, 369).

## References

- Beauregard, M., and Denyse O’Leary. 2007. *The Spiritual Brain: A Neuroscientist’s Case for the Existence of the Soul*. New York: Harper One.
- Bernstein, Alan. 1996. *The Formation of Hell: Death and Retribution in the Ancient and Early Christian Worlds*. Ithaca, NY: Cornell University Press.
- Cooper, John. 2000 [1989]. *Body, Soul and Life Everlasting: Biblical Anthropology and the Monism-Dualism Debate*. Grand Rapids, MI: Eerdmans.
- Cooper, John. 2001. “Biblical Anthropology and the Body-Soul Problem.” In *Soul, Body and Survival: Essays on the Metaphysics of Human Persons*, edited by Kevin Corcoran, 218–228. Ithaca: Cornell University Press.

- Cooper, John. 2007. "The Bible and Dualism Once Again: A Reply to Joel B. Green and Nancey Murphy." *Philosophia Christi*, 9(2): 459–469.
- Cooper, John. 2009a. "The Current Body-Soul Debate: A Case for Dualistic Holism." *Southern Baptist Journal of Theology*, 13(2): 32–50.
- Cooper, John. 2009b. "Exaggerated Rumors of Dualism's Demise: A Review Essay on Body, Soul, and Human Life." *Philosophia Christi*, 11(2): 453–464.
- Cooper, John. 2015. "Scripture and Philosophy on the Unity of Body and Soul: An Integrative Method for Theological Anthropology." In *The Ashgate Research Companion to Theological Anthropology*, edited by Joshua Farris and Charles Taliaferro, 27–42. Farnham, UK: Ashgate.
- Goetz, Stewart. 2012. "Is N. T. Wright Right about Substance Dualism?" *Philosophia Christi*, 14(1): 183–192.
- Goetz, Stewart, and Charles Taliaferro. 2011. *A Brief History of the Soul*. Oxford: Wiley-Blackwell.
- Green, Joel. 1998. "'Bodies – That Is, Humans Lives:' A Re-Examination of Human Nature in the Bible." In *Whatever Happened to the Soul? Scientific and Theological Portraits of Human Nature*, edited by Warren S. Brown, Nancey Murphy, and H. Newton Malony, 149–173. Minneapolis, MN: Fortress Press.
- Green, Joel. 2002. "Eschatology and the Nature of Humans: A Reconsideration of the Pertinent Biblical Evidence." *Science and Christian Belief*, 14(1): 33–50.
- Green, Joel. 2005. "Body and Soul, Mind and Brain: Critical Issues." In *In Search of the Soul: Four Views of the Mind-Body Problem*, edited by Joel B. Green and Stuart L. Palmer, 7–32. Downers Grove, IL: InterVarsity Press.
- Green, Joel. 2008. *Body, Soul, and Human Life: The Nature of Humanity in the Bible*. Grand Rapids, MI: Baker Academic.
- Heinaman, Robert. 1990. "Aristotle and the Mind-Body Problem." *Phronesis*, 35(1): 83–102.
- Rickabaugh, Brandon. 2012. "N. T. Wright and the Ontology of Human Persons." Paper presented at Society of Vineyard Scholars Conference, Biola University, April 19–21.
- Robinson, Howard. 1983. "Aristotelian Dualism." In *Oxford Studies in Ancient Philosophy*, edited by Julia Annas, vol. 1, 123–144. Oxford: Oxford University Press.
- Segal, Alan. 2004. *Life after Death: A History of the Afterlife in Western Religion*. New York: Doubleday.
- Taylor, John. 2001. *Death and the Afterlife in Ancient Egypt*. Chicago: University of Chicago Press.
- Walton, John. 2006. *Ancient Near Eastern Thought and the Old Testament*. Grand Rapids, MI: Baker Academic.
- Wright, N. T. 1992. *The New Testament and the People of God*. Minneapolis, MN: Fortress.
- Wright, N. T. 2003a. *The Resurrection of the Son of God*. Minneapolis, MN: Fortress Press.
- Wright, N. T. 2003b. "Rethinking the Tradition." In *For All the Saints? Remembering the Christian Departed*, 20–46. London: SPCK; Harrisburg, PA: Morehouse.
- Wright, N. T. 2011. "Mind, Spirit, Soul and Body: All for One and One for All Reflections on Paul's Anthropology in his Complex Contexts." Paper presented at Society of Christian Philosophers Regional Meeting, Fordham University, March 18. Accessed July 1, 2015. [http://ntwrightpage.com/Wright\\_SCP\\_MindSpiritSoulBody.htm](http://ntwrightpage.com/Wright_SCP_MindSpiritSoulBody.htm).

# The Strange Case of the Vanishing Soul

JOEL B. GREEN

At least at a popular level, the default assumption many readers have about the New Testament's anthropology takes the form of either a body-soul dualism or, reflecting the language of 1 Thessalonians 5:23, a body-soul-spirit trichotomy. That is, many, especially nonspecialists, imagine that the theological anthropology of the Christian Scriptures assumes and supports the view that humans are constituted by "parts," numbered as two or three. In this respect, N. T. Wright has observed:

We have been buying our mental furniture for so long in Plato's factory that we have come to take for granted a basic ontological contrast between "spirit" in the sense of something immaterial and "matter" in the sense of something material, solid, physical. (Wright 2008, 153–154)

Although Wright goes on to deny that such a view was characteristic of the dominant cosmologies of the first-century world of the apostles, this reference to our "mental furniture" helps to explain the ease with which moderns might find dualism in the Christian Bible, and especially in the New Testament.

Assumptions like these stand in sharp contrast with the position increasingly documented in New Testament translations of the last century or so, themselves a barometer of the positions New Testament scholars have taken with respect to biblical anthropology. In the Authorized Version of the New Testament, popularly known in the United States as the King James Version, published in 1611, the word "soul" appears 39 times, always as a translation of the Greek term *ψυχή* (*psychē*). According to Samuel Johnson's *A Dictionary of the English Language* of 1755, this English term referred to "the immaterial and immortal spirit of man," though Johnson notes that, in Shakespeare, who would have been writing about the time when the Authorized Version was being translated, "soul" referred to "vital principle," "interior power," or "essence" (Johnson 1785, pt. 2). With the publication of the American Standard Version of 1901, the translation of *ψυχή* (*psychē*) as "soul" had lost



some traction, with only 34 occurrences in the New Testament. Fifty years later, the Revised Standard Version (1952) provided only 27 appearances of the term “soul.” Published in 1989, the New Revised Standard Version maintained only 22 occurrences. Today’s New International Version, published in 2001, has 20; the 2011 edition of the New American Bible has 15; and the Common English Bible, also published in 2011, has three. In translations of the New Testament, the “soul” seems to be vanishing.

Generally, Bible translations today tend to use nonspecialist terms, so it is noteworthy that the decline in occurrences of the term “soul” has taken place during an era when the public meaning of the term has remained relatively static. According to contemporary Oxford dictionaries, “soul” is used first in the sense of “the spiritual or immaterial part of a human being or animal, regarded as immortal” (*Oxford Dictionaries*, s.v. “Soul.” <http://oxforddictionaries.com/definition/english/soul?q=soul>.) Apparently, New Testament translators have abandoned the term “soul” to popular usage, where it usually refers to an immaterial, immortal part of a human, while noting that the Greek term ψυχή (*psychē*) is not very well lexicalized in this way.

Those working in the philosophy of mind will have more specialized ways of thinking about the “soul,” using the term with reference to the *mind*, *personal identity*, and so on. It is therefore remarkable that New Testament translators have not moved in the direction of identifying Greek terms for “mind,” such as νοῦς (*nous*) or φρήν (*phrēn*), with the English term “soul.” Nor have they translated with the English term “soul” a host of other terms denoting what we today think of in terms of psychological capacities, such as σπλάγχνα (*splagchna*, “entrails,” then “seat of love or affection”) or καρδιά (*kardia*, “heart,” then “center of one’s inner life”).

In fact, as has often been noted in studies of New Testament anthropology, New Testament writers use a wide array of terms to discuss different capacities or aspects or functions, though not in discrete or denotative ways. In actual usage, καρδιά (*kardia*) can overlap with σπλάγχνα (*splagchna*), just as ψυχή (*psychē*) can overlap with νοῦς (*nous*). In fact, as Johannes Louw and Eugene Nida observe, in the hands of the New Testament writers, ψυχή (*psychē*) can refer to a “person” or “life” or “inner person” (Louw and Nida 1988, pt. 2–266); or as the standard Greek-English lexicon reports, in the hands of the New Testament writers, ψυχή (*psychē*) can refer to “life on earth in its animating aspect making bodily function possible” or the “seat and center of the inner human life in its many and varied aspects” or simply to a “person” (Bauer 2000, 1098–1100). Given these ways of understanding the term ψυχή (*psychē*), it may not be surprising that we have seen a remarkable decline in appearances of the word “soul” in English translations of the New Testament. Against this linguistic backdrop, perhaps the case of the vanishing soul is not so strange after all.

I should quickly add that the vanishing soul in New Testament translations is not the result of influence from the natural sciences in the modern era. For better or worse, New Testament scholars generally are known for their esoteric, in-house discussions, not for their forays into other areas of inquiry. In fact, the centuries-old hegemony of the historical-critical method in biblical studies has encouraged biblical scholars against bringing such modern insights to their study of these ancient texts. New Testament scholars want to grasp their objects of study from within the thought-world of the ancient Mediterranean world. Accordingly, the scholarly trajectory occupied by New Testament studies over the last century and more does not owe its direction or endpoint to pressure from the natural sciences.

In some ways, such siloization is unfortunate. After all, important shifts in one field are often indebted to insights from another. One has only to think of the enormous impact the social sciences have had in New Testament studies in the past three or four decades to get a sense of what I mean. Since the natural sciences, the neurosciences in particular, were moving along a parallel track in their understanding of the human person, we might wonder what an earlier collusion between the two fields might have produced.

How shall we explain the strange case of the vanishing soul? Let me briefly expand on three areas of inquiry – first, the significance of historical inquiry for situating the New Testament materials more securely within their first-century milieu as a prophylactic against colonizing New Testament perspectives, in this case, with foreign assumptions about theological anthropology; second, the significance of sociocultural forms of inquiry, including medical anthropology and social psychology, for shaping our understanding of humanity in the New Testament world; and third, a rereading of New Testament texts that served previously as taken-for-granted illustrations of the New Testament’s anthropological dualism.

## 28.1 The “Soul” and Historical Inquiry

First, then, we account for the significance of historical inquiry for situating the New Testament materials more securely within their first-century milieu as a prophylactic against colonizing New Testament perspectives with foreign assumptions about theological anthropology.

The transformation in thinking about New Testament anthropology is easy enough to document. At the end of the 1800s, the dominant views of Pauline anthropology, for example, were either that Paul imagined the human person in dichotomous (body-soul) terms or in trichotomous (body-soul-spirit) terms. This latter viewpoint is based on the famous text in 1 Thessalonians, of course, where Paul writes:

And the very God of peace sanctify you wholly; and I pray God your whole spirit and soul and body be preserved blameless unto the coming of our Lord Jesus Christ. (5:23, KJV)

Given this unambiguous phrasing, I find it surprising how easily our contemporaries outside of biblical studies who want nonetheless to work with Scripture continue to hold to some form of dualism, since it is not easy to see why one would allow the body-soul distinction to stand but collapse the spirit-soul distinction into a single category of things immaterial. What may be even more surprising, however, is Anthony Thiselton’s finding that, with few exceptions, the history of interpretation of 1 Thessalonians 5:23 has tended toward a wholistic, or monist, understanding of the text (Thiselton 2011, 161–75).

In any case, with regard to Pauline anthropology more generally, by the mid-1900s, New Testament scholars had largely rejected a dichotomous or trichotomous anthropology in favor of what I shall call a nonpartitive understanding of the human person – that is, the human person is a single whole, indivisible into parts; what I have otherwise referred to as anthropological monism, with its rejection of the need for an ontologically separate soul and/or spirit to account for human capacities and distinctives. This transformation is due especially to the influence of Rudolf Bultmann, whose theological analysis of the human as a creature who does not possess a body but who is a body would dominate subsequent

discussion (Bultmann 1955). Other scholars would poke and prod at Bultmann's understanding on this or that point, but this emphasis on the essential unity of human existence was championed by an array of prominent voices in New Testament scholarship in the latter half of the twentieth century. Thus, for example, John A. T. Robinson claimed in *The Body: A Study in Pauline Theology* that Paul's anthropology was essentially Hebrew in its wholistic and unified understanding (Robinson 1952, 11). F. F. Bruce, widely acknowledged as the dean of evangelical biblical studies in the twentieth century, concluded similarly that, in his anthropology, "Paul was a 'Hebrew born and bred.'" (Bruce 1971, 469). And the ever-influential Werner Georg Kümmel wrote in his study of *Man in the New Testament* that, with regard to Paul, we can speak only of the "complete" person (Kümmel 1963, 47). Writing toward the end of the twentieth century, New Testament theologian Udo Schnelle sought to correct Bultmann by urging that "a person has a body and is a body," (Schnelle 1996, 58) while claiming that Paul uses the term σῶμα (*sōma*, "body") to signify the human self (57). This list of New Testament scholars who have published significantly in this area and with similar conclusions could be expanded (see Green 2008, 3–16), but perhaps this is enough to demonstrate the direction historical inquiry has moved the conversation away from nineteenth-century views concerning Paul's two-part or three-part anthropology, in favor of a unified human being.

Moving beyond the general observation that historical inquiry has thus attempted to cast off the colonizing tendencies of theological and philosophical traditions that had emphasized body-soul or body-mind dualism, we can identify two primary factors that led to this reformulation of New Testament anthropology. The first is that New Testament scholars have generally learned more and more to take seriously the influence of Israel's Scriptures and its theological trajectories on early Christians whose writings now make up our New Testament – an influence that would include a portrait of the human creature, in the words of Brevard Childs, as "a complete entity and not a composite of parts" (Childs 1985, 199; Warne 1995). That is, if Israel's Scriptures rather than Plato's writings constitute the subtext of New Testament thought, then our assumptions about what words mean and our interpretive categories undergo a fundamental shift. Just as clinical psychologists and members of my church's adult education class will hear a word like "borderline" in quite different ways; just as a bass fisherman will understand the phrase "bank runner" or the term "bucketmouth" while the rest of us are left to our imaginations; in the same way, approaching New Testament texts with monist categories will shift what is seen and heard when compared with those who approach the text with dualist or trichotomous categories in tow. What we see depends a great deal on the prescription of our glasses or contact lenses, and historical inquiry in New Testament studies has effected a major change of prescription, so to speak.

If the first factor draws on the importance of Israel's Scriptures for New Testament thought, the second has to do with a rather wholesale rethinking of what it means to refer to the Greco-Roman background within which the New Testament materials were written. Against the tendency to imagine that Josephus or Philo represented a pervasive Platonic influence in the first-century world, we now know that Josephus and Philo are in some ways far from the center of Jewish engagement with Greco-Roman philosophy and, perhaps more importantly, that Platonic influence itself was not as dominant as was once imagined. The first-century Mediterranean world was characterized by diverse anthropologies, and the most prominent philosophical schools, Stoicism and Epicureanism, held decidedly non-Platonic views of the soul, both with regard to the soul's corporeal or material nature and

with regard to its immortality. The point is simply this: Against those who might imagine that, in their engagement with Greek thought, the New Testament writers developed a recognizably Greek approach to theological anthropology, we must recognize simply that it is not possible to speak reductively in this way of the world of Paul, John, or Luke. No singular conception of the soul held sway in the New Testament world, and the body-soul relationship was variously assessed among philosophers and physicians in antiquity.<sup>1</sup>

## 28.2 The “Soul” and Sociocultural Inquiry

We have drawn attention, first, to the significance of historical inquiry for situating the New Testament materials more securely within their first-century milieu as a prophylactic against colonizing New Testament perspectives with foreign assumptions about theological anthropology. We can push this point further, second, by developing further the direction that historical inquiry has taken us. I refer to the significance of sociocultural forms of inquiry, including medical anthropology and social psychology, for shaping our understanding of humanity in the New Testament world.

My first example of how historical inquiry has moved us in the direction of a wholistic anthropology comes in the form of medical anthropology, or ethnomedicine, which accounts for the ease with which Western readers of the New Testament Gospels read modern categories back into these texts. This is because, to a degree not often recognized, “sickness” is in the eye of the beholder, with the result that the identification and etiology of sickness, and the therapeutic interventions it warrants, are typically grounded in widely shared, culturally embedded understandings of health. If we can agree that “sickness” is an unwanted condition of self or substantial threat of unwanted conditions of self, then we can recognize that notions of health and sickness are tied to how a people measures human well-being. Accordingly, Western readers of the Bible have tended to focus on a diagnosis of a presenting problem and its resolution in terms oriented toward the physical body, an approach that reflects the Western medical tradition, but that often turns a blind eye to definitions of “healing” and “health” outside of the West, including those assumed in and supported by the biblical materials.

Robert A. Hahn of the US Centers for Disease Control and Prevention provides a useful antidote to the problem. He sketches a threefold taxonomy for intercultural study of healing. (1) *Disease accounts* focus on abnormalities located within the body, at or beneath the skin. The problem lies in the structure and functions of bodily organs or systems, so healing requires physical or biomedical intervention. (2) *Illness accounts* center on the body but also one’s networks of relationships and interaction with the larger social environment. The body is placed within a larger web of meaning that includes the embodied lives of persons in community. Healing might require physical intervention, but certainly must address the nesting of persons with others as the target of intervention. (3) *Disorder accounts*, without neglecting either the body or one’s networks of relationships, also attend to one’s relationship to the world at large, experienced as out of order. The recovery of well-being, in this case, would be tantamount to “putting the world back together,” or otherwise redressing a cosmic imbalance. Of course, these are ideal categories that, in the lived experience of a people, may overlap (Hahn 1995).<sup>2</sup>

If contemporary people of the West tend to think of disease preeminently in bodily terms (“disease accounts”), then they would also imagine that healing requires bodily or

biomedical intervention. People within biblical narratives, however, tend to think of sickness in more wholistic ways. The source of sickness for them lies not always and not only in the bodies of the sick, but also and sometimes especially in their social environments and in the larger cosmos (“illness” and “disorder accounts”).

Let us take as an example the case of “leprosy.” In the Gospels, “leprosy” only rarely, if ever, refers to true leprosy, or Hansen’s Disease, but instead includes a range of skin conditions (see, e.g., Omiya 2013, 517–518). According to Leviticus 13–14, these skin diseases are a sign of divine curse on a person and make a person “unclean” from a religious point of view. This is why someone diagnosed by a priest as a leper is relegated to the margins of human community:

The person who has the leprous disease shall wear torn clothes and let the hair of his head be disheveled; and he shall cover his upper lip and cry out, “Unclean, unclean.” He shall remain unclean as long as he has the disease; he is unclean. He shall live alone; his dwelling shall be outside the camp. (Leviticus 13:45–46, NRSV)

This “leprosy” is not life-threatening from a biomedical point of view, and accounts of leprosy in the Gospels are not concerned with the communication of a biological pathogen. However, using Hahn’s categories, the disorder of leprosy is communicable to others through physical contact. In this case the contagion is not a disease-causing microorganism, but the socio-religious status of ritual impurity.

Biblical accounts of leprosy thus demonstrate how religious, social, and physical considerations unite in a single disorder. Other examples lie close at hand, including persons who are demonized, and those who suffer from paralysis or generalized edema or even death. The issue is not (simply, or reductively) a “bodily” problem. It is not the case that a person’s “body” requires healing. Human health is not about “bodies” per se, as if bodies were separable from human lives. The situations in which we find Jesus intervening, and empowering his disciples to do the same, are more fully integrated human problems, and therapeutic interventions are aimed at the restoration of human health. Thus, having raised a dead man back to life, Jesus restores this man, “his mother’s only son,” to his widowed mother (Luke 7:11–17); and, after Jesus’s therapeutic intervention, the demonized Gerasene was found “sitting at Jesus’ feet, fully dressed, and completely sane,” and he was returned to his home with a vocation to “tell the story of what God has done for you” (Luke 8:26–39, CEB).

My first example of how historical inquiry has moved us toward a wholistic anthropology concerned insights from medical anthropology. My second example concerns the ease with which contemporary readers might transform words and phrases in the New Testament concerned with “inner” and “outer” into support for anthropological dualism, rather than working more from within the social psychology of the New Testament world. It is easy enough to see why this happens, given our inclination, as Charles Taylor notes in his important study of *Sources of the Self*, to posit both that a person has an “inner self” and that this “inner self” is the “true self” (Taylor 1989).

In their studies of Paul’s language of the “inner” person, however, both Hans Dieter Betz and Theo Heckel have rejected the view that, with this language, Paul is working within the framework of a body-soul dualism, with Heckel in particular emphasizing that Paul’s concern is with embodied life in this world (Betz 2000, Heckel 1993). More to our present point, though, is the way human “exteriority” – whether one is speaking of a person’s

countenance, the light of a person's eyes, or one's clothing – cannot be understood in terms separate from one's character, one's dispositions, one's inner life. This is a consequence of the social-psychological observation that, in the New Testament world, personal identity is outward-focused rather than inward.

Some rather obvious examples come to mind. In Galatians 3:27–29, to be “clothed in Christ” is not to wear Christ as an outer garment, a cover-up of “the real me,” for example, but expresses allegiance to Christ and the dissolution of socio-religious distinctions among different sorts of human beings – Jew and Gentile, slave and free, male and female. In Romans 13:12–14, putting on the armor of light and being clothed with the Lord Jesus Christ – these are not references to the personal façade I present to the public but are tantamount to honorable forms of believing, thinking, feeling, and behaving. As Klaus Berger puts it in his book, *Identity and Experience in the New Testament*:

when Jesus Christ is identified as the garment that one puts on, what is meant is that Jesus determines and defines what one represents in the eyes of one's contemporaries. (Berger 2003, 42)

The metaphor continues in other references to “getting dressed.” In 1 Timothy 2, “women,” we are told, “should dress themselves modestly and decently in suitable clothing, not with their hair braided, or with gold, pearls, or expensive clothes, but with good works, as is proper for women who profess reverence for God” (2:9–10, NRSV). Such a word of instruction makes sense in the ancient world only because clothing expressed not only status and wealth but *who one is*. This is why, in that world, a variety of literary forms, such as legal documents and philosophical treatises, and even images on coins, concerned themselves with appropriate dress. Items in this catalog – braiding of hair, gold, and fine clothing – would accordingly be interpreted as windows into a woman's being, displaying lack of self-control, immodesty, pompousness, even lewdness. Elsewhere, Peter addressed everyone, women and men, using the related metaphor of disrobing: “Take off (ἀποτίθημι, “to remove [clothing]”) every evil and every deceit, and pretenses and jealousies and all slander” (2:1; my translation).

Note the parallel way of thinking in texts that affirm the organic nature of the human being. According to John the Baptist, we know who they are that live conversionist lives by their behaviors: “Bear fruits worthy of repentance” (Luke 3:8, NRSV). According to Jesus, we know a tree is good because of its good fruit, whereas thorn bushes produce thorns, not figs (Luke 6:43–46). Or as James puts it, “Both fresh water and salt water don't come from the same spring, do they? My brothers and sisters, can a fig tree produce olives? Can a grapevine produce figs? Of course not, and fresh water doesn't flow from a saltwater spring either” (3:11–12, CEB). These are organic metaphors, not mechanical ones. Mechanical metaphors invite images of fitting disparate pieces together to produce a product, and allow the prioritization of parts. These images are foreign to notions of integration and essential unity characteristic of organic metaphors. (A qualitative difference exists between identifying the parts of a plant and the parts of an airplane.)

Illustrations of this kind do not negate the importance of the language of “inner” and “outer,” but they do beg for our understanding of this language outside categories associated with two-part or three-part views of the human person. They remind us that this way of speaking refers to aspects or dimensions of life and not to a distinction between immaterial and material or otherwise to a person's constitutive “parts.” References to the “outer

person” are not references to one’s corporeality, to one’s body as a discrete part of a person, but to the human in his or her entirety. For us, interiority often refers to the inner-directedness of a person and thus to his or her true self; within the New Testament writings, interiority may refer to a person’s core, but not so much the hidden, real self, but the self already transparent in its other-directedness.

Although our attention has been centered on the question of what we might call the constitution of the human person, the lines of inquiry we have followed press for reflection in other ways, too. These New Testament examples document the degree to which the human person is “twice embodied” – first, with reference to his or her whole being, indivisible in terms of his or her essential unity; and second, with reference to his or her existence within the social group(s) to which she or he belongs (McConville 2016, 59). Human beings must be understood not only in their wholeness, but also in their situatedness, including their relatedness.

### 28.3 Saving “Souls” in 1 Peter?

My concern has been the strange case of the vanishing soul. I have been exploring some of the reasons for the move in New Testament studies away from translations that project a portrait of the human person as having two “parts” – first, by drawing attention to the way historical inquiry has situated the New Testament materials more securely within their first-century milieu as a prophylactic against colonizing New Testament perspectives with foreign assumptions about theological anthropology; and second, by noting the significance of sociocultural forms of inquiry for shaping our understanding of humanity in the New Testament world. We now turn to the observation that New Testament texts are no longer regarded as transparently dualistic or trichotomous in their understanding of the human person. This is not to say that we cannot find interpretations that affirm dichotomous or trichotomous portraits of the human creature, but rather that such interpretations can no longer be regarded as the “obvious,” the “only,” or even the “best” readings. I have addressed a number of these texts in other publications (Green 1999, 2010), and will turn here to a brief sketch of the anthropological terminology of 1 Peter.

1 Peter is an interesting test case because of its six occurrences of the term *ψυχή* (*psychē*), a proportionately high number of such occurrences for a book in the Greek New Testament.<sup>3</sup> In one of those instances (1 Peter 1:9), we find the interesting phrase *ὑμῶν σωτηρίαν ψυχῶν* (*hymōn sōtērian psychōn*) used to describe faith’s aim: “the salvation of your souls” (KJV, NRSV). It is an attractive case, too, because Reinhard Feldmeier has recently argued in favor of Hellenistic Jewish influence on 1 Peter, to the effect that we find here “resemblances to a dualist anthropology” (Feldmeier 2008, 91).<sup>4</sup> With regard to 1 Peter 1:9 in particular, Feldmeier urges that the author of 1 Peter has identified the *ψυχή* (*psychē*) as “the anthropological correlate to God’s turning toward the world,” (Feldmeier 2008, 87) and so “the receptor of the saving salvific action,” (92) which survives death. I will show that Feldmeier’s general case has much to commend it, even if his primary claim has nothing to do with – and, in fact, does not require – body-soul dualism.

Feldmeier’s overall case is helped along, first, by his tentative argument that the apostle Peter did not author this letter and that it was written in the 80s CE. If one were to accept the letter’s own attribution to Peter and thus an earlier date for its composition, the idea that it reflected Hellenistic thought such as one finds in 4 Maccabees or Philo would face obstacles

Feldmeier's discussion bypasses. These include the usual dating of the composition of 4 Maccabees toward the end of the first century CE, and the difficulty of imagining how even a well-traveled Galilean like Peter could find himself immersed in the middle Platonism of an Alexandrian Jew like Philo.<sup>5</sup> Feldmeier's case is aided, second, by his practice of reading the anthropological terminology of 1 Peter against the backdrop of Philo's "clearly Hellenized anthropology" and, to a lesser degree, in relation to the anthropological perspectives of the Wisdom of Solomon and 4 Maccabees (Feldmeier 2008, 91). As Feldmeier is well aware, the philosophical climate within which 1 Peter was written was variegated, and this is true whether the letter was composed in the 60s or the 80s. Especially prominent among the options available in the Greco-Roman world are Epicureanism, which understood both mind and spirit as corporeal because they act on the body (and all entities that either act or are acted upon are bodies by definition), and Stoicism, which taught that everything that exists, including the human soul, is corporeal. Israel's Scriptures themselves project a monist anthropology. And the confluence of these two cultural streams – Greco-Roman and Hebrew – yielded a range of positions that could be called Hellenistic Jewish. Some would have been more clearly dualist (as Feldmeier observes), others monist (though Feldmeier does not mention these). Since simple appeal to the cultural matrix within which 1 Peter was written cannot resolve matters, the question remains: How has Peter portrayed the human person? As it is known to us from 1 Peter, was his a more dualist anthropology, or did his understanding of the human person reflect the monism of Israel's Scriptures (and/or of those prominent Greco-Roman philosophies that were more monist)?

We have already had reason to note that New Testament writers, Peter among them, had access to no denotative language by which to lexicalize their theologies of the human person. 1 Peter is neither a dictionary, nor does it provide us with its working definitions of key anthropological terms. Accordingly, we are left with the need to examine how the letter actually uses the relevant terms. Turning to this task, it would be an egregious oversight on my part if I failed to mention that we find nothing in 1 Peter to remind us of the turns of phrase found in Philo's middle Platonism, nothing analogous to Philo's transparent references to body and soul as discrete human essences, nothing approaching Philo's claims concerning the soul's sovereignty over the body, and nothing that identifies the body as a tent or shrine in which the soul might dwell (see, e.g., Philo, *On Creation of the World*). Accordingly, anyone wanting to find in 1 Peter a body-soul dualism analogous to what is so obvious in Philo would need either to presuppose Philo's influence on 1 Peter or to find similarly unambiguous linguistic and/or conceptual evidence.

σῶμα (*sōma*, "body") appears in 1 Peter only once, in 2:24: "He himself bore our sins in his body (σῶμα, *sōma*) on the cross, so that, free from sins, we might live for righteousness" (NRSV). Peter's claim comes at the apex of this presentation of Christ's suffering, which he presents to his audience not only as exemplary and effective but as manifestly corporeal (2:21–25). Elsewhere in the letter, we find that this Christ testified in advance through the prophets to his suffering and glory (1:10–11), and that this Christ will be revealed in glory (1:13; 4:13). Peter thus presents Christ as a transcendent figure who shares in God's identity, and yet who, in his advent, experienced bodily personhood and, as such, experienced personal suffering and death. Theologically, this speaks profoundly of the significance of human existence, including human, bodily suffering.

σὰρξ (*sarx*, "flesh," "body") appears in 1:24, where it signifies "humanity," and otherwise in 3:18, 21; 4:1 (twice), 2, 6 with reference to "life as a human." Thus, Christ was put to death "as a human" (3:18), just as he suffered "as a human" (4:1). This translation gains traction



from the parallel constructions in 4:2 (τὸν ἐπίλοιπον ἐν σαρκὶ βιώσαι χρόνον, “live the rest of their human lives,” CEB) and 4:6 (κριθῶσιν . . . σαρκί, “they were judged as humans,” CEB).

ψυχὴ (*psychē*, “life,” “vitality,” sometimes translated as “soul”) appears in 1:9, 22; 2:11, 25; 3:20; 4:19. In 2:11, ψυχὴ (*psychē*) is set in contrast to σαρκικός (*sarkikos*, “belonging to this world”), and so connotes “life as it is determined by or directed toward God.” ψυχὴ (*psychē*) never appears in 1 Peter in relation to (or as distinct from) either σῶμα (*sōma*) or σὰρξ (*sarx*), however. Christ is the guardian of the Christian’s ψυχὴ (*psychē*) in 2:25, just as God is guarding “you” for a salvation ready to be revealed at the last time (1:4–5; ὑμᾶς, *hymas*, “for you”). This parallelism makes it difficult to read ψυχὴ (*psychē*) in 2:25 in any way other than as an equivalent to the personal pronoun, that is, as a reference to the unified person in his or her entirety. Similarly, those who suffer entrust their lives (or themselves, τὰς ψυχὰς αὐτῶν, *tas psychas autōn*) to God (4:19). In 1 Peter 3:20, a plural form of ψυχὴ (*psychē*) refers to “persons,” Noah and his kin (and not only a part of each person); presumably, people, and not only their immaterial souls, need to be rescued through the flood. Given this range of texts, as well as the eschatological goal to which faith is directed in 1:9, it is easy to affirm the translation of this verse (κοιζόμενοι τὸ τέλος τῆς πίστεως ὑμῶν σωτηρίαν ψυχῶν) offered by the Common English Bible: “You are receiving the goal of your faith: your salvation.”<sup>6</sup>

How might we gloss the perspective 1 Peter has provided? We might say that, in Peter’s hands, σὰρξ (*sarx*) refers to “life as it reflects and/or pertains to this world,” whereas ψυχὴ (*psychē*) signifies “life as it reflects and/or pertains to the world to come.” This conclusion coheres well with Feldmeier’s proposal concerning ψυχὴ (*psychē*): “the anthropological correlate to God’s turning toward the world.” Peter’s perspective is dualistic, therefore, but not in the way Feldmeier might have us believe. Peter’s is an eschatological dualism – life in this world versus life in the age to come, or life as it is situated in this world versus life as it is directed toward God – rather than an anthropological one – body versus soul. Accordingly, with respect to his anthropology, we find ourselves with Peter more at home in Israel’s Scriptures than in the philosophy of Plato’s successors.

To be sure, Peter knows and works with what we might call a container metaphor, even if it is not the sort we find in Philo. With Philo, the body functions as a container for the bearer of God’s image, the soul. This is not what we find in 1 Peter. Peter begins the letter with reference to the location of his audience as “chosen strangers in the world of the diaspora,” a reference to their situatedness as foreigners (1:1–2, CEB). This is true even if he is concerned that their lives not be thus determined, or contained, in an ultimate sense by this location, as though the harassment they suffered comprised the last word to be spoken over them. Rather, he maps his audience in relation to Israel’s story and Christ’s career, that is, with reference to their journey as a pilgrim people, so that they find their true home “in Christ” (5:10, 14). “In Christ” – this is the sphere of influence Peter desires for his audience, for their way of life as sojourners. Their way of life in this world thus gains its true significance through its Christological redefinition, that is, with reference to the interpretive significance of the exemplary and redemptive journey of Christ through suffering to exaltation. Life in this world is a participation in Christ’s suffering, but the destination of this life’s journey is a share in God’s honor.

This portrait of humanity allows Peter to take with utmost seriousness the situatedness of his audience – not by urging these harassed believers to retreat into their genuine selves from bodily suffering and social banishment, as though their true selves might remain

untouched by the calamity of suffering; and not by reminding them that their bodily suffering would give way to God's pressing concern with their "souls." Rather, his portrait of the human provides life in this world its fullest significance, centering his emphasis on a faithful "manner of living" in the world. Peter's anthropology is thus twice embodied – bodily life, to be sure, indivisible in terms of a person's essential unity, but also full-bodied life among a people called to follow Christ's example, to follow in his footsteps (2:21). According to this portrait, human existence is shaped by the struggles accompanying those whose commitment to the suffering-and-risen Christ ensures lives out of step with society-at-large, and by the call to hold together and stand firm in their commitment to this Christ.

## 28.4 Conclusion

Over the past five centuries, those who translate the Greek New Testament for English readers have increasingly found it appropriate to do so without recourse to a human soul. This is not simply a case of linguistic slippage, but the consequence of sustained exploration of the social-historical milieu within which the New Testament writers lived and wrote. The fruit of that sustained exploration has been making its way into a wide range of studies adjunct to what we today call theological anthropology, with the result that texts that perhaps once seemed to support a partitive understanding of the human person do so less and less. In popular usage, the English term "soul" usually refers to an immaterial, immortal part of a human. Since the Greek term *σάρξ* (*sarx*) is not very well lexicalized in this way, New Testament studies has increasingly learned to do without the soul. This has allowed readers of the New Testament (and the Christian Bible more generally) more fully to explore theologically the nature of embodied human life understood not only in terms of human wholeness, but also in their situatedness, including their relatedness.

## Notes

1. See, for example, the opening chapters of Wright and Potter (2000); Hippocratic medicine (Beate Gundert); Plato (T. M. Robinson); Aristotle (Philip J. van der Eijk); a cluster of Hellenistic philosophers and physicians, from Epicurus to Galen (Heinrich von Staden); and Paul (Theo K. Heckel). See also the summary in Martin (1995, 3–37).
2. I have addressed the relevance of Hahn's (1995) typology to study of the New Testament world (Green 2013, 330–341).
3. In the Greek New Testament, only 3 John (with a single instance of *ψυχή* (*psychē*) among a total of 218 words (0.46%)) has a higher proportion than 1 Peter (six instances of *ψυχή* (*psychē*) in 1679 words (0.36%)).
4. Feldmeier (2008, 87–92) casts his argument as a response to Dautzenberg (1964).
5. In fact, the standard critical commentaries, even though they too assume that the apostle Peter did not author this letter, read its anthropology against the backdrop of Israel's Scriptures (see Achtemeier 1996, 104; Elliott 2000, 344).
6. See "the salvation of [you] lives" (Elliott 2000, 328); "your salvation" (Achtemeier 1996, 99).

## References

- Achtemeier, Paul J. 1996. *1 Peter, Hermeneia*. Minneapolis, MN: Fortress.
- Bauer, Walter. 2000. *A Greek-English Lexicon of the New Testament and Other Early Christian Literature*, revised and edited by Frederick William Danker, 3rd edn. Chicago: University of Chicago Press.
- Berger, Klaus. 2003. *Identity and Experience in the New Testament*. Minneapolis, MN: Fortress.

- Betz, Hans Dieter. 2000. "The Concept of the 'Inner Human Being' (ὁἷσω ἄνθρωπος) in the Anthropology of Paul." *New Testament Studies*, 46: 315–341.
- Bruce, F. F. 1971. "Paul on Immortality." *Scottish Journal of Theology*, 24: 457–472.
- Bultmann, Rudolf. 1955. *Theology of the New Testament*, 2 vols. New York: Charles Scribner's Sons.
- Childs, Brevard S. 1985. *Old Testament Theology in a Canonical Context*. Philadelphia, PA: Fortress.
- Dautzenberg, G. 1964. "Σωτηρία ψυχῶν (1 Peter 1, 9)." *Biblische Zeitschrift* (n. s.) 8: 262–276.
- Elliott, John H. 2000. *1 Peter: A New Translation, with Introduction and Commentary. Volume 37, Part 2: Anchor Bible*. New York: Doubleday.
- Feldmeier, Reinhard. 2008. *The First Letter of Peter*. Waco, TX: Baylor University Press.
- Green, Joel B. 1999. "Restoring the Human Person: New Testament Voices for a Wholistic and Social Anthropology." In *Neuroscience and the Person: Scientific Perspectives on Divine Action*, edited by Robert John Russell, Nancey Murphy, Theo Meyering, and Michael A. Arbib, 3–22. Vatican City State: Vatican Observatory; Berkeley, CA: Center for Theology and the Natural Sciences.
- Green, Joel B. 2008. *Body, Soul, and Human Life: The Nature of Humanity in the Bible* (Studies in Theological Interpretation). Grand Rapids, MI: Baker Academic.
- Green, Joel B. 2010. "What about . . . ? Three Exegetical Forays into the Body-Soul Discussion," *Criswell Theological Review* (n.s.) 7: 3–18.
- Green, Joel B. 2013. "Healing and Health Care." In *The World of the New Testament: Cultural, Social, and Historical Contexts*, edited by JOEL B. Green and Lee Martin McDonald, 330–342. Grand Rapids, MI: Baker Academic.
- Hahn, Robert A. 1995. *Sickness and Healing: An Anthropological Perspective*. New Haven, CT: Yale University Press.
- Heckel, Theo K. 1993. *Der innere Mensch: Die Paulinische Verarbeitung eines Platonischen Motivs* (Wissenschaftliche Untersuchungen zum Neuen Testament 2, no. 53). Tübingen: Mohr Siebeck.
- Johnson, Samuel. 1785. *A Dictionary of the English Language*, 2 vols, 6th edn. London: H. G. Collins.
- Kümmell, Werner Georg. 1963. *Man in the New Testament*. London: Epworth.
- Louw, Johannes P., and Eugene A. Nida, eds. 1988. *Greek-English Lexicon of the New Testament Based on Semantic Domains*, 2 vols. New York: United Bible Societies.
- Martin, Dale B. 1995. *The Corinthian Body*. New Haven, CT: Yale University Press.
- McConville, J. Gordon. 2016. *Being Human in God's World: An Old Testament Theology of Humanity*. Grand Rapids, MI: Baker Academic.
- Omiya, Tomohiro. 2013. "Leprosy." In *Dictionary of Jesus and the Gospels*, edited by JOEL B. Green, 2nd edn. Downers Grove, IL: IVP Academic.
- Robinson, John A. T. 1952. *The Body: A Study in Pauline Theology* (Studies in Biblical Theology, 5) London: SCM.
- Schnelle, Udo. 1996. *The Human Condition: Anthropology in the Teachings of Jesus, Paul, and John*. Minneapolis, MN: Fortress.
- Taylor, Charles. 1989. *Sources of the Self: The Making of the Modern Identity*. Cambridge, MA: Harvard University Press.
- Thiselton, Anthony C. 2011. *1 and 2 Thessalonians through the Centuries* (Blackwell Bible Commentaries). Oxford: Wiley-Blackwell.
- Warne, Graham J. 1995. *Hebrew Perspectives on the Human Person in the Hellenistic Era: Philo and Paul*. Lewiston, NY: Mellen.
- Wright, John P., and Paul Potter, eds. 2000. *Psyche and Soma: Physicians and Metaphysicians on the Mind-Body Problem from Antiquity to Enlightenment*. Oxford: Clarendon.
- Wright, N. T. 2008. *Surprised by Hope: Rethinking Heaven, the Resurrection, and the Mission of the Church*. New York: Harper One.

# Debating the Incarnation

# Dualism Offers the Best Account of the Incarnation

LUKE VAN HORN

Christian materialists claim that we are material objects. In many assessments of this claim, the relevance of the doctrine of the Incarnation has been underappreciated. If we are material objects, the Incarnation could not have occurred. Substance dualism faces no such difficulty, however, so Christians should be dualists.

The Incarnation, as I understand it, is the doctrine that the second person of the Trinity (hereafter SPT) became fully human (as human as you or I) while remaining fully divine (as divine as the Father and Holy Spirit).<sup>1</sup> Thus, Kenotic theories are not really analyses of the Incarnation and shall be ignored in what follows.<sup>2</sup>

There are debates over what *materialism* and/or *physicalism* are. I do not intend to wade into these disputes. In what follows, I shall use “materialism” and “physicalism” interchangeably. By these terms I intend to refer solely to the claim that human beings are entirely material objects. Thus, I am not referring to broader claims that the universe – or reality – is purely material or physical.<sup>3</sup> Neither am I taking sides on whether or not various types of property/event dualism and emergentism count as versions of materialism/physicalism. Even if some of these views posit “nonphysical” properties, as long as the human beings alleged to have these properties are material objects, such views count as materialist/physicalist, at least in this chapter.<sup>4</sup>

I will not attempt to address every possible type of materialist Christology. While application of worm or stage theory to the Incarnation would be interesting, no one to my knowledge has advanced such views in print. Critiquing theories endorsed by no one is not the best use of limited space, so I will only interact with accounts physicalists have actually defended.<sup>5</sup> Space limitations also prohibit discussion of Peter van Inwagen’s relative identity account of the Incarnation. Although his account is not explicitly materialist, he is likely the most prominent Christian materialist today and his model is clearly available to materialists. Fortunately, relative identity has been discussed extensively elsewhere (see van Inwagen 1995; Hughes 2009, 298–300; Rea 2003).

## 29.1 Materialist Christology

Trenton Merricks has presented the most prominent defense of a materialist Christology to date (Merricks 2007). His account is quite simple. As an animalist, he claims that at the point of Incarnation, SPT became an animal, a composite material object. This is in the most literal sense: SPT ceased being an immaterial object at the moment of Incarnation and transformed into a wholly material object composed of quarks and electrons. Once incarnate, no immaterial substance is involved, whether “the divine nature” or a soul.<sup>6</sup> This is a radical view (hereafter labeled animalist Christology, or AC) which faces a number of objections.

God is omnipotent and omniscient. Thus, an orthodox model of the Incarnation must maintain the omnipotence and omniscience of SPT throughout the period of humiliation.<sup>7</sup> Merricks must then insist on the possibility of a human organism being omniscient. But is this possible? Robin Le Poidevin has argued in the negative:

For it is not possible for the characteristic properties of divinity, including omnipotence and omniscience, to be realised by the states of a limited human body. God the Son may be able to bring anything about, but not simply by means of the powers of a recognisably human body. (Le Poidevin 2009b, 710)

“And insofar as that being’s mental states are realized by states of a normal human brain, it is surely impossible for such a being to be omniscient” (Le Poidevin 2009a, 170). Jonathan Hill has argued similarly:

But if Christ is a human body, then a human body must be omnipotent and omniscient. Is that possible? A human body does things by moving itself and affecting the environment directly around itself, and it knows things by storing information in its brain. And a human brain could not have the capacity to store the infinite amount of knowledge required for omniscience. (Hill 2012, 11)

I won’t claim that these are knockdown objections, but it is certainly difficult to conceive of a human brain not only storing an infinite amount of information but also having that information accessible as knowledge. In any case, even if an account of how this could work is possible, it would have to cover all stages of SPT’s biological development. It is difficult enough to believe that a human infant could be omniscient, but claiming that an embryo without a soul knows everything strains credulity too far.<sup>8</sup>

Another difficulty for AC is that it requires that a necessary being can be wholly composed of contingent parts, but many may contend that anything wholly composed of contingent parts must be contingent. Worse, since these parts are also created, it seems that, on AC, SPT becomes a creature, for whatever is wholly composed of created parts is created. Richard Cross, another proponent of AC (or something similar), has attempted to refute this objection (Cross 2003, 302–306). He offers a two-pronged response: (1) perhaps Lynne Rudder Baker’s metaphysics is correct, in which case SPT, while incarnate, is constituted by, but numerically distinct from, the human organism that is his body, allowing SPT to be uncreated while his body is created; (2) SPT has properties that are inconsistent with being created (such as necessary existence and eternity), so being wholly composed of created parts must be consistent with being uncreated.

Neither of these strategies is particularly helpful. Baker's metaphysics of co-located, constituted objects is extremely controversial, which would make materialist Christology vulnerable to the objections to that metaphysics.<sup>9</sup> Further, it is unclear that Baker's view is even compatible with AC. If SPT merely came to be "constituted" by his body, then he only became *related to* rather than *became* a material object. Baker distinguishes between the *derivative* and *nonderivative* having of a property.<sup>10</sup> A constituted object like a statue or a person has physical properties like mass and color derivatively. To have a property derivatively is to be constituted by something that has that property.<sup>11</sup> Further, constituted objects are actually mereological atoms. While we may say that a person's hand is a part of the person, what this cashes out to is merely that the person is constituted by something else (it's body) that has the hand as a part (Baker 2007, 187). This application of Baker's metaphysics does not save AC. Instead, it suggests that SPT merely becomes an extended simple related to a human body, yet failing to acquire most of the properties of material objects. This looks very much like a substance dualist account of the Incarnation, not AC.<sup>12</sup>

The second prong of Cross's defense is that SPT has properties like necessary existence which are inconsistent with being created. This is simply a bald assertion of AC's coherence in the face of an objection to its coherence. It does nothing to undermine the intuition that things entirely composed of creatures are themselves creatures. To adapt Brian Leftow's comment on a related objection to AC, "It would be theft rather than honest toil to say that [SPT] has the divine nature and [that] this" deflects any charges of incoherence (Leftow 2015, 79).

In addition to theological objections to AC, there are purely metaphysical difficulties.<sup>13</sup> For example, one might plausibly think that immateriality entails nonspatiality and that nothing nonspatial could acquire a spatial location. Similarly, an immaterial, nonspatial entity is unextended, but one might think that an unextended object could not come to be extended. Further, AC requires that the possibility of a simple – an entity with no proper parts whatsoever – becoming composite.<sup>14</sup> But this seems impossible. This intuition is well-expressed by Kevin Corcoran:

I suspect there are no conclusive non-question-begging arguments for the claim that physical objects are essentially physical. Nevertheless, the consequences of its denial strike me as simply unacceptable. For if the denial is true, then it is possible for a thing at this world that is spatially extended and weighs two tons (an elephant, say) to exist (that very thing!) at another world without any mass whatever and without, in fact, being a physical object at all. (Corcoran 1999, 18n3; cf. Taliaferro and Goetz 2008, 310)

Corcoran is likely correct that there are no conclusive arguments for simples being essentially simple, and attempts to argue for it may be less obviously correct than the intuition itself. Nevertheless, if an argument is needed, consider one inspired by the paradox of increase (Olson 2006).

- 1 A simple object *x* incorporates *y* as a part at time *t*.
- 2 Necessarily, nothing can have a single proper part.
- 3 At *t*, *x* is composed of *y* and the complement of *y* (i.e., all of *x* except for *y*).
- 4 The complement of *y* exists prior to *t*.
- 5 Prior to *t*, *x* has all of the same parts as the complement of *y*.
- 6 Necessarily, two distinct objects cannot have all of the same parts at the same time.

From these premises, we can derive that  $x$  is identical to one of its proper parts, a contradiction. For the original paradox of increase (involving composite objects acquiring parts), Merricks is most likely to favor denying (3), as it would be a consequence of the doctrine of arbitrary undetached parts (DAUP), and a good case can be made against this principle (van Inwagen 1981; Merricks 2001, 48–53). However, when applied to simples, (3) would seem to be true, even if DAUP is false. Premises (4), (5), and (6) also appear to be true. But then (1) must be false; simples cannot gain parts.

Merricks could claim that this sort of argument shows that a simple cannot incorporate a single proper part at a time, but it could incorporate multiple proper parts at once. Premise (5) is obviously true if  $x$  is incorporating a single part, but perhaps it is false if  $x$  acquires two or more parts at  $t$ . This kind of response strikes me as weak. For instance, it would require the falsity of the following principle:

- 7 If, at  $t$ ,  $x$  incorporates  $n$  parts (and loses no parts), then, at  $t$ ,  $x$  has at least  $n + 1$  proper parts.

In any ordinary case, the principle seems true. For example, if a carrot nose is added to a snowman made of three large snowballs at  $t$ , then at  $t$  the snowman has four proper parts. Moreover, the proposed way out has bizarre consequences. If a simple tries to acquire a single part, it will fail, but if it tries for two or more it might succeed? That is like saying that a composite object with only two proper parts cannot lose one of its parts without ceasing to exist, but it could lose both and still exist. Worse, the parts must be incorporated at exactly the same time. If the simple tries to incorporate one part a picosecond before the other, it will fail. But could the possibility of incorporating a new part really hang on such a short period of time?<sup>15</sup>

Again, none of the above objections are knockdowns, but their combined force strikes me as sufficient to discredit AC as a plausible physicalist account of the Incarnation.<sup>16</sup> If Christians want to be physicalists, they should look for other ways to integrate Christology with materialism.

One attempt at such an alternative is to endorse what Thomas Flint has dubbed the “Model T” version of so-called “concretist” analyses of the Incarnation (Flint 2011b, 71). Concretists tend to think of the Incarnation in mereological terms, claiming that SPT came to stand in a mereological relation to a particular human being. According to Model T, at the moment of Incarnation, SPT acquired a human organism as a proper part. This organism is not itself a person, but it is otherwise a human being just like you and me. Nothing in this model requires that human organisms have souls, so materialists are free to adopt it.<sup>17</sup>

At first blush, this model appears significantly different from AC, since it does not claim that SPT transforms himself into a human organism (HO). Rather, HO is merely a proper part of SPT, the “human nature” that complements SPT’s divine nature in the Incarnation. Upon closer examination, however, AC and Model T appear to be the same account of the Incarnation. Prior to the Incarnation, SPT is a mereological simple. Upon the acquisition of HO as a part, SPT is composed of nothing more than HO (there is no “God-matter” that composes SPT in addition to HO). Model T thus reduces to AC and faces all of the same theological and metaphysical challenges. If one attempts to avoid this result by denying that SPT is a simple prior to the Incarnation, the traditional paradox of increase must still be faced:



- 8 At  $t$  (a time prior to the Incarnation),  $SPT = W$  (whatever composes him prior to the Incarnation).  
 9 At  $t_2$ ,  $SPT = (W + HO)$ .

Again, from these premises we can derive that SPT is identical to one of his proper parts, a contradiction.<sup>18</sup>

If advocates of Model T can somehow avoid the above metaphysical objection and distinguish their account from AC, they face a theological challenge common to all versions of Model T, whether materialist or not. They must avoid sliding into the heresy of Nestorianism, which says that the Incarnation involved two persons, SPT and a distinct human person. At first glance, Model T appears to be an explicitly Nestorian model, since it claims that HO, a human being just like you and me, is a part of SPT. Since we are persons, it appears that there are two persons in the Incarnation, SPT and HO.

Although not an advocate of Model T, Brian Leftow has offered on its behalf the following principle of natural kind maximalism:

- 10 Given a set of parts composing at time  $t$  a member of a natural kind (e.g., cat), no subset of that set composes at  $t$  a member of the same natural kind (Leftow 2002, 282).

If *person* is a natural kind, then it is impossible that HO be a person, at least while part of SPT, and the threat of Nestorianism is averted. However, I don't find this strategy convincing. Natural kind maximalism faces plausible counterexamples in organisms like two-headed pigs. Even if the principle is restricted to persons rather than natural kinds generally, it is not obviously true. On the contrary, it seems clearly false, since an explicitly Nestorian version of Model T seemingly provides a clear counterexample. Michael Burke, while advocating that no proper part of a *human* person is a person, suggests that the more general claim about all persons is likely false.

It would be risky to claim that no person-parts are persons. It would be riskier still to claim that it's *impossible* for person-parts to be persons. Perhaps somewhere there are multicellular persons composed of unicellular persons. Or, if there aren't, perhaps there *could* have been. (Burke 2003, 112)

To avoid counterexamples of this sort, Burke defends an alternative definition of maximalism:

Kind/property/term/concept  $C$  is *maximal* just in case necessarily, no identity-sufficient [proper] part of  $C$  is itself a  $C$  . . . Something is an *identity-sufficient* part of a  $C$  just in case the particles composing the part would immediately compose *that* very  $C$ , if the complement of the part suddenly . . . ceased to exist. (Burke 2003, 112–113)

To illustrate, Descartes-minus (the part of Descartes that is all of him but his left leg; van Inwagen 1981) is an identity-sufficient part of Descartes. Burke's account allows that Descartes-minus is not a person and that *human personhood* is maximal, even if *personhood* more generally is not. But this account is of no use to advocates of Model T, for it is not true that HO would be the very same person as SPT should HO's complement cease to exist.<sup>19</sup> Rather, HO would be a mere human being, a creature, not SPT. Since HO would not be SPT, Burke's account does not entail that HO is not a person while part of SPT. I won't claim that there are not (or could not be) other accounts of personhood maximalism as defensible as Burke's, but I am not aware of any advocates of Model T who have presented such. As things

stand, it appears that appealing to maximalism has not yet produced a credible defense against Nestorianism.

Timothy O'Connor and Philip Woodward have suggested a different strategy for claiming that HO is not a person. Recognizing that HO having the full panoply of mental capacities, experiences, and intentions would appear to make HO a person, they suggest the following:

When God the Son took on a human body as a part, the emergence base for that human body's mental states was expanded. It then included not only the types of causal powers that would ordinarily be sufficient to generate an experiencing subject and agent at the center of a dynamic phenomenal/intentional manifold. It also contained divine causal powers that masked the causal powers responsible for the emergence of a proprietary human subject. Conscious mental state[s] nevertheless emerged, but absent a proprietary human subject, they emerged as mental states of the larger individual, the divine-human composite. (O'Connor and Woodward 2014, 233–234)

This proposal appears to succeed in dispelling any appearance of personhood for HO. However, in doing so, O'Connor and Woodward have undermined one of the key motivations for Model T. One of the attractions of Model T is that it claims that a human being just like you and me is a part of SPT, allowing SPT to have all of the properties of humanity while retaining all of his divine attributes. But according to O'Connor and Woodward HO is not a human being like you and me. HO appears to entirely lack a mental life. HO experiences and wills nothing. Rather, SPT is the subject of all experiences and the origin of every act of will. True, HO is not entirely nonmental, since its brain produces sensations in SPT, but HO does not experience them. Garrett DeWeese has critiqued similar strategies by nonmaterialist proponents of Model T:

The unintended result of this line of thinking is that Christ's human will/mind/consciousness becomes little more than a theoretical entity with no observable consequences in the life of Christ. Christ's exemplary role as a perfect man simply evaporates. (DeWeese 2007, 133)

Moreover, O'Connor and Woodward's proposal ironically appears to advocate a form of Cartesian dualism: the brain produces experiences in the soul (SPT in this case), and the soul exercises its will, causing effects in the brain. In avoiding Nestorianism, O'Connor and Woodward appear to have unintentionally abandoned materialist Christology.

One final strategy for avoiding Nestorianism is offered by Glenn Andrew Peoples.<sup>20</sup> Following Brian Leftow, he suggests that SPT may have "gotten to" HO at a time sufficiently early (perhaps at conception) to prevent HO from becoming a person distinct from SPT.

The result is a fully functioning human body whose life (i.e., whose timeline) was the life of the incarnate Son of God rather than whatever life it would have constituted. It may even turn out that a materialist view of human beings has a marked advantage over substance dualism here. It is easy to think that a human body might fail to be a human being or a person, but it is much more difficult to think of a simple, immaterial human soul that is not a person. What else can a Cartesian soul be but a person? (Peoples 2015, 338)

I am inclined to agree with Peoples that a Cartesian soul cannot fail to be a person, and so believe that dualist versions of concretism tend to be Nestorian. However, I fail to see how materialism does any better avoiding this heresy. It is indeed easy to think of human bodies

in states which fail to satisfy whatever the criteria are for personhood (even if personhood begins at conception, doubtless it ends at death). However, a “fully functioning human body” is surely a person (assuming materialism), at least after suitable development of the brain. If HO is “fully functioning,” then he has a mental life like yours and mine, even if he was “gotten to” (a terribly vague notion) at conception. What else can a human organism with such a mental life be but a person?<sup>21</sup>

Proponents of Model T face a dilemma. In order to claim that SPT was fully human in the Incarnation, they must claim that HO had a mental life like yours and mine, which is Nestorianism. To avoid Nestorianism, it seems they must limit HO’s mental life, a move which fails to make SPT fully human, courting substance dualism. But if the cost of avoiding heresy is dualism, why not pay it and do one’s Incarnational theorizing from an explicitly dualist framework?

## 29.2 Dualist Christology

In an earlier paper, I have defended a dualist account of the Incarnation (Van Horn 2010). In brief, the model assumes Cartesian dualism and a position on souls according to which *rational soul* is a natural kind. What differentiates a human soul from a Wookiee soul is simply its body. A disembodied soul is neither human nor Wookiee.<sup>22</sup> In the Incarnation SPT became a human being by becoming embodied in a particular HO. In so doing he became a human soul. This involved no fundamental transformation on his part,<sup>23</sup> allowing him to remain fully divine while simultaneously becoming fully human.

Objections to this sort of Incarnational model tend to be fairly weak, at least when compared to those faced by materialist Christology. For example, Anna Marmodoro and Jonathan Hill object that dualism is an “undesirable commitment.”

Models of this kind are committed to substance dualism . . . However, the fact that this model depends upon such a theory of the mind is surely problematic in itself. The difficulties with substance dualism are well known, to the extent that few philosophers today defend it. (Marmodoro and Hill 2008, 115–116)

True, dualism is a minority position among philosophers today, but this should carry little weight among Christians, for whom dualism has always been the majority view. Further, the various objections to dualism tend to be greatly overrated. As William Hasker has commented:

I want to say up front that many of the popular objections against dualism are mistaken, unfair, or just plain bad philosophy. For instance, the well-worn objection that mind and matter cannot interact because they are different kinds of substances has my vote for being the most overrated philosophical objection of all time. (Hasker 2010, 95–96)

This chapter is not the place to address the standard objections to dualism, but thorough responses to all of them are readily available elsewhere.<sup>24</sup>

Perhaps the most powerful objection is offered by Trenton Merricks.<sup>25</sup> He argues that dualists cannot provide an adequate account of embodiment that entails SPT being uniquely embodied in HO while neither the Father nor Holy Spirit are themselves embodied (Merricks 2007, 284–288). I am inclined to agree that several accounts he critiques are flawed and should not be endorsed by dualists. However, the following account

appears to meet both of his conditions for a successful analysis of embodiment: *S* is embodied in *x* iff *S* is causally disposed to have direct causal power over *x* and *x* is causally disposed to cause effects in *S*.<sup>26</sup> No object other than HO is disposed to cause effects in SPT, so SPT is uniquely embodied in HO. Further, no objects are disposed to cause effects in the Father or Holy Spirit, so neither is embodied. Thus, this account of embodiment is consistent with the Incarnation.

Merricks objects that this type of account makes disembodiment impossible:

Consider a disembodied soul, whose former body has died. If that body were in sense-experience-causing conditions – conditions presumably requiring it to be alive – then I suppose the soul would have the appropriate experiences. After all, the nearest counterfactual situation in which, for example, Lincoln’s body is now alive is presumably, given dualism, a situation in which Lincoln’s soul is embodied. (Merricks 2007, 288n10)

Merricks is focusing on the wrong counterfactual conditions. While it may be true that *If Lincoln’s body were alive, he would be embodied*, the proponent of the above account of embodiment thinks that Lincoln is embodied only if a counterfactual like *If Lincoln’s body is punched, he feels pain* is true. That counterfactual is currently not true, so Lincoln is not currently embodied.

Lynne Rudder Baker has objected that dualist models of the Incarnation like mine “court Apollinarianism” (Baker 2011, 53). Peter van Inwagen defines this heresy as claiming that

Christ did not have a human mind or spirit or rational soul – that he lacked something that is essential to human nature – and that God or some “aspect” of God . . . was united to the human body of Jesus of Nazareth in such a way as to “be a substitute for” or perform the function of the human mind or soul or spirit. (van Inwagen 1998, 727)

I plead, “Not guilty.” While it is easy to see how a dualist Apollinarian account of the Incarnation could be constructed, on my model Christ did have a human mind, spirit, and rational soul. That soul is SPT, who at the moment of Incarnation became, rather than substituted for, a human soul. He became a fully human being just like you and me, so there was no essential aspect of humanness that Christ lacked.<sup>27</sup>

## 29.3 Conclusion

To adapt a frequent comment of Plantinga’s, “Materialism is all the rage these days.” This is true even among Christian theologians and philosophers, so doubtless we will see new, more elaborate versions and defenses of materialist Christology in the future. Nevertheless, these projects are misguided. The Incarnation, the central doctrine of the Christian faith, is a very poor fit with materialism. It fits quite well with Cartesian dualism, however, so Christians should be dualists.<sup>28</sup>

## Notes

1. My arguments are not intended to assume divine temporalism, so proponents of divine timelessness are welcome to adjust my wording where appropriate.

2. Kenotic theories claim that in some way SPT was not fully divine while incarnate.
3. I also do not intend to distinguish between material and physical objects.
4. I can phrase my claim using substance terminology as, for the purposes of this chapter, any view that claims that the human substance is a material/physical substance is a physicalist view, even if that view asserts that the physical substance has nonphysical properties.
5. For a critique of some alternatives available through metaphysical schemes like perdurantism, see Flint (2011b, 72–79).
6. Merricks would not deny that the incarnate SPT *has* the divine nature, but he would deny that this is any sort of substance. Rather, if it is anything, it is some sort of abstract object that SPT *exemplifies*, but is not a part or constituent of SPT.
7. This is the time during which SPT was incarnate but prior to his ascension to glory.
8. Christ's death presents difficulties as well (see McCall 2015, 210–211).
9. Merricks rejects views like Baker's in his *Objects and Persons*, so this part of Cross's defense is unavailable to him.
10. Baker (2007, 167–168). For her formal account of the constitution relation itself, see pp. 161–165.
11. Baker insists that having a property derivatively is to really have the property (Baker 2007, 169), but this appears to just be an assertion. The property of *being related to something else that has property p* does not entail *being p*.
12. Dean Zimmerman has for these sorts of reasons charged that Baker's metaphysics is really a form of substance dualism:

This reinforces my suspicion that her view is dualism-in-disguise. Organisms and aggregates of matter cannot, presumably, lose all of their physical parts at once; and there are limits on the ways in which the subsequent physical states of organisms and aggregates may evolve out of earlier ones. Baker's persons are free of such constraints. They can, miraculously, jump from one body to another, losing the shape and size and so on of the one body, and instantaneously acquiring those of the other, whatever they might be. Not even a miracle could allow mere hunks of matter or organisms to perform such feats. I would say that, if the current size and shape and physical makeup of an object puts no necessary constraints upon the immediately subsequent size and shape and physical makeup of that object, then the object does not really have that size, shape, or makeup – however appropriate it is to ascribe them to it in ordinary contexts on the basis of relations to things that really have them . . . She thereby becomes a Cartesian dualist with a complicated theory of the relation (“constitution”) that unites nonphysical persons and bodies. (Zimmerman 2004, 340–341)

13. The next few paragraphs are adapted from Van Horn (2010, 339–340).
14. Cross appears to attempt to deny that SPT is simple prior to the Incarnation, claiming that “Any Chalcedonian Christology has to suppose that no divine person is necessarily simple.” However, his explanation of this comment is that “any divine person is a composite of essence and personal property.” This is an equivocation on the term “composite,” since essences and personal properties, as Cross makes clear, are not parts of the divine persons in the mereological sense of “part” (Cross 2003, 303–304).
15. Is it even an empirical fact that there is a single moment at which parts unite to form a new whole, such as a new human organism (whether at conception or later)?
16. If one is inclined, as I am, to endorse mereological nihilism (the claim that there are no composite objects), then this provides even further reason to reject AC. While few have endorsed this view, there are powerful arguments in its favor, many provided by Merricks himself (he declines to apply his arguments to human organisms, but they apply there just as well as they do to pools of water and baseballs). If there are no composite objects, then AC is ruled out automatically. See Merricks 2001, 1–83; van Inwagen 1990, 21–80; Rea 2001, 134–141.
17. Timothy O'Connor and Philip Woodward, both materialists, endorse this model (O'Connor and Woodward 2014, 232–234).
18. For versions of this argument, see Le Poidevin 2009a, 177–180; Flint 2011b, 72–79. Flint considers the standard responses to the paradox of increase, such as co-located objects and denying DAUP, but finds them problematic when applied to the Incarnation.
19. Perhaps this is a trivial truth, since I recognize this may be a counter-possible, for it may not be possible for HO's complement (W) to cease to exist (I only say “may be,” since it's not clear to me what W could be, and thus not clear whether or not it could cease to exist).

20. It is not clear whether or not Peoples endorses Model T. He clearly proposes a materialist version of concretism, but he never clearly specifies if HO is a part of SPT or whether they are both proper parts of a larger whole (a view labeled by Flint as "Model A"). Nevertheless, it is clear that a proponent of Model T could pursue the strategy Peoples offers (see Peoples 2015, 336–338; Flint 2011b, 79).
21. Perhaps Peoples is thinking that being a person is partially dependent upon external relations, such as standing in the assumption relation to SPT (basically Flint's strategy). This would allow HO to be intrinsically just like you and me yet fail to be a person. However, this is a mere technicality based upon an ad hoc analysis of personhood. It would allow advocates of Model T to avoid the label of Nestorianism, but their model would still contain all of the objectionable elements of the heresy. As Leftow explains, "it is not hard to see why this should be a heresy: if [SPT] is one person and Jesus is another, different person, God did not become man in Jesus, but instead merely entered some sort of intimate association with a man." This conclusion is not avoided by pointing out that HO was "gotten to" and thus stands in a relation that technically disqualifies him from being a person, while being in every way like the objectionable distinct man Leftow refers to. If HO has a mental life like yours and mine, then Model T is guilty of Nestorianism or something near enough (Leftow 2011, 30; Flint 2011a, 202–205).
22. J. H. W. Chan classifies this as the "relational view" of souls (Chan 2015, 356–357).
23. For this view to have a plausible analysis of Christ's incarnate mental life (at least prior to his ascension), it may have to be wedded to a hypothesis such as Andrew Loke's "Divine Preconscious Model" (Loke 2009), wherein some of Christ's knowledge is confined to his preconscious. He would be omniscient, but some knowledge would not be consciously occurrent. This can be given a kenotic spin, but that is unnecessary if the knowledge of nonincarnate divine persons is not constantly occurrent. I see no good reason to believe that God is always thinking about everything.
24. See, for example, Collins (2008); Lycan (2009); Plantinga (2007); Bailey, Rasmussen, and Van Horn (2011).
25. The next paragraph adapts material from Van Horn (2010, 334–335).
26. This account is inspired by Peter Unger, but modified to handle possible counterexamples involving paralyzing drugs (Unger 2006, 456–460).
27. For further defenses of dualist models like mine against the Apollinarian charge, see DeWeese (2007, 146–148) and Crisp (2009, 50–61). Crisp objects that, while models like mine do not have to be Apollinarian, they do require Monothelitism, also heretical. Space restraints preclude an adequate response to this objection, but my response would be two-pronged: (a) It isn't clear that my model entails Monothelitism; (b) even if it did, this isn't particularly worrisome, since I am a Protestant. For fuller discussion see Plantinga (1999, 185); DeWeese (2007, 148–149); and Wessling (2013).
28. I am grateful to Dallas and Lynne van Horn, familial proofreaders than which none greater can be conceived.

## References

- Bailey, Andrew M., Joshua Rasmussen, and Luke van Horn. 2011. "No Pairing Problem." *Philosophical Studies*, 154: 349–360. DOI: 10.1007/s11098-010-9555-7.
- Baker, Lynne Rudder. 2007. *The Metaphysics of Everyday Life: An Essay in Practical Realism*. Cambridge: Cambridge University Press.
- Baker, Lynne Rudder. 2011. "Christian Materialism in a Scientific Age." *International Journal of Philosophy of Religion*, 70: 47–59. DOI: 10.1007/s11153-010-9283-0.
- Burke, Michael B. 2003. "Is My Head a Person?" In *On Human Persons*, edited by Klaus Petrus, 107–125. Frankfurt: Ontos.
- Chan, J. H. W. 2015. "A Cartesian Approach to the Incarnation." In *The Ashgate Research Companion to Theological Anthropology*, edited by Joshua R. Farris and Charles Taliaferro, 355–367. Burlington, VT: Ashgate.
- Collins, Robin. 2008. "Modern Physics and the Energy-Conservation Objection to Mind-Body Dualism." *American Philosophical Quarterly*, 45: 31–42.
- Corcoran, Kevin J. 1999. "Persons, Bodies, and the Constitution Relation." *The Southern Journal of Philosophy*, 37: 1–20. DOI: 10.1111/j.2041-6962.1999.tb00854.x.
- Crisp, Oliver D. 2009. *God Incarnate: Explorations in Christology*. New York: T&T Clark International.

- Cross, Richard. 2003. "Incarnation, Omnipresence, and Action at a Distance." *Neue Zeitschrift für Systematische Theologie und Religionsphilosophie*, 45: 293–312. DOI: 10.1515/nzst.2003.020.
- DeWeese, Garrett J. 2007. "One Person, Two Natures: Two Metaphysical Models of the Incarnation." In *Jesus in Trinitarian Perspective*, edited by Fred Sanders and Klaus Issler, 114–153. Nashville, TN: B&H Publishing.
- Flint, Thomas P. 2011a. "Molinism and Incarnation." In *Molinism: The Contemporary Debate*, edited by Ken Perszyk, 187–207. Oxford: Oxford University Press.
- Flint, Thomas P. 2011b. "Should Concretists Part with Mereological Models of the Incarnation?" In *The Metaphysics of the Incarnation*, edited by Anna Marmodoro and Jonathan Hill, 67–87. New York: Oxford University Press.
- Hasker, William. 2010. "On Behalf of Emergent Dualism." In *In Search of the Soul: Four Views of the Mind-Body Problem*, edited by Joel B. Green, 75–100. Eugene, OR: Wipf and Stock.
- Hill, Jonathan. 2012. "Incarnation, Timelessness, and Exaltation." *Faith and Philosophy*, 29: 3–29. DOI: 10.5840/faithphil20122911.
- Hughes, Christopher. 2009. "Defending the Consistency of the Doctrine of the Trinity." In *Philosophical and Theological Essays on the Trinity*, edited by Thomas McCall and Michael C. Rea, 293–313. Oxford: Oxford University Press.
- Leftow, Brian. 2002. "A Timeless God Incarnate." In *The Incarnation*, edited by Stephen T. Davis, Daniel Kendall, and Gerald O'Collins, 273–299. Oxford: Oxford University Press.
- Leftow, Brian. 2011. "The Humanity of God." In *The Metaphysics of the Incarnation*, edited by Anna Marmodoro and Jonathan Hill, 20–44. New York: Oxford University Press.
- Leftow, Brian. 2015. "Against Materialist Christology." In *Christian Philosophy of Religion: Essays in Honor of Stephen T. Davis*, edited by C. P. Ruloff, 65–94. Notre Dame, IN: University of Notre Dame Press.
- Le Poidevin, Robin. 2009a. "Identity and the Composite Christ: An Incarnational Dilemma." *Religious Studies*, 45: 167–186. DOI: 10.1017/S003441250800975X.
- Le Poidevin, Robin. 2009b. "Incarnation: Metaphysical Issues." *Philosophy Compass*, 4: 703–714. DOI: 10.1111/j.1747-9991.2009.00222.x.
- Loke, Andrew. 2009. "On the Coherence of the Incarnation: The Divine Preconscious Model." *Neue Zeitschrift für Systematische Theologie und Religionsphilosophie*, 51: 50–63. DOI: 10.1515/NZST.2009.004.
- Lycan, William G. 2009. "Giving Dualism Its Due." *Australasian Journal of Philosophy*, 87: 551–563. DOI: 10.1080/00048400802340642.
- Marmodoro, Anna, and Jonathan Hill. 2008. "Modeling the Metaphysics of the Incarnation." *Philosophy & Theology*, 20: 99–128. DOI: 10.5840/philtheol2008201/25.
- McCall, Thomas H. 2015. "'I Am My Body?' Physicalism, Identity, and the Metaphysics of the Incarnation." *Philosophia Christi*, 17: 205–211.
- Merricks, Trenton. 2001. *Objects and Persons*. Oxford: Clarendon Press.
- Merricks, Trenton. 2007. "The Word Made Flesh: Dualism, Physicalism, and the Incarnation." In *Persons: Human and Divine*, edited by Peter van Inwagen and Dean Zimmerman, 281–300. Oxford: Clarendon Press.
- O'Connor, Timothy, and Philip Woodward. 2014. "Incarnation and the Multiverse." In *God and the Multiverse: Scientific, Philosophical, and Theological Perspectives*, edited by Klaas Kraay, 227–241. New York: Routledge.
- Olson, Eric. 2006. "The Paradox of Increase." *The Monist*, 89: 390–417.
- Peoples, Glenn Andrew. 2015. "The Mortal God: Materialism and Christology." In *The Ashgate Research Companion to Theological Anthropology*, edited by Joshua R. Farris and Charles Taliaferro, 331–343. Burlington, VT: Ashgate.
- Plantinga, Alvin. 1999. "On Heresy, Mind, and Truth." *Faith and Philosophy*, 16: 182–193. DOI: 10.5840/faithphil199916221.

- Plantinga, Alvin. 2007. "Materialism and Christian Belief." In *Persons: Human and Divine*, edited by Peter van Inwagen and Dean Zimmerman, 99–141. Oxford: Clarendon Press.
- Rea, Michael C. 2001. "How to Be an Eleatic Monist." *Philosophical Perspectives*, 15: 134–141. DOI: 10.1111/0029-4624.35.s15.7.
- Rea, Michael C. 2003. "Relative Identity and the Doctrine of the Trinity." *Philosophia Christi*, 5: 431–445.
- Taliaferro, Charles, and Stewart Goetz. 2008. "The Prospect of Christian Materialism." *Christian Scholar's Review*, 37: 303–321.
- Unger, Peter. 2006. *All the Power in the World*. Oxford: Oxford University Press.
- Van Horn, Luke. 2010. "Merricks's Soulless Savior." *Faith and Philosophy*, 27: 330–341. DOI: 10.5840/faithphil201027333.
- van Inwagen, Peter. 1981. "The Doctrine of Arbitrary Undetached Parts." *Pacific Philosophical Quarterly*, 62: 123–137.
- van Inwagen, Peter. 1990. *Material Beings*. Ithaca, NY: Cornell University Press.
- van Inwagen, Peter. 1995. "Not by Confusion of Substance, but by Unity of Person." In *Reason and the Christian Religion: Essays in Honor of Richard Swinburne*, edited by Alan G. Padgett, 201–226. Oxford: Clarendon Press.
- van Inwagen, Peter. 1998. "Incarnation and Christology." In *Routledge Encyclopedia of Philosophy*, edited by Edward Craig, vol. 4, 725–732. London: Routledge.
- Wessling, Jordan. 2013. "Christology and Conciliar Authority: On the Viability of Monothelitism for Protestant Theology." In *Christology, Ancient and Modern: Explorations in Constructive Dogmatics*, edited by Oliver D. Crisp and Fred Sanders, 151–170. Grand Rapids, MI: Zondervan.
- Zimmerman, Dean W. 2004. "Reply to Baker." In *Contemporary Debates in Philosophy of Religion*, edited by Michael L. Peterson and Raymond J. VanArragon, 338–341. Malden, MA: Blackwell.



# The Word Made Flesh

## *Dualism, Physicalism, and the Incarnation*

TRENTON MERRICKS

The Incarnation is beautiful and mysterious, awe-inspiring and humbling. The metaphysics of “embodiment” is, in comparison, drab and petty. And so a chapter on the Incarnation that focused on God the Son’s relation to his body would be like a chapter on the history of music that focused on the kazoo. Nevertheless, I shall ask: “How is the incarnate God the Son related to his body?”

This is not the most important question about the Incarnation. Nor is it a traditional question. For example, it is not (at least not obviously) a question about human nature or a question about divine nature or even a question about the union of the two. It is not a question explicitly addressed in creedal discussions of the Incarnation. And so, if (for example) this chapter were an exegesis of the Chalcedonian Definition, I might have no business asking this question, let alone answering it. But, though I intend to stay within the definition’s parameters, such exegesis is not my project.

My project starts by assuming that God the Son, in virtue of being incarnate, is related to his body just as you and I are related to our respective bodies. This assumption opens up a way to explore the Incarnation, a way in addition to examining the creeds. For, given this assumption, one’s view on how each of us is related to his or her body should dictate one’s view on how the Son is related to his body. Conversely, if an account of embodiment is untenable in the case of the Son and his body, it is untenable in our case as well. And so my starting assumption opens up not just a way to investigate the Incarnation, but also an Incarnation-based way to investigate the relation between person and body in general.

Although controversial, my starting assumption is quite plausible. For the incarnate Son is fully human and so, presumably, human in the same sense that you and I are. Part of being human, at least in this life, is having a body. And so, presumably, the Son has a body in the same sense that you and I do. More generally, he is related to his body just as each and every other human is related to his or her body.<sup>1</sup>

### 30.1 Substance Dualism

The word “soul” can be used in many ways. Throughout this chapter, I shall use “soul” to mean a thing or object or substance that has mental properties but lacks physical properties. Because a soul has mental properties, a soul can believe that the sun is shining, hope that rain will come soon, and be appeared to redly. Because it lacks physical properties, a soul has no mass, fails to be extended in space, and reflects no light.

Substance dualists (hereafter “dualists”) believe in souls. Indeed, dualists say that each human person just is – is identical with – a soul.<sup>2</sup> Obviously, if we are souls, we are not bodies. Nor do we have bodies as parts. Nevertheless, even dualists believe that, in this life at least, we “have bodies.”

According to the typical dualistic picture, a soul’s having a body is partly a matter of that soul’s having direct causal control over that body. For example, when I – suppose I am a soul – intend that my left arm move, the left arm of “my body” moves. Of course, I can indirectly cause things outside of my own body. I could indirectly cause your arm to rise by my lifting your arm with my hand. But this is not a case of direct causal control, since I cannot make your arm rise simply by intending that it does.<sup>3</sup>

Given typical dualism, the union of soul and body is partly constituted by the soul’s having direct control over the body. In addition, the union of soul and body involves the body’s influencing the soul. For example, when a blue piece of paper is in front of a body – a body with its eyes open, in plenty of light, and so on – that body causes “its soul” to know that a blue piece of paper is located there.

Dualists are explicit about all this. Thus Richard Swinburne:

A person has a body if there is a chunk of matter through which he makes a difference to the material world, and through which he acquires true beliefs about that world. Those persons who are men have bodies because stimuli landing on their eyes or ears give them true beliefs about the world, which they would not otherwise have; and they make differences to the world by moving arms and legs, lips and fingers. Our bodies are the vehicles of our knowledge and operation. The “linking” of body and soul consists in there being a body which is related to the soul in this way. (Swinburne 1986, 146)

According to Swinburne, a person’s standing in the appropriate causal relations to a body – or a “chunk of matter” – is all there is to a soul’s having a body.<sup>4</sup> Given this picture, embodiment is a cluster of relations, all of which involve some sort of epistemic access or direct control. The more of these embodiment-constituting relations that hold between an immaterial person and a body, the “more embodied” that person is (see Swinburne 1986, 151). And so, according to this picture, embodiment can come in degrees. (For example, a soul could leave its body gradually, as one and then another embodiment-constituting relation “shuts off.”)

I take this to be the most natural dualistic account of embodiment. Below I shall develop objections to this account. I shall then consider other accounts, including even one that says that *having a body* is primitive and unanalyzable.<sup>5</sup> But my starting point is Swinburne’s account, which I believe reflects the most familiar dualistic understanding of having a body.

### 30.1.1 *The soul's influence on a body*

The typical dualist says that a person's *having a body* just is her standing in the relevant causal relations to that body. This leads to the most familiar objection to dualism. As Daniel Dennett says:

The standard objection to dualism was all too familiar to Descartes himself in the seventeenth century, and it is fair to say that neither he nor any subsequent dualist has ever overcome it convincingly. If mind and body are distinct things or substances, they nevertheless must interact; the bodily sense organs, via the brain, must *inform* the mind, must send to it or present it with perceptions or ideas or data of some sort, and then the mind, having thought things over must *direct* the body in appropriate action (including speech) . . . but anything that can move a physical thing is itself a physical thing. (Dennett 1991, 33–35)

There are a number of replies the dualist could make to the “standard objection.” I shall focus on one that is particularly fitting in the present context. This chapter explores how a commitment to the Incarnation bears on how we understand embodiment. (It also explores, conversely, how theories of embodiment bear on our understanding of the Incarnation.) The Incarnation entails theism. But theism – with its nonphysical miracle-working creator God – entails that the nonphysical can causally influence the physical. So, given the Incarnation, the standard objection to dualism ought to be judged unconvincing.

The Incarnation helps the dualist out of a familiar problem, providing a decisive reason to reject the premise that the physical and the nonphysical cannot causally interact.<sup>6</sup> But in accepting this help, the dualist takes a poisoned pawn. For, as I shall argue throughout much of this chapter, the Incarnation threatens to undermine the dualist's notion of embodiment, thereby undermining dualism itself.

Consider, for example, dualism's claim that having a body is partly a matter of having direct control of a body. Thus, God the Son's having the body of Jesus is partly constituted by his having direct control of that body. But that implication of dualism, and so dualism itself, seems to be mistaken. For the Son's being embodied cannot be partly constituted by his having such control over the body of Jesus, lest to that same extent he – along with the Father and the Spirit – have every body that ever was. After all, each divine person, being omnipotent, has direct control over each and every body.<sup>7</sup>

The dualist might reply that while God *has* direct control of each human body, God does not *exercise* such control.<sup>8</sup> (Of course, God continually sustains everything in existence, but that is another matter.) For example, while God could raise my arm simply by intending that it rise, God does not do so, at least not typically. On the other hand, God the Son regularly exercises direct control over the body of Jesus. This avoids the above objection, this reply concludes, because a soul's having a body is not (partly) a matter of that soul's merely *having* direct control of that body, but is instead (partly) a matter of that soul's *exercising* such control.

This reply makes embodiment a matter of a soul's exercising control over a body. And so it implies that whenever one is not intending bodily actions, one is not embodied, or at least not embodied to the extent that embodiment is a matter of the soul's influence on the body. But that implication cannot be right. For my failing to intend bodily actions does not render me totally disembodied. Nor does it even render me somewhat less embodied than I would otherwise be. After all, embodiment does not wax and wane with everyday occurrences, such as my now intending to raise my arm, my now failing to intend any bodily action at all.

For these reasons, I conclude that embodiment is not even partly a matter of the *exercise* of direct control. Rather, insofar as the soul's influence on the body is concerned, embodiment is a matter of the soul's *having* direct control. This allows one to be fully embodied even when intending no bodily actions. (And it implies, quite plausibly, that embodiment is the precondition for, rather than the result of, exercising direct control.) Of course, this returns us to the problem already noted. Insofar as embodiment is having direct control, each divine person is thereby embodied in each human body. Indeed, it seems that each divine person is thereby embodied in each physical object.

### 30.1.2 *The body's influence on a soul*

A soul's having a body is not merely its having direct control of that body. That is only one "direction" in the embodiment equation. The other "direction" involves the body's influence on the soul. As Swinburne says above, a person's body is that physical object "through which he acquires true beliefs about the world."

God has direct and immediate knowledge of everything in and around every body. And so insofar as having a body is having knowledge of what is in and around that body, each person of the Trinity has each and every body. This is all by itself bad enough. And it threatens the Incarnation. For, to the extent that the Son has every body, he does not have the body of Jesus in particular.

I think this point is basically correct. But, to be compelling, it needs to be developed further. Moreover, it is open to more objections than the previous point, the point that the Son's omnipotence gives him the sort of direct control over each and every body that is – according to the dualist – the other "half" of embodiment. So let me consider some objections and offer some replies while, at the same time, clarifying the dualist's problem.

*Objection 1:* Each person of the Trinity is omniscient. Thus each knows "everything." Nevertheless, each can know some things the others do not. For example, only the Father knows "I am the Father." More to the point, only the Son knew "I am walking on water." The Son's having the body of Jesus is partly constituted by his knowing such things, things which are appropriately correlated with that body.

*Reply:* The dualist denies that human persons are bodies. She denies that human persons have physical parts, such as feet. As a result, she must say that when the Son truly thinks "I am walking on water," this is a shorthand way of thinking "my body is walking on water." This in turn is shorthand for "the body of Jesus is walking on the water and the body of Jesus is my body." Such beliefs presuppose that the Son has the body of Jesus. (In this regard, they are on a par with the Son's belief "I have the body Jesus.") Thus, they cannot even partly constitute his having that body.

*Objection 2:* It is one thing to know something. It is another for that knowledge to be caused by a body. Although each person of the Trinity knows what is happening in and around each body, it is not because events in that body cause this knowledge in God. And so it is false that insofar as embodiment involves a body's causing knowledge in a person that the Father, the Son, and the Spirit thereby have each and every body.

*Reply:* Consider the Platonic claim that God is self-sufficient and unchanging, thus not possibly influenced by goings-on in the physical world. If this claim were correct, bodily events could not cause knowledge in God. And so the above objection would stand. But dualistic Incarnation would not. For God the Son could not be causally influenced by what goes on in the body of Jesus. And so, at least insofar as embodiment is a matter of the body's

influence on the person, God the Son could not have the body of Jesus. Therefore, in order to give the dualist a fighting chance, I shall reject this Platonic picture of God.<sup>9</sup>

Besides, I really do think this picture is mistaken. I assume that, typically, God knows something is happening because it is happening, and not the other way round. God knows what is happening in my body because it is happening there. Moreover, God knows that when particular experiences in my body are caused in particular ways, certain things are happening in the world around that body. Thus, God knows about goings-on in the world because of events in my body. (Of course, God also knows about those goings-on directly.) So it seems that events in my body cause knowledge of the world in each person of the Trinity. At any rate, it is hard to see a principled way of ruling out causation in this case without thereby ruling out something to which the Christian dualist is committed – events in the body of Jesus causing knowledge in God the Son.

*Objection 3:* A body does not deliver only *propositional* knowledge to its soul. A body also provides *sensory experiential* knowledge. Such knowledge – for example, knowledge of what it is like to see a red-tailed hawk – essentially involves sensory experience, which in turn essentially involves having a body (see Aquinas, *Summa Theologica*, 1a.77.8). God the Father and God the Spirit, lacking bodies, lack sensory experiential knowledge. God the Son, however, has sensory experiential knowledge. And we can parlay this knowledge into a way in which the Son is uniquely related to the body of Jesus, a way that at least partly constitutes the Son's having that body.

*Reply:* Dualists typically endorse the possibility of my having no body, yet everything's seeming to me just as it actually does. (Thus begins Descartes's famous argument for dualism in the *Meditations*.) And so dualists typically think that, possibly, a disembodied immaterial being has sensory experiences. As a result, dualists should say that such a being can have experiential knowledge and, therefore, they should not endorse the above objection.

Besides, presumably, the omniscient God's knowledge of creation is not far poorer than ours. And so, presumably, each person of the Trinity knows, for example, what a red-tailed hawk looks like and what *Eine kleine Nachtmusik* sounds like. Indeed, insofar as we dare speculate on such a thing, I would say that God's knowledge is so rich that each divine person knows what it is like to have your body, what it is like to have the body of Jesus, and even what it is like to be a bat.

*Objection 4:* It was a mistake to focus on sensory experiential *knowledge*. Let's consider, instead, sensory *experiences* (see Cross 2003, 301). To see the distinction, consider that you may not now see anything red, though you nevertheless now know what it is like to see red. Seeing red is one thing; knowing what it is like to see red is another. In general, it is one thing to know what an experience is like and quite another to have that experience.

To have a body is to have sensory experiences caused by that body. For example, if a body's eyes are open and a sheet of red paper is held in front of it, then that body may cause "its soul" to see red, and to see it "from the perspective of" that body. God the Son's experiences are caused in this way by exactly the body of Jesus and no other. (God the Father and God the Spirit may not have sensory experiences, and even if they do, their experiences are not caused by a body.)

*Reply:* This reply makes embodiment a matter of a body's actually causing a person to have experiences. And so it implies that whenever one is not having experiences caused by a body, one is not embodied. But I object that my body's failing to cause experiences in me should not render me disembodied. My soul might leave my body when I die, but not when

I dreamlessly sleep! Moreover, lack of experiences caused by a body should not even result in my being somewhat less embodied. To repeat an earlier point, embodiment does not wax and wane with changes of the sort we encounter every day. Thus, I conclude that embodiment is not even partly a matter of a body's actually causing experiences in a soul. (Presumably, embodiment is instead a precondition for having experiences caused by a body.)<sup>10</sup>

### 30.1.3 *Kenosis and embodiment*

The most familiar and straightforward dualistic account of embodiment says that to have a body is to have direct control over, and epistemic access to, that body. But this account stumbles over theism and the Incarnation. So let's consider another account.

Suppose that standing in the relevant relations of control and access to a body is not what it is to have that body. Rather, the dualist might say, to have a body is both to stand in those relations to that body and *to fail to stand in those relations to any other body*. Embodiment, thus understood, is not merely a matter of being "positively" related to a body. It is also a matter of being appropriately limited. It is a matter of lacking control over, and epistemic access to, any other body.

The Father and the Spirit, being omnipotent and omniscient, are not appropriately limited. That is, they stand in the relevant relations of control and epistemic access to each and every body. So this "revised account of embodiment" keeps the Father and Spirit from having bodies. But – by the same token – it robs the Son of his body.

More carefully, this revised account robs the Son of his body if he too is omnipotent and omniscient. But suppose the Incarnation involved a "kenosis." According to Gottfried Thomasius – whose *Person and Work of Christ* (1852–1861) contains the first explicit defense of a kenosis – Christ abandoned his divine attributes from birth until resurrection (see McGrath 2001, 377–378). Given a kenosis, the revised account of embodiment might allow the Son to have a body. For, given a kenosis, the Son might stand in the relevant relations of control and epistemic access to only the body of Jesus.

Without a kenosis, the revised account makes it impossible for the Son to have a body. So, given that the Son came to have the body of Jesus, one cost of the revised account is a kenosis. I shall not raise any objections to a kenosis.<sup>11</sup> Nevertheless, I have three objections to the revised account of embodiment.

The revised account implies that no omnipotent and omniscient person can have a body.<sup>12</sup> Given a kenosis, the Lord emptied himself of omniscience and omnipotence at birth and regained these attributes at resurrection. Thus the revised account, combined with a kenosis, has the comic and absurd implication that, upon the resurrection of his body, the Son became disembodied. This – along with the revised account's implying that the Son is forevermore disembodied – is the first reason to reject the revised account.

Suppose a soul starts with a single body and then acquires the appropriate control over and knowledge via a second body, which results in the soul's having two bodies. This idea – one soul's having two bodies – seems possible. Yet the revised account of embodiment renders that idea impossible. (According to that account, a soul has a body if and only if that soul stands in the relevant relations to that body and *fails to stand in those relations to any other body*.) This is the second reason that the dualist should reject the revised account.

The revised account allows the Father to be related to the body of Jesus in each and every way that the Son is, while insisting that the Son, but not the Father, has that body.

(According to that account, the Father does not have that body because of how he is related to other bodies.) But, I object, the union of a person and a body ought to be wholly a matter of their relations to each other, not instead partly a matter of how they fail to be related to other things. The revised account does not respect this. This is the third reason to reject that account.

### 30.1.4 *Relations R and X*

The dualist should reject the revised account of embodiment. So let's return to the original account. That account says that to have a body is to have direct control of, and to enjoy epistemic access to, that body. I have argued that the most obvious and plausible ways of spelling out *direct control* and *epistemic access*, when combined with the dualist's account of embodiment, get the "wrong results." (The wrong results include, among others, your and my lacking bodies when intending no actions and the Spirit's having each and every body for his own.) Of course, I have not examined every possible candidate for what control or access might amount to. And so one might fear that I have overlooked a candidate that gets "all the right results."

So, for the sake of argument, let's grant that there is a relation – call it "Relation R" – that gets all the right results. That is, R holds between God the Son and only the body of Jesus; R fails to hold between any body and any other divine person; R holds between each of us and exactly each of our respective bodies; and, finally, R is intuitively embodiment-constituting because it is a kind of direct control, a kind of epistemic access, or a combination of both.

Suppose that R, which is an embodiment-constituting relation, holds between the Son and the body of Jesus alone. Even so, the moral of the chapter so far is that many more embodiment-constituting relations hold between the Son and each and every body. Thus the dualist cannot say both that God the Son is fully and completely embodied in the body of Jesus and also that God the Son has no other body at all, without qualification. At best, even granting relation R, the Son might be "slightly more embodied" in the body of Jesus than he is in your body or in my body or in a teacup. But that's not good enough. (Moreover, even granting R, the Father and the Spirit are only "slightly less embodied" than you or I.)

This objection presupposes that embodiment is a cluster of relations, only one of which is relation R. The dualist might, however, reject this presupposition. She might say, instead, that embodiment *just is* relation R. Since we have stipulated that R gets "all the right results," this account of embodiment gets the right results as well.

We don't yet know what relation R is, other than that it will be some variety of direct control and/or epistemic access. Insofar as we do not know what R is – given the identification of R with embodiment – we do not know what embodiment is. Thus this suggestion renders embodiment somewhat mysterious.

Moreover, until we know more about R, we have no reason to think that R is intrinsically any more suited to be *the* embodiment relation than any other relation of epistemic access or direct control. The claim that R *just is* embodiment therefore privileges, in an ad hoc manner, exactly one out of many relations, all of which intuitively constitute embodiment. It would be better to take R as a crucial ingredient of embodiment than to take it to be the whole shooting match.

The dualist might deny that embodiment is reduced to, or analyzed in terms of, any apparently embodiment-constituting relations, including R. That is, she might deny that embodiment is analyzed in terms of any relation of direct control or epistemic access. She

might say, rather, that embodiment is a primitive, unanalyzable relation that holds between a person and a body. This too would allow the dualist who believes in the Incarnation to get all the “right results,” simply by *fiat*.

But this move, even more than the previous, renders embodiment an occult relation. For once we make this move, then we do not know what the dualist means when she says that each of us “has a body” (she does not mean that we have direct control of a body; she does not mean that we have epistemic access to a body; she does not mean that we have a combination of control over and access to a body, and so on). To simply assert that embodiment is some “relation X” makes embodiment completely mysterious and so is utterly implausible.

Moreover, for all we have said so far, it is possible for me to stand in all the seemingly relevant control and epistemic relations to a body without being X-related to it. Conversely, I could be X-related to a body without being related to it by any control or epistemic relations. *Ex hypothesi*, something is my body if and only if I am X-related to it. And so – at least for all we have said so far – my body might be, for example, the one typically believed to belong to Queen Elizabeth II, the body now in Buckingham Palace. But no account of embodiment should make it possible for that to be my body and not that of Her Majesty, given that HM has (and I lack) causal control over, and epistemic access to, that body.

The defender of relation X could embellish her account to rule out such absurdities. She could insist that although X is not reduced to the relevant relations of control and access, each of these relations – including R, so that we get the cases right – is necessary for X’s holding and all of them together are sufficient.

Embellishing the X-account in this way has three advantages. First, there is no chance I have the Queen’s body. Second, if X supervenes on relations of control and access, it may not be so mysterious after all. Finally, this account allows one to deny that embodiment comes in degrees. And so its defenders can resist some of my earlier objections. To take just one example, even if God the Father is related to my body by every apparently embodiment-constituting relation except for R, it is false – given the embellished X-account – that there is a degree to which he *has* my body.

Let’s focus on this third advantage. The embellished X-account tells us that a soul, standing in all the relevant relations to a body except for R, is absolutely disembodied. This raises the troubling thought that – even though I control a body through which I have knowledge of the world – I might actually be disembodied. This troubling thought is a mere symptom of the real problem. The real problem is that this account puts too much weight on R.

This account says that a soul related to a body by every relevant relation of epistemic access and control except for R is totally disembodied. But I reply that, if I am a soul related to a body by all the relevant relations save R, then surely I am embodied to a significant degree, though perhaps not as embodied as other slightly more plugged-in souls. The X-account – even when embellished – is not plausible. And, besides, we have no reason to think that there really is any relation R, any apparently embodiment-constituting relation that really does get “all the right results.”<sup>13</sup>

### 30.1.5 *Embodiment and incarnation*

For the sake of argument, let’s assume that the dualist can handle the objections above. Let’s assume, in particular, that she can account for the Son’s having exactly the body of Jesus. Even so, I shall argue, the Incarnation casts doubt on dualism.



This second Incarnation-based objection to dualism begins by considering “Apollinarianism.” Here is how Peter van Inwagen describes this heresy:

*Apollinarianism* (after Apollinarius, ca. 310–ca. 390) holds that Christ did not have a human mind or spirit or rational soul – that he lacked something that is essential to human nature – and that God or some “aspect” of God (such as the divine *Logos*) was united to the human body of Jesus of Nazareth in such a way as to “be a substitute for” or perform the function of the human mind or soul or spirit. (van Inwagen 1998, 727)

Regarding the heresy of monophysitism, Swinburne says:

Monophysitism, holding that the Incarnate Christ had only one nature, normally understood that to be the divine nature . . . He had a human body; and the connection with [that body] that leads to the sensory desires – pain, thirst, etc. So this is not Docetism, the view that Christ’s body was mere “appearance” and Christ did not really suffer. But it is what the century before Chalcedon knew as Apollinarianism, the view that the Incarnation consisted in the Word of God acquiring a human body but not a “rational soul.” (Swinburne 1994, 224)

With all this in mind, let us turn to:

*The heretical theory:* God the Son is fully divine. But he is not fully human. Nevertheless, ever since the virgin conception and birth over two thousand years ago, he has been related to the body of Jesus just as a normal human soul is related to its body. So God the Son controls the body of Jesus. Moreover, he knows what happens in and around that body. He even has experiences such as hunger and pain and seeing red caused by that body.

Arguing that this or that metaphysics of the Incarnation is heretical can be tricky business (see Plantinga 1999). Nevertheless, it is safe to say that the Heretical Theory is aptly named. For the heretical theory explicitly asserts that God the Son is not fully human. This is a failure of doctrine. It is not, however, a failure of logic. That is, there is nothing incoherent in the claim that a nonhuman divine person is related to a human body in the ways a normal soul is related to its body. God the Son could “play the role” with respect to a body that, according to dualism, is typically played by a human soul, and the Son could do this without thereby becoming human.

So one moral of the heretical theory is that having a body, as understood by the dualist, is not sufficient for being human. Nor would it seem to be necessary. For dualists typically allow that you and I can continue to exist – and continue to be human – after our body dies, even before resurrection. Thus, given dualism, having a human body seems to be neither necessary nor sufficient for being human.<sup>14</sup>

The Incarnation is the Son’s becoming human. Given dualism, this cannot be a matter of the Son’s coming to have a human body. So the dualist must say that the Son, in addition to coming to have a body, also became human. I suppose that, for the dualist, to be human is to be a human soul. So the dualist must claim that the Son, while remaining divine, became a human soul.

It is neither incoherent nor obviously heretical to say that God the Son’s becoming human just is his becoming a human soul. But I would prefer an account of the Incarnation according to which the Son’s coming to have a human body is at least a necessary condition for his becoming human. Dualism, as we have seen, is not such an account. Dualism makes

the Son's becoming human one thing and his becoming embodied something else altogether. This is my second Incarnation-based objection to dualism.

Besides, whether or not embodiment is absolutely necessary for the Incarnation, God the Son does have the body of Jesus. And the fact that the Son has a body – and the Father and the Spirit do not – undermines the standard dualistic account of embodiment, a straightforward account in terms of knowledge and control. Given theism and the Incarnation, the dualist must exchange the straightforward account for something or other ad hoc or implausible or darkly mysterious. (Among the unattractive options are that embodiment turns completely on “relation R” and that embodiment wanes when sleeping.) This, of course, was my first, and more important, Incarnation-based objection to dualism.

In light of this objection, one could conclude that Christians should be dualists who defend one or another ad hoc or implausible or darkly mysterious claim about embodiment. Similarly, in light of my second objection, one could conclude that the Son could have become human without ever having a body. But, rather than jump to these conclusions, I suggest that we consider another approach altogether.

### 30.2 Physicalism

It seems pretty obvious that you have physical properties. You have a height and a weight; you take up space; you have a shape. But only physical objects have physical properties. For to be a physical object just is to be a thing that has physical properties.<sup>15</sup> Given all this, it seems pretty obvious that *physicalism* – the claim that each of us is a physical object – is true.

(Not everyone will agree. Dualists do not think it is obvious that we have physical properties. Indeed, dualists think we lack such properties. For each of us, according to the dualist, is a soul. And souls have no physical properties. Souls have neither height nor weight, shape nor size.)

Physicalism says that we are physical objects. Consider the human-shaped object sitting in your chair and wearing your clothes.<sup>16</sup> Let's call that human-shaped and living and breathing object “your body.” The sort of physicalism I defend says that you are identical with your body (see Merricks 2001a). That is, you are that human-shaped thing sitting in your chair and wearing your clothes. You just are that living, breathing organism.

Physicalism has a straightforward account of embodiment. You have a body if and only if you are identical with that body. I assume that, in the Incarnation, God the Son is related to the body of Jesus just as you and I are related to our respective bodies. So, given physicalism, God the Son, in the Incarnation, is identical with the body of Jesus. That is, in becoming human, he became a body.

Some might object that saying that God the Son became a physical object is deeply inappropriate. Some might object that this is akin to saying that that which spoke the universe into existence is a mere doorstep or a lace doily. But, I reply, to say that the Son became a physical object is just to say that he came to have physical properties. Saying that God the Son became a physical object is no more impious than saying that God came to be such that we could literally touch and see him. The only scandal here is that of the Incarnation itself.

Moreover, the claim that God the Son is identical with the body of Jesus does not mean that God the Son is *merely* a physical object, in the sense that his only properties are

physical. Indeed, the physicalist need not say that any human person, divine or otherwise, is merely a physical object. For while the physicalist says that a human person has physical properties, she does not insist that a human person has only physical properties. Persons also have mental properties. And physicalism, as I shall understand it, is consistent with a physical person's mental properties' being *sui generis*, being irreducible to physical properties.<sup>17</sup> (Thus my sort of physicalism is consistent with "property dualism" about the mental.) Moreover, for all I know, we might have properties that are neither mental nor physical. And the same goes for the incarnate Son. *Being the Lord of Glory* is not obviously a mental property, but it is a safe bet that it is not a physical property either.

The claim that human persons are physical organisms is consistent with a variety of views about mental properties. It is also consistent with a variety of views about the further details of human nature. Consider, for example, Aquinas's view. He denied that a human person is a soul that interacts with a numerically distinct body (see *Summa Contra Gentiles*, II.57). Indeed, according to Aquinas, the substantial form of the person is identical with the substantial form of the person's living body (*Summa Theologica*, Ia.76). And so, according to (at least one way of reading) Aquinas, a human person in this life is identical with a living body. And that is physicalism.<sup>18</sup>

Alvin Plantinga says:

Consider again the doctrine of the Incarnation, that characteristic and nonnegotiable Christian teaching according to which the second person of the Trinity became Incarnate and dwelt among us. As I understand the scripture and the creeds (Nicene, Athanasian, the Chalcedonian formulation), this involves the second person of the Trinity's actually becoming human. The Logos became a human being, acquiring the property necessary and sufficient for being human. Prior to the Incarnation, however, the second person of the Trinity was not a material object, but an immaterial being. If, however, as materialists assert, to be a human being is to be a material object, then the second person of the Trinity must have become a material object. If he has remained a human being, furthermore, he is presently a material object. But then an immaterial being became a material object; and this seems to me to be impossible. It is clearly impossible, I'd say, that the number seven or the proposition that  $7 + 5 = 12$ , or the property of self-exemplification, all of which are immaterial objects, should become, turn into, material objects. It is less clearly impossible, but still impossible, it seems to me, that the second person of the Trinity – that personal being with will and intellect and affection – should turn into a material object. (Plantinga 1999, 186)

Plantinga focuses on the physicalist's account of *being human*. But I have been concerned with her account of *having a body*. (I shall address physicalism and *being human* below.) Nevertheless, if Plantinga is right that a nonphysical person cannot possibly become physical, then the physicalist's account of embodiment rules out the Son's coming to have the body of Jesus.

Brian Leftow considers the version of physicalism, according to which "the body just is the person. On this version, the Son owns [body] B only if the Son *becomes* B; only if an immaterial item becomes material." Like Plantinga, Leftow says: "This does not seem possible" (Leftow 2002, 284). So, like Plantinga, Leftow thinks that physicalism rules out the Son's coming to have a body.<sup>19</sup>

"Kind-essentialism" says that if something is a member of a natural kind, then it is essentially a member of that kind. Presumably, physical objects constitute a natural kind.

And so, given kind-essentialism, physical objects are essentially physical objects. Nothing can start out lacking an *essential* property and then later acquire it. Given all this, kind-essentialism implies that something that starts out as a nonphysical object cannot possibly become a physical object. In this way, kind-essentialism threatens the physicalist's account of the Incarnation.

Presumably, human souls would constitute a natural kind. Recall that dualists must say that God the Son became a human soul, though he did not start out that way. Thus kind-essentialism also undermines dualistic Incarnation. Indeed, kind-essentialism undermines the Incarnation on any view. For surely if there are natural kinds, human beings constitute one. Kind-essentialism therefore says that *being human* is an essential property of all humans. But God the Son became human, though he did not start out that way.

Believers in the Incarnation must reject kind-essentialism. Once kind-essentialism is rejected, it is hard to see why the nonphysical God the Son could not become a physical human organism.<sup>20</sup> Perhaps this is the sort of thing that might not seem possible merely upon reflection, given no relevant revelation. But the same thing goes for God the Son's becoming human. This is the mystery. Once we've accepted that possibility, we should accept whatever else comes along with it, including – if part of being human is having physical properties – the Son's coming to have physical properties, that is, coming to be a physical object.

The dualist might still resist. She might object that – even given the Incarnation – God the Son's becoming a human organism seems impossible while his becoming a human soul does not. Presumably, this objection presupposes that there is a “bigger difference” between the divine and (alleged) physical humans than there is between the divine and (alleged) nonphysical humans. But, in reply, the difference between God the Son and each of us is staggering. The difference between a nonphysical human person and a physical human person is comparatively trivial. If we believe that God the Son became a human being, we have swallowed the camel. To insist that God the Son could not possibly become a *physical* human is to strain out a gnat.

Once we accept the possibility of God the Son's becoming a human being, there remains no good objection to the possibility of his becoming a physical human being. So the Incarnation does not support dualism over physicalism. Quite the contrary. When it comes to the Incarnation, physicalism has two advantages over dualism.

To see the first advantage, recall that dualism – given its account of embodiment – has trouble making sense of God the Son's having exactly the body of Jesus. (It also had trouble affirming that the Father and the Spirit lack bodies.) But physicalism has no such trouble. Physicalism's account of embodiment is that a person has a body if and only if she is identical with that body. Given this account, we can easily state what it is for the Son to have exactly the body of Jesus: God the Son is identical with the body of Jesus and with no other. (And since neither the Father nor the Spirit is identical with a body, the physicalist's account of embodiment tells us that neither has a body.)

To see physicalism's second advantage over dualism, recall that the dualist says that *to be human* is *to be a human soul*. And so dualism makes Christ's becoming human (that is, the Incarnation) one thing, but his becoming embodied something else. But it would be nice to have an account of the Incarnation that required Christ to become incarnate. And we can have such an account, given physicalism. The most straightforward account says that *to be human* just is *to be a human organism*. Christ's becoming human and his coming to have a body – his becoming incarnate – would then be one and the same thing.

There may, however, be a problem with this most straightforward account. To begin to see this potential problem, consider Gregory of Nazianzus's famous anti-Apollinarian remark:

If anyone has put their trust in [Christ] as a human being lacking a human mind, they are themselves mindless and not worthy of salvation. For what has not been assumed has not been healed; it is what is united to his divinity that is saved . . . Let them not grudge us our total salvation, or endue the Saviour only with the bones and nerves and mere appearance of humanity. (Quoted in McGrath 2001, 362)

The chief objection to Apollinarianism is soteriological. To fully redeem humanity, Christ must be fully human. He must not be merely physically human – as the Apollinarians said – but also mentally human. (I think this is the point of the creedal insistence that Christ has a “rational soul”; see Kelly 1978, 296–297.) The moral of all this, for our purposes, is that being a human organism – even a human organism with mental properties – might not be sufficient for being “mentally human,” for having a “human mind.” And if it is not, then the claim that *to be human* just is *to be a human organism* is simply false.

Suppose, for the sake of argument, that being a human organism with mental properties is insufficient for having a human mind. (Presumably, being a human soul with mental properties is also insufficient.) Then the physicalist should say that having a human mind requires thinking in certain ways, having various experiences, and so on (see Swinburne 1994, 208). She should say, that is, that to have a human mind is *to be a human organism with a certain sort of mental life*.

The physicalist might say that to be human is simply to be a human organism (with mental properties). Or, instead, the physicalist might say that to be human is to be a human organism with a certain sort of mental life. Either way, physicalism makes becoming identical with (and so having) a body necessary for becoming human. Either way, according to physicalism, the Incarnation – that is, the Son's becoming human – requires his becoming embodied. I say that the Incarnation should be dependent on God the Son's becoming embodied. So, I say, we have another point in favor of physicalism over dualism, since dualism implies the possibility of the Incarnation without embodiment.

I have been treating physicalism as the claim that each human person is identical with a human organism. But a better (and more inclusive) definition of “physicalism” might be that each human person is a physical object, though not necessarily a human organism. Thus construed, physicalism of course rules out dualism. But it is consistent with a wide range of views. It is consistent with – but does not entail – our being organisms. It is consistent with our being brains. It is consistent with each of us being co-located (but not identical) with a living human body (see Baker 2000). And so on.

Each of these versions of physicalism has its own account of embodiment. According to one, a person has a body if and only if she is a (brain that is) part of a human body. According to another, a person has a body if and only if she is co-located with a body. And so on. These accounts of embodiment have none of dualism's problems with Christ's having exactly one body or with the other divine persons having none. For nothing in the omniscience or omnipotence of God suggests, for example, that the persons of the Trinity are proper parts of, or are co-located with, each and every body. Moreover, defenders of each of these accounts can insist that, to become human, Christ had to become a physical object of some sort and so had to become incarnate.

### 30.3 Conclusion

I assume that we are not events or properties, but rather objects or things or substances. Given that we are objects of some sort, there is no question that we are objects with *mental* properties; obviously we are. The only real question is whether we are objects with *physical* properties. If we are, we are physical objects. If we are not, we are nonphysical objects. Given that we are objects of some sort, the only options are physicalism and dualism.<sup>21</sup>

Our options are physicalism and dualism. Which are we to endorse? The Incarnation points us toward physicalism. For the physicalist, unlike the dualist, can insist that becoming embodied is necessary for becoming human; she can insist that the Incarnation requires the Son to become incarnate. Moreover, and more importantly, the physicalist – but not the dualist – can easily and straightforwardly account for God the Son’s having the body of Jesus and no other.

Of course, physicalism does not solve every puzzle or answer every question regarding the Incarnation. To take just one example, physicalism is silent on how to reconcile Christ’s divinity with his apparently not knowing the hour of his return (Matthew 24:36). So physicalism is not a cure-all with respect to the Incarnation. Nevertheless, it does cure something, doing away with the embodiment ills brought on by dualism. This gives Christians a good reason to be physicalists.<sup>22</sup>

### Acknowledgments

This chapter first appeared in *Persons, Human and Divine*, edited by Peter van Inwagen and Dean Zimmerman (Oxford: Oxford University Press, 2007), pp. 281–300, and is here reprinted by permission of the author and Oxford University Press.

The initial paper of this chapter was presented at the Pew Workshop on the Metaphysics of the Human Person (February 2004) and to the Butler Society at the University of Oxford (March 2004). Thanks to both audiences for helpful comments. Thanks also to Mike Bergmann, Jim Cargile, Joseph Jedwab, Brian Leftow, Mark Murphy, Mike Rea, Richard Swinburne, Patrick Toner, and Thomas Williams.

### Notes

1. A venerable theory of the Incarnation – arguably, the historically dominant theory – seems to reject this assumption. This is the theory that God the Son, in the Incarnation, “took up” an “individual human nature.” This individual human nature is supposed to be intrinsically just like a complete human person. Indeed, it would have been a human person had it not – perhaps *per impossibile* – been taken up by God the Son. (A theory along these lines is associated with, for example, Aquinas, Scotus, and Ockham; for a discussion, see Freddoso 1986.) But I have a hard time seeing how the individual human nature fails to be a human person (as it must, lest this theory be Nestorian). Moreover, it is hard to see how “taking up” an individual human nature makes God the Son human in the same way you and I are human; and if he is not human just as we are, I do not see how he could be fully human.
2. Some dualists deny that a person is identical with a soul, saying instead that each human person is a composite of soul and body. This is a minority view among dualistic philosophers, and for good reason. For, if there are souls, they have mental properties. Persons have mental properties, too. So the dualist who denies that a person is identical with a soul must say that there are two objects with mental properties (a person and her soul) where normally we think there is one. (For more objections to the claim that a person is a composite of soul and body, see Merricks 2001a, 47–53, esp. 48n9, and Olson 2001.)

3. My intending to raise my arm causes events in my brain, starting a chain of causes, which result in my arm's rising. So I may not cause my arm to rise *directly*. But the relevant point is that, according to the dualist, the only physical events that I directly cause are events in my own body, including my own brain. Along similar lines, suppose my hands are tied behind my back – or the nerves in my arms are injured – so that I cannot raise my arm just by intending to do so. This does not render me disembodied, since my soul could still directly cause events in my brain.
4. This account of embodiment seems to rule out one's causing a cup, which is not part of one's body, to levitate simply by intending that it does. And it rules out someone's having knowledge of the physical world via the body of another. Perhaps the dualist might revise this account to permit the possibility of a bit of magic here, a little clairvoyance there. Nevertheless, her account of embodiment should rule out one's consistently having direct control over, and knowledge by way of, a body that is not one's own. If one found oneself regularly enjoying the control of, and knowledge via, a body, the dualist should say that one thereby *has* that body.
5. Hasker (1999) takes a soul's having a body to be that body's generating that soul. This bodes ill for the Incarnation – surely the body of Jesus does not generate God the Son – so I shall set Hasker's account aside.
6. Indeed, everyone – not just the theist – ought to find the standard objection to dualism unconvincing. Compare: If God exists, then something nonphysical (God) causes physical events; nothing nonphysical can cause physical events; therefore, God does not exist. This argument seems neither better nor worse than the standard objection to dualism. But this argument is question-begging, or so close to question-begging as makes no difference. So the moral is that the standard objection to substance dualism likewise begs the question. (That is, the argument that takes “no causal interaction between the nonphysical and physical” as a *premise* seems to beg the question against the substance dualist. Opponents of substance dualism who *argue* for the impossibility of such interaction – such as Kim 2001 and Sosa 1984 – need not beg the question.)
7. I shall move back and forth between claims like “God has control” and “the Father, the Son, and the Spirit have control.” But I think that the sense in which the Trinity has a property like *having control* is not the same as the sense in which each divine person has that property. For more on this, see Merricks (2006, sect. VII).
8. Another reply says that God's control of physical objects is totally unlike our control of our bodies. As a result, God does not have direct control of bodies, at least not in the sense that you and I do. Defending one version of this reply, Brian Leftow (1997, 120) says “In sum, God cannot in fact move matter by basic acts (again, with perhaps the exception of the Incarnation).” Suppose that God's moving matter in the way that we move our bodies is impossible. Then, given the dualist's account of embodiment, God the Son's becoming incarnate is impossible too. So this reply will not help the dualist. Further, this reply denies that God is a nonphysical entity who causes physical events in the sense of “cause” that nonphysical souls cause physical events. And so, given this reply, the dualist loses her theism-based rejoinder to Dennett's “standard objection.” Thus, insofar as the dualist thinks the theism-based rejoinder is a good one, she should reject this reply.
9. Another Incarnation-based objection to this Platonic view of God is that in taking on humanity, God the Son underwent some sort of change. (Senor 1990 defends this objection; Leftow 2002 responds.)
10. Let me address a couple of the most obvious strategies for tweaking Objection 4 in light of my reply. One strategy says that having a body is a matter of the *ability* to have experiences caused by that body. I object that this would give the Father, the Spirit, and the Son each and every body, since each divine person, being omnipotent, is able to have experiences caused by any body he chooses.

A second strategy says that a soul's having a body is a matter of that soul's being such that, were that body in such and such a condition, then that soul would have thus and so experience. (Arguably, even when unconscious, the nearest counterfactual situation in which my body is in sense-experience-causing conditions is also one in which I have the corresponding experiences.) But, I object, whether one is embodied ought to be a matter of how things actually are, not a matter of how they would be, had things gone differently. Moreover, consider a disembodied soul, whose former body has died. If that body were in sense-experience-causing conditions – conditions presumably requiring it to be alive – then I suppose the soul would have the appropriate experiences. After all, the nearest counterfactual situation in which, for example, Lincoln's body is now alive is presumably, given dualism, a situation in which Lincoln's soul is embodied. But this should not imply – as it seems to given this second strategy – that Lincoln's disembodied soul now has a body.

11. The most serious objection to a kenosis says that to be divine just is to have the appropriate array of divine attributes; to shed those attributes is to thereby shed divinity; thus a kenotic Christology denies the Lord's divinity. (One possible reply is that divinity requires only attributes such as being-omniscient-unless-freely and-temporarily choosing-to-be-otherwise; Morris 1986, 99–107.)
12. That is, such a person cannot have a body *if more than one body exists*. If exactly one body exists, then such a person can have a body. Indeed – given the revised account – that body will automatically be the body of every omnipotent and omniscient person. This result is another flaw in the account, since surely a divine person could create a world with one body without thereby becoming embodied.
13. Some dualists believe that an embodied person is a composite of soul and body. Their resources for accounting for when a soul and body are thus united – for when a soul and body are related by *composing a person* – are no different from those of the standard dualist. (Indeed, Swinburne, whose account of the union of soul and body has been our touchstone, holds that a person is composed of soul and body; see Swinburne 1986, 146.) So this sort of dualism provides no way around the arguments above.
14. This is not surprising. Dualistic embodiment is a matter of causal relations that one bears to something contingently, to something not identical with oneself. It would be odd if one's very humanity were a matter of such relations.
15. More carefully, to be a physical object is to have physical properties *and fail to have a nonphysical object (like a soul) as a part*. If we were composed of both a body and a soul, we would have physical properties, but would not be physical objects. But, as already noted, there are good reasons to set aside the view according to which we are composites of soul and body. So I set it aside.
16. I say there is exactly one such object. Below we shall note a version of physicalism according to which more than one such object exists, each wholly co-located with the others.
17. Moreover, the physicalism here is physicalism about *human persons*, saying only that we humans are physical; it does not say that everything is physical.
18. Physicalism suggests that a person does not exist between death and resurrection. But perhaps a physical person – a human organism – could become nonphysical (and presumably nonhuman) at death and continue to exist in such a state until becoming physical again at resurrection. This seems to be Eleonore Stump's (2002) understanding of Aquinas's view. I think Aquinas can also be read as saying that the person does not exist between death and resurrection, but only the person's substantial form (see Objection 5 and his reply in *Summa Theologica* IIaIIae.83.11).
19. It is odd that Leftow thinks that the Incarnation is inconsistent with a normal human person's being identical with a body. For Leftow (2002) endorses a picture of the Incarnation according to which the Son is *not* related to the body of Jesus like each of us is to our own bodies. (So Leftow rejects the assumption that opens this chapter.) Thus – given Leftow's view of the Incarnation – an ordinary human person's being identical with a body would not imply that the Son becomes identical with a body in the Incarnation. Moreover, Leftow (2001) himself seems to identify a person in this life with her body, supplementing this with a Thomistic theory of the nature of living human bodies.
20. C. Stephen Evans rightly insists that we cannot say “the identity of Jesus as the Son of God is grounded in bodily continuity, since the incarnation is a change from a bodiless state to an embodied state” (Evans 2002, 269). Some might worry, however, that if we say that God the Son became identical with a human body, we are somehow committed to “a bodily theory of personal identity” that rules out his having existed without a body. But this worry is misplaced. Just so long as *being a body* is a contingent property of what has it, it is possible that that very thing – that very body – could have existed even though it was not a body. (Compare: I am identical with a professor; but this very thing – this very professor – could have existed (and did exist) without being a professor.)
21. A couple of points will clarify my hasty argument for “only physicalism or dualism.” (1) My argument ignores the option of our being composites of soul and body. But since that option just is a form of dualism, that option is consistent with the argument's conclusion, and with the points to follow in this section of the chapter. (2) Somewhat misleadingly, I would (in the context of this chapter) classify idealism as a form of dualism. For given idealism, we are mental entities that have no physical properties. Given idealism, having a body will presumably be a matter of being associated in the right ways with the relevant ideas. I suspect that, however the “right ways” are cashed out, this will mimic what the standard substance dualist says about embodiment. And so I suspect that idealism is vulnerable to this chapter's objections to dualism.
22. The hope for eternal life, which in scripture is often expressed in terms of hope for the resurrection of the body, gives Christians another good reason to be physicalists (see Merricks 1999; see also Merricks 2001b).



## References

- Baker, Lynne Rudder. 2000. *Persons and Bodies*. Cambridge: Cambridge University Press.
- Cross, Richard. 2003. "Incarnation, Omnipresence, and Action at a Distance." *Neue Zeitschrift für Systematische Theologie und Religionsphilosophie*, 45: 293–312.
- Dennett, Daniel. 1991. *Consciousness Explained*. Boston: Little, Brown.
- Evans, C. Stephen. 2002. "The Self-Emptying of Love: Some Thoughts on Kenotic Christology." In *The Incarnation*, edited by Stephen T. Davis, Daniel Kendall, and Gerald O'Collins, 256–272. Oxford: Oxford University Press.
- Freddoso, Alfred J. 1986. "Human Nature, Potency and the Incarnation." *Faith and Philosophy*, 3: 27–53.
- Hasker, William. 1999. *The Emergent Self*. Ithaca, NY: Cornell University Press.
- Kelly, J. N. D. 1978. *Early Christian Doctrines*, rev. edn New York: HarperCollins.
- Kim, Jaegwon. 2001. "Lonely Souls: Causality and Substance Dualism." In *Soul, Body, and Survival: Essays on the Metaphysics of Human Persons*, edited by Kevin Corcoran, 30–43. Ithaca, NY: Cornell University Press.
- Leftow, Brian. 1997. "Divine Action and Embodiment." *Proceedings of the American Catholic Philosophical Association*, 71: 113–124.
- Leftow, Brian. 2001. "Souls Dipped in Dust." In *Soul, Body, and Survival: Essays on the Metaphysics of Human Persons*, edited by Kevin Corcoran, 120–138. Ithaca, NY: Cornell University Press.
- Leftow, Brian. 2002. "A Timeless God Incarnate." In *The Incarnation*, edited by Stephen T. Davis, Daniel Kendall, and Gerald O'Collins, 273–299. Oxford: Oxford University Press.
- McGrath, Alister E. 2001. *Christian Theology: An Introduction*, 3rd edn. Oxford: Blackwell.
- Merricks, Trenton. 1999. "The Resurrection of the Body and the Life Everlasting." In *Reason for the Hope Within*, edited by Michael Murray, 261–286. Grand Rapids, MI: Eerdmans.
- Merricks, Trenton. 2001a. *Objects and Persons*. Oxford: Clarendon Press.
- Merricks, Trenton. 2001b. "How to Live Forever Without Saving Your Soul: Physicalism and Immortality." In *Soul, Body, and Survival: Essays on the Metaphysics of Human Persons*, edited by Kevin Corcoran, 183–200. Ithaca, NY: Cornell University Press.
- Merricks, Trenton. 2006. "Split Brains and the Godhead." In *Knowledge and Reality: Essays in Honor of Alvin Plantinga*, edited by Thomas Crisp, Matthew Davidson, and David Vander Laan, 299–326. Dordrecht, Netherlands: Springer.
- Morris, Thomas V. 1986. *The Logic of God Incarnate*. Ithaca, NY: Cornell University Press.
- Olson, Eric. 2001. "A Compound of Two Substances." In *Soul, Body, and Survival: Essays on the Metaphysics of Human Persons*, edited by Kevin Corcoran, 73–88. Ithaca, NY: Cornell University Press.
- Plantinga, Alvin. 1999. "On Heresy, Mind, and Truth." *Faith and Philosophy*, 16: 182–193.
- Senor, Thomas. 1990. "Incarnation and Timelessness." *Faith and Philosophy*, 7: 149–164.
- Sosa, Ernest. 1984. "Mind–Body Interaction and Supervenient Causation." *Midwest Studies in Philosophy*, 9: 271–281.
- Stump, Eleonore. 2002. "Aquinas' Metaphysics of the Incarnation." In *The Incarnation*, edited by Stephen T. Davis, Daniel Kendall, and Gerald O'Collins, 197–220. Oxford: Oxford University Press.
- Swinburne, Richard. 1986. *The Evolution of the Soul*. Oxford: Clarendon Press.
- Swinburne, Richard. 1994. *The Christian God*. Oxford: Clarendon Press.
- van Inwagen, Peter. 1998. "Incarnation and Christology." In *Routledge Encyclopedia of Philosophy*, edited by Edward Craig, vol. 4, 725–732. London: Routledge.

# Debating the Resurrection

# Materialism Most Miserable

## *The Prospects for Dualist and Physicalist Accounts of Resurrection*

JONATHAN J. LOOSE

Whether we are to live in a future state, as it is the most important question which can possibly be asked, so it is the most intelligible one which can be expressed in language. Yet strange perplexities have been raised about the meaning of that identity or sameness of person, which is implied in the notion of our living now and hereafter, or in any two successive moments. And the solution of these difficulties hath been stranger than the difficulties themselves.

Joseph Butler (1897 [1736], 317)

### 31.1 The General Resurrection

Stephen Davis's detailed assessment of the doctrine of the general resurrection suggests that it is the claim that those who have died will persist into a subsequent, embodied life by means of a divine miracle (Davis 2010, 108). The assumption on which this chapter is focused is that in resurrection a relation of numerical identity (henceforth, identity) must hold between the person who died and the person who is raised such that they are the very same individual. A concern to establish this has attended discussion of the general resurrection from earliest times. For example, early Christian theologians tended to join the vast majority of people in accepting anthropological dualism, taking the view that a human person is identical to or partly composed of a simple immaterial substance. At death the body decays and disappears while the soul departs and exists for an interim period in a disembodied state.<sup>1</sup> On the Last Day the very same (identical) human person stands again as a result of the miraculous reuniting of the soul with the identical body in resurrection. The relevant particles are reassembled to compose the body, which is suitably transformed to be fit for heavenly existence. This reassembly model is so characteristic of the period that Davis labels it the patristic theory in

contrast to alternatives that do not depend on reassembly, which he describes as modern (Davis 1993, 95).

The patristic theory suffers from problems that partly explain why recent discussion focuses on modern alternatives. For example, if sufficient matter required for my resurrection were annihilated or – more likely – came to be part of other human bodies prior to the Last Day then it (and hence my body) would not be available for resurrection. Also, given the continual replacement of bodily matter, it is possible that at a later stage of earthly life the material simples composing my body would not overlap with those that composed it at an earlier stage. Thus at the Resurrection both earlier and later bodies could be reassembled simultaneously, standing side by side each fully and equally qualified to be my resurrection body. Yet identity is a reflexive, symmetrical and – crucially – a *transitive* relation and so two bodies that are each identical to my body must also be identical to each other, which is absurd. Perhaps one solution is to claim that only the later body is qualified since it composed me most recently. However, this would make the identity of a reassembled body at the Resurrection dependent on the absence of a more recent candidate, which would also be absurd since it is a necessary rather than a contingent fact that a thing is the thing that it is and not something else. It seems safe to say that whether or not a particular body is my body cannot depend on the nonexistence of a better qualified candidate although, as we shall see, not everyone agrees.

Given these problems with the patristic view one might expect modern theorists to relinquish the bodily identity requirement and rely on the continued existence of the soul to preserve personal identity. This would not entail that human persons naturally survive death. Even if not essentially physically dependent, we are clearly contingently dependent on our bodies for our functioning during earthly life and so it remains possible that in the natural course of events immaterial human persons would cease to exist when the body dies. Survival would nevertheless be possible as an achievement of omnipotence, consistent with the Christian understanding of resurrection as divine gift. However, this promising move does not remove the Patristic commitment to dualism and – as van Inwagen (1998) notes – while dualist and materialist alike must face the problems generally associated with their positions, the problems often thought to be associated with dualism continue to make it unpopular. Thus a number of Christian scholars resist dualism energetically. For example, Baker describes souls as *surds* in nature (Baker 2007), and Wright has claimed that, “we do not need what has been called ‘dualism’ to help us over the awkward gap between bodily death and bodily resurrection” (Wright 2011). However, materialist accounts face what van Inwagen has called a “special philosophical problem about personal identity” (van Inwagen 1998). It is no easy task to show that a wholly material person at the Resurrection could be identical to a wholly material person who previously died. If the dualist can explain personal identity across the bridge of death and the materialist cannot then – to say the least – dualism has a significant advantage in demonstrating the possibility of resurrection. Indeed materialism may turn out to be inconsistent with resurrection belief.<sup>2</sup> So we consider the prospects for both dualist and materialist theories, but focus on the latter.

### 31.2 Dualism and Personal Identity

The dualist’s account of resurrection depends on the possibility that the identity of a person over time is preserved by the persistence of a simple immaterial substance with no necessary

connection to a particular physical or psychological career. While there is little to be said about the persistence of simples, one common challenge to the claim that human persons persist as essentially nonphysical beings is that personal identity must consist in the criteria that we normally use when making judgments about it, namely the “continuities of mental or physical properties or of the physical stuff (that is, the bodily matter) of which they are made” (Swinburne 2012, 105). This is the complex view of personal identity over time. In contrast, substance dualism requires the alternative simple view that personal identity consists in a “quite separate further fact” (Parfit 1982) beyond these empirical continuities. Are there reasons to reject the simple view?

It is extremely difficult to hold that the simple view is impossible on the grounds that it is inconceivable that personal identity could consist in a “further fact.” First-person knowledge of what it means to be a subject of experience is central to the way in which we develop the ability to produce and understand sentences involving personal identity and this experience does not depend on particular continuities of mental or physical properties. Furthermore, the belief that human persons are naturally dependent on such empirical continuities is both unpopular and unnatural. Concepts of soul are ubiquitous throughout history and across human societies (including those of the Bible’s authors).<sup>3</sup> Psychological evidence supports the claim that even from infancy the default conception of human persons is that we are less physically constrained than material objects and in early childhood the default belief is that persons are capable of surviving death and disembodiment and doing so with a range of mental states intact.<sup>4</sup> The conceivability of disembodied personal existence has been considered sufficiently robust to constitute a premise in modal arguments for the truth of dualism (e.g., Taliaferro 1994),<sup>5</sup> and disembodied experience is probably conceivable as indicated by speculative accounts such as Price’s (1964 [1953]) famous vision of a world of experienceable and communicable mental images through which disembodied persons might be known to one another.<sup>6</sup>

The ubiquity of default, dualist views about human persons strongly suggests that it is by no means inconceivable that personal identity over time would consist in the continued existence of the same soul rather than the physical and psychological continuities that we make use of when making judgments about it.

However, the seeming conceivability of the claim that a person’s identity consists in a “further fact” does not establish its possibility if further reflection reveals sound arguments demonstrating that the view entails a contradiction. Swinburne presents a typical example of just such an argument, developed from Locke:

[Advocates of the complex view] may claim that “Socrates is the same person as the mayor of Queenborough, but has none of the same brain, memory or character as the mayor,” together with what they may claim to be a necessary truth “no one should be punished for any act which they cannot remember doing” entails “both {the mayor should be punished for any immoral act of Socrates} and not-{the mayor should be punished for any immoral act of Socrates}.” (Swinburne 2012, 113)

As Swinburne points out, however, there are ready objections to this particular argument and the soundness of such arguments is typically difficult to establish to the satisfaction of both defenders and opponents of the simple view. For example, we need only ask why the statement that “no one should be punished for any act which they cannot remember doing” is a necessary truth. Since the statement seems possibly false (as even Leibniz seemed to

think)<sup>7</sup> the inconsistent conclusion is avoided. It is perhaps the difficulty of demonstrating the inconsistency of the simple view that leads some of its opponents to press its implausibility instead (e.g., see Shoemaker 2012). However, the simple view is not implausible per se for the immaterialist who also holds that the “further fact” in which identity consists is the ongoing existence of a particular simple immaterial substance.

Even if the simple view is neither inconceivable nor demonstrably inconsistent, what of the complex theorist’s positive claim that since we obviously depend on physical and psychological criteria when making judgments about personal identity it must be the case that it consists in these criteria? Perhaps, as Flew claimed, “persons are what you meet” (see Price 1964 [1953], 287). The simple theorist holds that Flew’s comment – while true in one sense – is false in the most fundamental sense. To elaborate on an earlier point, our underlying conception of personal identity over time comes not from third-person observation but from direct first-person knowledge of what it is to be a persisting subject of experience and each of us learns to talk about personal identity by associating various expressions with that knowledge. Such knowledge is, of course, psychological; however, this is not a psychological account of identity in the sense that is often discussed. We are not talking about psychological continuities as constitutive of identity, but certain experiences that give direct knowledge of identity. Thus, most fundamentally, *persons are what we know ourselves to be*.

Since judgments about the identities of others cannot involve first-person knowledge, they must be made on the basis of the evidence of those observable physical and psychological continuities that we have learned are normally closely associated with our own direct experiences of personal persistence. This close association would be expected on any version of dualism that takes soul and body to be integrated and most contemporary dualists hold that persons, while ontologically separable in principle, are best understood holistically.

It seems, then, that there is no compelling reason per se to hold that personal identity cannot consist in a “further fact” beyond physical and psychological continuities. Positively, it should also be noted that the simple view exerts a very strong pull because it is consistent with the strongly and widely held conviction that personal identity must be *determinate*. Joseph Butler and Thomas Reid famously distinguished different senses of identity, holding that personal identity is correctly understood in a determinate “strict and philosophical” (or “perfect”) sense as opposed to the “loose and popular” sense that is employed when referring to other things.<sup>8</sup>

In introducing and naming the distinction between simple and complex views Parfit – himself a complex theorist – emphasized the strength of determinacy’s pull, writing that it is the simple view that is adopted by “the great majority of those who have thought about the question” (Parfit 1982, 227). He claimed that “most of us believe, and nearly all of us are inclined to believe, the Simple View” (Parfit 1982, 228) because problem cases reveal a strong intuition that whatever the degree of observable continuity between me and a future person, it can only ever be the case that that person either *is* or *is not* me. Yet, if personal identity consists in something that can be a matter of degree, such as the extent to which physical and psychological characteristics are continuous over time, then scenarios could be conceived in which the relevant continuities are present to a level that leaves the question of personal identity indeterminate even for an observer otherwise in possession of all the facts. In such a scenario it would seem just as likely that personal identity is preserved as that it is not and arguments could be made either way. Parfit claims that “nearly all of us” would

prefer to claim that personal identity consists in a “further fact” rather than accept that it could be like this. Swinburne notes the important role that thought experiments involving such ambiguous scenarios have had in demonstrating the implausibility of complex views of personal identity and persuading philosophers of the truth of the simple view (Swinburne 2012). When attempts are made to reduce personal identity to nonpersonal continuities that admit indeterminate situations, the simple theorist may well claim that the perfect sense of identity is being confused with its loose and popular counterpart.<sup>9</sup> The importance of ambiguous situations will become clear when considering materialist, modern accounts of resurrection, to which we will turn next.

To summarize, substance dualism accommodates personal identity across the bridge of death in accordance with the simple view. The simple view is not inconceivable, underlying as it does the way in which we typically come to understand sentences involving personal identity and being consistent with the ubiquitous default view that persons are not physically constrained and may survive death. Neither does the simple view obviously entail a contradiction that would undermine the inference from conceivability to possibility. Importantly it is also consistent with the widely held intuition that personal identity must be determinate and it accommodates the important evidential role we ascribe to physical and psychological continuities when making judgments about the identity of other persons in normal circumstances.

Given all this, the dependence of the dualist account of resurrection on a simple view of personal identity seems unproblematic. Indeed, given the ubiquity of the simple view of personal identity, we could invoke a principle of credulity and hold that since it seems to almost all people to be correct and given that there is no compelling reason to reject it, then it probably is correct. Dualism then seems to offer an excellent account of the possibility of resurrection.

However, it is important to consider the prospects for a coherent materialist alternative. After all, not only is substance dualism unpopular among philosophers but a small number of people report that from the first their intuitions about human persons have swum against the tide and have been strongly materialistic, or that it is difficult in early childhood to understand how a human person might be absent from his dead body (e.g., Corcoran 2005, 69). Such “antecedent physicalists” do not see that dualism has any advantage based on early intuitions about the nature of persons or the way in which concepts and language about personal identity are acquired. We thus consider some popular materialist understandings of resurrection, considering first the account that is the progenitor of the current debate.

### 31.3 The Simulacrum Model

Peter van Inwagen rejects the patristic theory for reasons already given, but also because of the importance he accords to causal continuity in personal identity. To illustrate this, he considers a manuscript completely burned up some time after the death of its human author. Later on, God subsequently gathers up the dispersed particles and reassembles the manuscript. While materially indistinguishable from the original, the reassembled manuscript has a divine rather than human causal origin and is thus not identical with it. What goes for manuscripts goes for material human persons. If a body on the Last Day is to be *me*, then it must be both materially and causally continuous with me. However, according to

van Inwagen this requirement does not preclude the possibility of resurrection. To see this we must recognize two things. First, there is a short period after death during which a corpse is not so badly damaged that it cannot return to life:

a former corpse in which the processes of life have been “started up again” may well be the very man who was once before alive, provided the processes of dissolution did not progress too far while he was a corpse. (van Inwagen 1978, 119)

Second, God can extend this short postmortem window of opportunity to cover the period from death to the Last Day:

Perhaps at the moment of each man’s death, God removes his corpse and replaces it with a simulacrum which is what is burned or rots. Or perhaps God is not quite so wholesale as this: perhaps He removes for “safekeeping” only the “core person” – the brain and central nervous system – or even some special part of it. These are details. I take it that this story shows that the resurrection is a feat an almighty being *could* accomplish. I think this is the *only* way such a being could accomplish it. Perhaps I’m wrong, but that’s of little importance. What *is* important is that God could accomplish it this way or some other. (van Inwagen 1978, 121)

This, then, is the simulacrum model of the general Resurrection. Van Inwagen regrets that his materialism brings him into conflict with the dualistic anthropology of the Church fathers and subsequent Christian history (and he thus holds that, even if false, dualism cannot be a pernicious belief). However, he takes his view to be consistent with the anthropology of Old and New Testaments<sup>10</sup> and draws a parallel between the body preserved between death and resurrection and Paul’s notion of a “naked kernel” (van Inwagen 1995, 486).

Whatever the merits of the simulacrum account of resurrection it has proved unpersuasive. Thus, in a later postscript, van Inwagen changed his original claim, arguing that this is not the only way in which a divine being could accomplish resurrection, although we lack the conceptual resources to understand the alternatives. The model thus demonstrates the metaphysical possibility of the Resurrection given a materialist account of human persons; it is a “just-so story”:

Although it serves to establish a possibility, it probably isn’t true. (And it is easy to see why someone might think it was preposterous, although it might be questioned whether any of us is in an epistemic position to make a judgment of this sort.) (van Inwagen 2009, 327)

In addition to its seeming preposterousness, the account raises at least two other major problems: First, it entails a problematic conception of God; second it is at best uncertain whether it offers the animalist an account of resurrection after all.

The most oft-cited reason to reject van Inwagen’s account is that it entails divine deception, raising the serious concern (shared with Descartes) that it thereby entails a defective concept of God.<sup>11</sup> If bodies are replaced with simulacra at death, then God systematically deceives the bereaved about what it is that is buried or cremated and about which they grieve; a deceit compared by Zimmerman to the theory that God placed dinosaur bones in the earth simply to deceive us about its age (Zimmerman 1999, 196). Van Inwagen’s brief reply to this objection is that in producing a situation in which humans



form false beliefs under optimal conditions God does not of necessity do something morally objectionable and the purpose of providing simulacra is to give a valuable counterfactual demonstration of what death would mean if we were “left to the situation we had earned for ourselves” without Christ (van Inwagen n.d.).

It is reasonable to claim that an act that causes a false belief may not be an act of deception if it is not part of the actor’s purpose to bring it about. However, simulacra are provided solely for the purpose of affecting the experience and beliefs of humans and in a way that seems to require the false belief. The most potent symbol of sin’s limitation of life is found in the interment of the very matter that composed the deceased at the moment of death, because in this event the bereaved are faced with the irretrievable terminus of that life and hence the full horror and finality of death without Christ and (counter-factually) the miraculous nature of Christ’s achievement. The term “simulacrum” can mean “an unsatisfactory imitation” and the knowing burial of a simulacrum would indeed be unsatisfactory by comparison. However, even if some case can be made that simulacra could serve this purpose without divine intent to create a false belief, it seems highly likely that God has another purpose that requires it. The disappearance of bodies as they are snatched at death would seem to provide such obvious evidence of divine activity that it may remove the cognitive freedom for people to choose whether or not to believe. Hence, simulacra cover God’s tracks and preserve this freedom. Divine deception remains a problem.

Deception is an important issue, but does the simulacrum model even amount to an account of resurrection? This animalist model requires that personal identity over time depend on the continuation of a particular life, while resurrection seems to require a gap in that life between death and the Last Day. Careful consideration of the way in which a life might cross the gap shows that the model fails to offer the Christian animalist a sure and certain metaphysical possibility of resurrection; perhaps even pushing her toward a simple view of personal identity over time and most probably toward dualism. To see this requires looking at the simulacrum model in some detail.

Van Inwagen’s moderate account of composition<sup>12</sup> holds that if a group of material simples is to possess the unity required to compose an object then they must be caught up together in a *life* and persistence at different times requires simples that are caught up in that particular life at those times. Hence, if I am to appear at the resurrection, the simples that will compose me then must somehow be caught up in the *very same* life as the simples that compose me now and a model of the possibility of resurrection must show that this can occur.

The question of how lives persist is difficult to answer because the nature of life remains controversial and van Inwagen notes that the task of defining it should be left in the hands of biologists.<sup>13</sup> However, he nevertheless emphasizes the standard view that lives are self-maintaining, homeodynamic events.<sup>14</sup> Therefore, at any moment the simples caught up in a life must possess the individual properties and relations necessary for that life to continue into the future. If these simples come to be dissociated and are thus incapable of causing the life to persist, then that life has ceased in the sense of *disruption*:

We may be confident that the life of an organism which has been blown to bits by a bomb or which has died naturally and has been subject to the normal, “room-temperature” processes of biological decay for, say, fifteen minutes has been disrupted. (van Inwagen 1990, 147)

A disrupted life has ceased irretrievably and so cannot be followed by resurrection. The seeming advantage of the simulacrum model, then, is that the preservation of the corpse

ensures that disruption never occurs. The corpse does not decay and dissolve, but is removed and preserved until the Last Day.

One way to make sense of this idea of preservation is to notice the obvious similarity to situations of cryogenic freezing (e.g., see Hasker 1999, 223). Van Inwagen elsewhere discusses a hypothetical successful case of the cryogenic freezing and thawing of a cat. While the cat is frozen there are no chemical or biological processes ongoing within it and when thawed it appears unharmed with all its vital signs intact. The question is whether the life of the cat that was frozen is identical to the life of the cat that was subsequently thawed. In discussing this question, van Inwagen makes a very important comment: "It is not altogether clear that the life of the cat ceases when it is frozen" (van Inwagen 1990, 146). The absence of biological and chemical processes might suggest the cessation of life, but it is attractive to think that it remains present since it might have been:

*squeezed* into the small-scale physical processes (the orbiting of electrons and the exchange of photons by charged particles). Its life became the sum of those sub-chemical changes that underlie and constitute chemical and large-scale physical unchange. . . . I . . . would describe the frozen cat as a living corpse. (van Inwagen 1990, 147; emphasis added)

If the frozen matter was indeed a living corpse then the cat clearly continued to exist through this period. Cryogenic freezing is a case of the removal from the organism's environment of the resources that it needs to manifest the chemical and biological processes associated with its life, but without removing the capacity for it to do so again when those resources return. It is thus attractive to claim that the organism's life remains pent up within it in the way that van Inwagen suggests. Importantly, in actual cases of cryogenic freezing it is usual to consider the organism to be alive despite the absence of the normal chemical and biological processes. For example, cryogenically frozen embryos or dehydrated tardigrades<sup>15</sup> are considered alive because they are viable, possessing the capacity for vitality (Luper-Foy 2016).

So identity may be preserved in virtue of the persistence of a life event, albeit in an unusual manner. This seems the most likely explanation of the simulacrum scenario. It may be that the short period of time during which death seems to be reversible and during which preservation can take place is a period in which the life is actually *there*.<sup>16</sup> The problem here is that resurrection first requires death and if life is not lost then death has not occurred:

It is part of the Christian faith that all men who share in the sin of Adam must die. What does it mean to say that I must die? Just this: that one day I shall be composed entirely of non-living matter; that is, I shall be a corpse. (van Inwagen 1978, 120)

If lives are "squeezed" during preservation then there is no time at which I am composed of nonliving matter and the model amounts to an account of mere resuscitation. An account is required on which preservation involves the loss of life.

It goes without saying that the clearest account of death – namely irreversible disruption – is of no help to the simulacrum model. This is a serious problem since, despite ongoing debate about the notion of life, the claim that disruption is likely to be essential to death is common. As De Grazia observes:

The qualifier "irreversible" is important . . . If the body of an organism stops functioning, even for a long time, but the condition is later reversed so that function resumes, it is presumably incorrect to say that the organism died before returning to life. (De Grazia 2014, 83)

So it is highly likely that the simulacrum model fails because loss of vitality without irreversible disruption is not death and without death a return to vitality is not resurrection.

Nevertheless, we should not reject the model too quickly. If lives squeezed into living corpses cannot offer an account of resurrection, then perhaps there is a way that a life may cease without irretrievable disruption. This is "suspension":

a life has been suspended if it has ceased and the simples that were caught up in it at the moment it ceased retain, owing to the mere absence of disruptive forces, their individual properties and their relations to one another. (van Inwagen 1990, 147)

Could the life of an organism be preserved at death this way? Unlike squeezing, suspension envisions the *loss* of life and unlike disruption this loss may be followed by a subsequent return. However, it remains extremely unlikely that suspension offers the right account of preservation. As explained above, when an organism is preserved in a viable state in the absence of disruptive forces then the most likely and frequent explanation is that the life remains present.

In the unlikely event that lives are suspended rather than squeezed at death, some development of the principles governing composition and persistence is required to show that the organism continues to exist in the absence of its life and that the life of the revived organism is the same as the life of the organism that died despite a temporal gap. In the case of persistence, van Inwagen's developed principle holds that "if a life is going on at  $t_1$  and  $t_3$ , then for any time  $t_2$  between  $t_1$  and  $t_3$  there must be objects whose activity at  $t_2$  constitutes *or results from* that life" (van Inwagen 1990, 149; emphasis added). By allowing that a later life that merely *results from* an earlier one may be identical to it across a temporal gap, the principle accommodates the strong intuition that the revived organism would be identical to the organism that died. However, if our understanding of the persistence of lives is sharpened in this way, it risks undermining strong intuitions in other difficult cases.

Consider metamorphosis. The physical facts seem to suggest strongly that metamorphosis is a case in which the life of one organism (a caterpillar) comes to an end and the life of a distinct organism (a butterfly) begins. If temporal gaps in lives are impossible then this view can be defended on the grounds that there is an interim period during the pupal stage when the life of the caterpillar has ceased and the life of the butterfly has not yet started. However, it seems more difficult to defend the claim that the processes within the chrysalis allow us to say that the life of the butterfly does not even *result* from the life of the caterpillar. This example demonstrates the difficulty of offering an account of the persistence of life that avoids creating conflict between strong intuitions about distinct, difficult cases. These difficulties arise only in the case of suspension and thus reinforce the view that it is better to think of preserved lives as squeezed into the small-scale physical processes underlying temporary chemical and biological unchange.

Despite what has been said in favor of squeezing over suspension, there is a larger issue. It is important to be mindful of van Inwagen's indicative remark that it is *not altogether clear* whether or not a life ceases in such circumstances. There is ambiguity surrounding the persistence of a life in the simulacrum situation and it is not obvious that this ambiguity would be removed simply by a more detailed knowledge of the physical facts. If not, then this indeterminacy could be a linguistic matter; a consequence of our semantic indecision given the lack of a perfectly precise understanding of life. If such knowledge is not beyond our epistemic limits to attain then perhaps we will attain it and thereby know whether the

simulacrum model entails a gap in the life of an organism, whether identity would be preserved across a gap, and how this would be consistent with solutions to other difficult problems about the persistence of lives. Without this knowledge it remains very difficult for the animalist to be certain about the possibility of resurrection in light of the simulacrum idea.

In fact, we can go further and suggest – given the previous discussion – that the simulacrum model might offer the kind of ambiguous state of affairs that Parfit recognized drives many to the simple view and which Swinburne urges should do so given the strong intuition that personal identity cannot be indeterminate. Van Inwagen concludes that lives are metaphysically vague entities admitting genuine borderline cases.<sup>17</sup> If so then ambiguity around the presence of a life through a period of preservation could reflect a relation of “indefinite identity.” It would not be true to say either that the person at the resurrection is or is not identical to the person who died. This emphasizes not only the failure of the simulacrum model but also the inability of animalism to offer an account of personal identity accommodating strong intuitions about its determinate nature.

I have argued that the seemingly preposterous simulacrum model fails to offer animalists such as van Inwagen an account of the metaphysical possibility of resurrection. It is most likely an account of resuscitation, but the situation is ambiguous and highlights the inability of animalism to provide a determinate account of personal identity, putting it at a disadvantage in comparison with the dualist’s simple view. However, this model has regularly been cited by Christian physicalists as a conceptually coherent account of the metaphysical possibility of resurrection with the problem of deception being the remaining sticking point. It is for this reason that Zimmerman offered an alternative account aiming to remove the need to involve God in the systematic deception implied by last-minute body-switching but otherwise utilizing the same metaphysical assumptions (Zimmerman 1999, 2010).

### 31.4 The Falling Elevator Model

Zimmerman suggests that instead of snatching away the body at the moment of death, God could give the simples that make up the body of a person, [A], the power to bud (or, in the original version, to fission) such that there come to be two identically structured sets of simples; one in this world [C], and one in the next [B]. Each of these products inherits the life-preserving causal relation from [A]. Thus, the self-sustaining causal process that had been passed down a single path during [A]’s earthly life would now continue down two separate and unrelated paths in two different worlds. However, crucially, we might say that the budding is only *singly successful*, since [C] in this world immediately goes on to constitute a nonliving corpse while [B] in the next world, suitably healed, functions as the sole and therefore successful candidate for the continuation of the pre-fission life (i.e.,  $A = B$ ). Zimmerman named this idea the falling elevator model because it describes a last-minute escape from annihilation, just as a cartoon character might escape death from a falling elevator by having it stop an inch from the ground so that he can step out of it! The difficulties with this “budding” account of resurrection are first that it is inconsistent with the nature of identity and second that it does not remove the deception problem after all.

The first problem is that the identity of [B] in the next world depends on the fate of [C] in this world. Imagine by contrast a *doubly successful* budding process in which both [B] and [C] are living bodies. In that case we would reasonably conclude that [A] continues life on

earth as [C] (i.e.,  $A = C$ ) and the product of budding in the next world [B] is a different person. Thus, whether or not [A] continues into the next world as [B] *depends* on the fate of [C]. As noted at the outset, it seems absurd that this should be the case, and identity cannot function this way. Noonan emphasizes this with his “only X and Y” principle (“OXY”) (Noonan 2003; see also Williams 1956) which may be stated in terms of [A] and [B] as follows:

Whether a later individual [B] is identical with an earlier individual [A] can depend only on facts about [A] and [B] and the relationships between them: it cannot depend upon facts about any individuals other than [A] or [B] [such as C]. (Noonan 2003, 129)

To deny OXY is to deny the necessity of identity. If two things are identical, then they are identical in every possible world. Yet, we have already seen that given the denial of OXY, the question of whether [A] and [B] are identical is dependent on the fate of [C]. So identity is being treated as a contingent rather than a necessary relation.

Another consequence of the nature of identity is that the previous suggestion that a doubly successful budding might lead to the continuation of [A]’s earthly life is mistaken. In this scenario it seems that both budding products have equal claim to identity with [A] and so either both continue [A]’s life or neither does. Given the transitivity of identity and that necessarily one thing cannot be two things, the answer cannot be “both.” Hence, bizarrely, [A] has budded out of existence. The falling elevator’s clash with the necessity and transitivity of identity, if real, is fatal for it.

In order to salvage the budding account, Zimmerman argues that a closest-continuer account of identity can be adopted. Broadly speaking, a closest-continuer account of identity denies OXY and thus allows that there might be multiple competitors for identity with a prior entity. The winner is the competitor more strongly continuous with the prior entity according to the metric specified by the theory in question. Closest-continuer theories are clearly articulated and defended by Nozick (1981).

Zimmerman argues that all materialist accounts must make use of closest-continuer theories and that the consequences of doing so are merely odd rather than absurd. However, Hasker disagrees with the former point, arguing that Zimmerman introduces a weakness into his model by making this move (Hasker 2011). It also seems too quick to claim that the consequences of the closest-continuer theory can be dismissed as acceptable oddities; the view has consequences that count significantly against its acceptability.

Consider again the singly and doubly successful budding situations described above. In the second (doubly successful) situation, the budding event, along with all of the prior events that make up the previous life of [A] are the progenitors of a new person in the next world. In the first (singly successful) situation, those same events fail to be the progenitors of a new person since in that case the next-worldly product of budding is the continuation of a person who already existed. Thus intrinsically identical sets of events produce different things. This seems to be an obviously absurd claim and thus its consistency with the closest-continuer theory serves to emphasize strongly the unacceptability of that theory (see Noonan 2003, 134).

We might also ask whether on the closest-continuer account situations such as those just compared can really be described as different. Of course, it would be absurd to claim that a situation in which a next-worldly person is identical to [A] is no different to the situation in which the next-worldly person is not identical to [A], but it turns out that this is what is

being claimed. Consider Geach's distinction between "mere Cambridge changes" understood as the gaining of a property without undergoing any real change. For example, if my son grows taller than me then I gain the property of becoming shorter than my son, not because I have changed but because he has. While the change in him is real, mine is a "mere Cambridge change." Given this distinction, consider again the situation of [B] in the next world. Whether or not [B] continues the life of [A] depends on the fate of the causally unrelated, and spatiotemporally distinct budding product [C]. Thus, from [B]'s point of view, the differences in the situation dependent on the fate of [C] can only be mere Cambridge differences. There are no real differences between the situations analogous to my son growing taller than me, but only Cambridge differences analogous to my becoming shorter than my son. However, normally two situations are considered identical even if there are Cambridge differences between them. In other words, Cambridge changes are not normally regarded as events (Noonan 137). We are unable to claim the situations are different even though they involve different persons. This obviously absurd claim again emphasizes the unacceptability of the closest-continuer account that is consistent with it.

If all materialist accounts of resurrection must involve a closest-continuer theory of identity then absurdity abounds. We might well agree with Butler that we are approaching the strange problem of personal identity across the bridge of death by proposing yet stranger conceptions of identity. However, Zimmerman's reason for putting forward the budding account is to provide van Inwagen an account of resurrection that avoids the systematic deception of the bereaved. More worrying, then, is the fact that the account fails to do even this.

When first introduced, the account seems to show that at the moment of budding the identical causal connection from [A] to [B] and [A] to [C] establishes that [A]'s corpse is left on earth (= [C]) while [A] enters the next world (= [B]). Hence there is no deception. However, even a closest-continuer theory does not allow that one thing can be two things and so if the budding process takes [A] to the next world then [C] in this world is not [A]'s corpse. Hasker makes this clear through a discussion of what happens to [A]'s proper parts at the moment of budding. It is deeply implausible to think that [A] could be present in the next world while being composed of wholly new parts, as is made clear by the following dialogue:

"That's a fine new axe you have there!"

"Oh, no – it's the same old axe I've been using for many years. But it just came back from the repair shop, where they fitted it with a new handle and a new axehead." (Hasker 2011, 90)

However, if [A]'s proper parts are transferred with [A] to the next world, then [C] is neither [A]'s corpse nor an object composed of the organs and other parts that previously composed [A], since these are all in the next world. The only advantage that [C] has over the simulacrum is that [C] is composed of material simples that were previously caught up in the life of [A]. However, contrary to appearance, [C] fails to be composed of parts that once belonged to [A], for these have been taken to the next world and are still composing [A] (= [C]). It seems, then, that in both van Inwagen's and Zimmerman's models a simulacrum is left on earth while the deceased is transported to the next world. What distinguishes them is the way in which this is achieved. In van Inwagen's account the simulacrum is a divine construction, while on Zimmerman's it is the product of the divine provision of a causal power for the body to jump to the next world. In neither case do the bereaved bury or

cremate the body of the individual who died. The common objection to van Inwagen's account is not dealt with by the budding account.

The prospects for animalism and resurrection seem miserable. Van Inwagen's account makes God a deceiver, is probably an account of mere resuscitation and presents a seemingly ambiguous scenario that emphasizes the relative inadequacy of complex theories of personal identity. Zimmerman's alternative requires a counter-intuitive closest-continuer theory of identity and seems ultimately to cover up rather than remove van Inwagen's deception problem. Even if either of these accounts were to succeed, they are intended not as acceptable proposals but only as demonstrations of metaphysical possibility, offering little hope that resurrection is in fact something that in fact occurs given the nature of the actual world.

Surely the animalist believer in resurrection must seek an alternative materialist metaphysic. Jacobs and O'Connor (2010) seek to apply the budding scenario while avoiding a closest-continuer account of identity. Their materialist metaphysic holds that a complex object possessing features not exhaustively constituted by the features of its parts (e.g., a human being) will additionally possess an *emergent particularity* as a nonmereological constituent. Identity over time will thus depend not only on immanent causal connections but also on the persistence of this emergent particularity. At budding, the individual goes wherever the particularity goes and so OXY is not denied. Given symmetrical fissioning the destination of the particularity at budding is indiscernible but, they claim, there is a fact of the matter and there could plausibly be a built in bias toward the particularity making it to the next world. However, the resurrection case is not symmetrical and if the differences between the this-worldly and next-worldly budding products are taken to be identity-relevant, then the particularity is overwhelmingly likely to fail to make the jump to the next world (see Hasker 2011). Even if it does, the deception problem remains.

### 31.5 Constitutionalism

One further prominent materialist alternative to animalism is based on Wiggins's view that two things of different kinds can be at the same place at the same time (Wiggins 1968). Consider first a statue (David) and the lump of marble of which it is made (Piece). Statue and Piece exist as distinct objects because they have distinct modal properties (e.g., if the matter were re-shaped into a perfect sphere then only Piece would survive; David would not). In that case the relation between David and Piece is one of neither identity nor mere coincidence but of *constitution*, where constitution is a philosophical term of art describing a *sui generis* relation beyond coincidence that – unlike identity – is asymmetric and irreflexive.

In her influential account of constitution, Baker explicates the notion in terms of circumstances and primary kinds (Baker 2000). An object in the Accademia Gallery might possess an extrinsic property that is the type of relation to an artworld that statues must possess (circumstances) and in virtue of this it is a statue (primary kind). Every individual is a member of just one primary kind. Thus, when a thing (*x*) of one primary kind comes to be in circumstances favorable to another primary kind a new thing (*y*) of that latter kind comes into existence. In this way *y* is constituted by (and not identical to) *x*.<sup>18</sup>

On Baker's view, human persons are constituted objects. When a thing of the primary kind "human organism" comes to possess a certain property it is in circumstances favorable

to the primary kind “human person” and it thereby comes to constitute a new thing of that kind. The property in question is a *first-person perspective*, which Baker describes as the property of being able to think of oneself as oneself; to think of oneself without making use of any name, description or demonstrative. This self-conscious reflexive understanding is the defining characteristic of persons for Baker and is not possessed by other conscious animals.<sup>19</sup> The advantage of such a theory when it comes to resurrection is that human persons, while necessarily embodied, do not necessarily have the bodies they in fact have. The career of an individual person may involve periods of constitution by a number of different objects (Baker 2007, 338). Furthermore, since a first-person perspective cannot be fissioned or duplicated there cannot be multiple claimants for identity with a particular individual at the Resurrection and so a closest-continuer theory is avoided.

However, this intriguing theory fails to provide an informative account of personal identity over time. It may be seen as a virtue shared with dualistic accounts that personal identity is not reducible to the nonpersonal continuities relied upon by the complex theorist. However, while dualist accounts ground personal identity in the persistence of a simple immaterial substance (a soul), the uninformative nature of the constitution view has a different and problematic source.

Baker holds that “a person exists when and only when her first-person perspective is exemplified” (Baker 2012, 182). So in order to talk about personal identity over time we must talk about the exemplification of that property over time. However, a first-person perspective is an unusual property since each exemplification is unique to the thing that exemplifies it. To grasp this, consider two properties of the Taj Mahal: it is white and it is identical to itself. While whiteness is independent of the existence of the Taj Mahal, the property of “being identical to the Taj Mahal” is not; it presupposes the Taj Mahal’s existence. So, on pain of circularity, *impure* properties such as “being identical to *x*” cannot be numbered among the constituents of *x*. First-person perspectives are also impure properties. The property of “thinking of myself as myself” presupposes my existence and so cannot be a constituent of me.

The impurity of this property explains why the constitution view is unable to offer an account of what it is that I am<sup>20</sup> and indicates that the features that characterize a first-person perspective should be accounted for in another way. Moreland has observed that the having of a first-person perspective involves being a point from which the world is viewed. It could thus be reduced to a sentient (viewing) kind of substance (point) that possesses the properties characteristic of persons. In other words, we can recognize that a first-person perspective is the thing that a person is rather than a property that a person has. However, if this is the case then the property of being a first-person perspective becomes redundant, since “first-person perspective” just describes the situation in which an ordinary mental property (e.g., being painful) is exemplified by a substantial personal ego. No further constituent is required or needed. As Moreland writes:

The first-person perspective is just a way of describing/referring to an ontologically prior substantial, sentient person with ordinary mental properties to which that perspective can be reduced. (Moreland 2009, 133)<sup>21</sup>

While the seeming benefits of Baker’s account for an account of resurrection are very similar to those of dualism, the view lacks an informative account of personal identity for the pernicious reason that it is built on a property of human persons that cannot bear their



weight: the first-person perspective has no place as a constituent of a human person. Thus the constitution view of persons is unsuccessful in offering the materialist an account of how an individual might die and yet appear again at the Resurrection.

While dualism offers a coherent account of personal identity consistent with the possibility of persistence across the bridge of death, the claim that materialism can do likewise seems flimsy indeed. Animalism aims to offer just-so stories adequate to establish the metaphysical possibility of resurrection by unknown means (which is not much to hope for) but fails to do even this. I therefore conclude that dualism offers a better account of the resurrection than materialism.

To conclude, it seems the materialist ontological commitments of some Christian philosophers should lead them to be of all people “most miserable” (1 Cor. 15:19b). However, the present reassessment in the philosophy of mind of the difficulties commonly associated with various forms of both materialism and dualism has involved an increasing turn away from materialism toward moderate forms of dualism, a renewed interest in the nature of the human subject and a reinvigorated intra-mural discussion about the form dualist theories ought to take (see the rest of this volume). It seems, then, that these Christian philosophers may feel increasingly able to reject materialism, and that this should give them reason for cheer.<sup>22</sup>

## Notes

1. The idea that there is an interim period between death and the Resurrection during which persons exist in an intermediate state (whether or not purgatorial) without either earthly or resurrection bodies is widely assumed by both Protestants and Catholics and could be considered a core assumption of resurrection belief along with those cited by Davis. Existence during the interim period would be impoverished compared with embodied life. During the Patristic period the question of whether or not this impoverishment amounted to the temporary nonexistence of the person was a matter of debate.
2. Since materialist anthropology is certainly inconsistent with the possibility of a conscious, disembodied intermediate state, Christian materialists are typically under pressure to present a theological picture that excludes this (see Chapter 27 and Chapter 28, this volume; see also Wright 2003).
3. A thread of clear affirmation of the cultural universality of “soul concepts” runs through nineteenth- and twentieth-century scientific anthropology. For example, one of the founders of scientific anthropology in the nineteenth century, Edward Burnett Tylor, considered the “doctrine of souls” to be among the basic beliefs on which the social and religious practices of all “primitive” societies are built, while in the twentieth George Peter Murdock listed “soul concepts” as one of those items that “occur, so far as the author’s knowledge goes, in every culture known to history of ethnography” (Burnett Tylor 1958; Murdock 1945).
4. The ubiquity and naturalness of soul belief is demonstrated very clearly through recent work in developmental psychology exploring the conceptions that young children and even infants have of the nature and possibilities of living and dead persons (Bering and Bjorklund 2004; Kuhlmeier, Bloom, and Wynn 2004). These studies indicate an early bias toward viewing human persons as agents as opposed to material objects, as less physically constrained than material objects, and as capable of surviving death with some mental states intact, including desire, emotional, and epistemic states. Kuhlmeier remarks that, “appreciation that people are just objects may be a developmental accomplishment” (Kuhlmeier, Bloom, and Wynn 2004) while Bloom believes that “we implicitly endorse a strong substance dualism of the sort defended by philosophers like Plato and Descartes” (Bloom 2009, 149).
5. It is not surprising that the widespread human tendency to dualistic anthropology is reflected in the dominant anthropology of Christian history. The Christian theist has additional reasons to affirm dualism, most importantly belief in an intermediate state.
6. Price’s account is indicative rather than adequate. While it allows that the communication of telepathic apparitions between persons would allow individuals to be identified, it does not show how one individual would be aware of another in order to initiate the required telepathic apparition. What is required is some frame of reference comparable to a spatial one. This problem disappears on theism since an independent,

inter-subjectively accessible, stable image world can be sustained by God such that images of particular individuals appear to others in appropriate ways. As Price recognizes, this would be something like Berkeley's idealistic vision and Price asks, "Could it be that these idealist metaphysicians have given us a substantially correct picture of the next world, though a mistaken picture of this one?" Whether or not this picture is the right one, it seems to be conceivable.

7. Consider Leibniz' response to Locke in which he says, "I would not wish to deny . . . that I am that *I* who was in the cradle, merely on the grounds that I can no longer remember anything that I did at that time . . . Thus, if an illness has interrupted . . . my . . . consciousness, the testimony of others could fill the gap in my recollection. I could even be punished on this testimony if I had done some deliberate wrong . . . Which this illness had made me forget" (Leibniz 1953 [1686], 237).
8. "[Identity] has no fixed nature when applied to bodies, and very often questions about it are questions about words. But identity when applied to persons has no ambiguity and admits not of degrees or of more or less. It is the foundation of all rights and obligations and of all accountableness, and the notion of it is fixed and precise" (Reid 1941 [1785]).
9. For example, Chisholm took talk involving the identity of things in a loose and popular sense to be a stretching of language motivated by the practical need for convenient ways to talk about entities that are, in a strict sense, merely fictional (Chisholm 1976, ch. 3).
10. This is controversial at best given that this view cannot accommodate the existence of a conscious intermediate state (see Cooper 2000).
11. "By 'God' I mean the very being the idea of whom is within me, that is, the possessor of all the perfections which I cannot grasp, but can somehow reach in my thought, who is subject to no defects whatsoever. It is clear enough from this that he cannot be a deceiver, since it is manifest by the natural light that all fraud and deception depend on some defect" (Descartes 1996 [ca. 1640], 35).
12. When it comes to the existence of composite material objects, van Inwagen's view is moderate in the sense that he rejects both nihilism (the view that there are no composite material objects, but only groups of material simples variously configured) and universalism (the idea that any group of material simples whatever jointly compose a material object).
13. Conceptions of life typically focus on one of three key properties: the ability to reproduce, the ability to undergo Darwinian evolution, or metabolism. Closely related to the idea of life as metabolism is the view that life is the property of being a self-maintaining and self-sustaining chemical system. This is the view put forward by Schrödinger in his influential classic, *What is Life?* (1969) (Bedau 2014).
14. Lives are thus considered to be like storms. Van Inwagen writes of the Great Spot on Jupiter: "The storm moves across the surface of the world, drawing swirls and clots of atoms into it and expelling others, always maintaining its overall structure. One might call it a homeodynamic event" (van Inwagen 1990, 86–87).
15. The *tardigrade* or water bear is a tiny creature that will typically exist undamaged in a completely dehydrated state with no vital signs for many years.
16. Van Inwagen provides another example that he takes to indicate the presence of life when certain seemingly essential parts of the process are missing. He considers a man whose heart has stopped beating noting that: "when the heart stops beating, the human organism will sometimes cause its arterial walls to contract, in a valiant and pathetic attempt to cause the blood to circulate; this indicates that the cells that compose the stricken man are still caught up in a continuing homeodynamic event" (van Inwagen 1990, 146).
17. In accepting this, van Inwagen resists powerful arguments against the possibility of vagueness. This is due to the necessary and reflexive nature of the identity relation, as famously argued by Evans (1978) and Salmon: "Suppose there is a pair of entities  $x$  and  $y$  . . . such that it is vague . . . whether they are one and the very same thing. Then the pair  $\langle x, y \rangle$  is quite definitely not the same pair as  $\langle x, x \rangle$ , since it is determinately true that  $x$  is one and the very same thing as itself. It follows that  $x$  and  $y$  must be distinct. But then it is not vague whether they are identical or distinct" (Salmon 2005, 244). The situation may be even worse for identity over time than it is for identity at a time. See Noonan's discussion comparing Shoemaker's "A Hall and B Hall" with Parfit's "Club" (Noonan 2003, 110–112).
18. The constitution relation is very intimate since *all* properties may be possessed by *both* objects, but in different ways since the properties of one may be *derived* from the properties of the other (Baker 2000, 46–58). Thus, if David has the property of being admired for artistic perfection *essentially*, then Piece has the property of being so-admired only *derivatively* because Piece constitutes David.
19. For example, consider the second occurrence of "I" in the sentence, "I wonder whether I will be tired at the end of the day?" This sentence indicates not only the subject's ability to have thoughts, but also the ability to realize

that one's thoughts are one's own. In contrast, an animal may be able to have (in some way) the thought "I am tired," but this does not indicate a first-person perspective, since the expression does not require thinking of oneself abstractly as an entity that may play a role in a range of situations.

20. The inherent circularity of impure properties in accounting for the constituents of a particular has often been used when objecting to bundle theories of the self, but the point challenges constitution theory. Since constitution theory holds that I exist in virtue of the persistence of a particular attribute, it may be thought of as a kind of single-attribute bundle theory.
21. Baker would presumably not object to this on metaphysical grounds, given that her principal objection to dualism is not that it is incoherent but that souls do not fit in with the natural world as we understand it.
22. I am grateful to Thomas Atkinson, Michael Lacewing and Angus Menuge for feedback on a previous draft of this chapter.

## References

- Baker, Lynne Rudder. 2000. *Persons and Bodies: A Constitution View*. Cambridge: Cambridge University Press.
- Baker, Lynne Rudder. 2007. "Persons and the Metaphysics of Resurrection." *Religious Studies*, 43(3): 333–348.
- Baker, Lynne Rudder. 2012. "Personal Identity: A Not-So-Simple Simple View." In *Personal Identity: Complex or Simple?* edited by Georg Gasser and Matthias Stefan, 179–191. Cambridge: Cambridge University Press.
- Bedau, Mark A. 2014. "The Nature of Life." In *The Cambridge Companion to Life and Death*, edited by Steven Luper-Foy, 13–29. Cambridge: Cambridge University Press.
- Bering, Jesse M., and David F. Bjorklund. 2004. "The Natural Emergence of Reasoning about the Afterlife as a Developmental Regularity." *Developmental Psychology*, 40(2): 217–233.
- Bloom, Paul. 2009. "Religious Belief as an Evolutionary Accident." In *The Believing Primate: Scientific, Philosophical, and Theological Reflections on the Origin of Religion*, edited by Jeffrey P. Schloss and Michael J. Murray. Oxford: Oxford University Press.
- Burnett Tylor, E. 1958. *Religion in Primitive Culture*. New York: Harper.
- Butler, Joseph. 1897 [1736]. "Of Personal Identity." In *The Works of Joseph Butler*, edited by W. E. Gladstone, vol. 1, 317–325. Oxford: Clarendon Press.
- Cooper, John W. 2000. *Body, Soul and the Life Everlasting: Biblical Anthropology and the Monism-Dualism Debate*. Grand Rapids, MI: Eerdmans.
- Corcoran, Kevin J. 2005. "A Constitutional Response." In *In Search of the Soul: Four Views of the Mind-Body Problem*, edited by Joel B. Green and Stuart L. Palmer, 153–176. Downers Grove, IL: IVP Academic.
- Chisholm, Roderick M. 1976. *Person and Object: A Metaphysical Study*. London: Allen & Unwin.
- Davis, Stephen T. 1993. *Risen Indeed*. London: SPCK.
- Davis, Stephen T. 2010. "Resurrection." In *The Cambridge Companion to Christian Philosophical Theology*, edited by Charles Taliaferro and Chad Meister, 108–123. Cambridge: Cambridge University Press.
- De Grazia, David. 2014. "The Nature of Human Death." In *The Cambridge Companion to Life and Death*, edited by Steven Luper-Foy, 80–98. Cambridge: Cambridge University Press.
- Descartes, René. 1996 [ca. 1640]. *Meditations on First Philosophy: With Selections from the Objections and Replies*, edited by J. Cottingham (Cambridge Texts in the History of Philosophy). Cambridge: Cambridge University Press.
- Evans, Gareth. 1978. "Can There Be Vague Objects?" *Analysis*, 38(4): 208.
- Hasker, William. 1999. *The Emergent Self*. Ithaca, NY: Cornell University Press.
- Hasker, William. 2011. "Materialism and the Resurrection: Are the Prospects Improving?" *European Journal for Philosophy of Religion*, 3(1): 83–103.
- Jacobs, Jonathan D., and Timothy O'Connor. 2010. "Emergent Individuals and the Resurrection." *European Journal for Philosophy of Religion*, 2(2): 69–88.

- Kuhlmeier, Valerie A., Paul Bloom, and Karen Wynn. 2004. "Do 5-Month-Old Infants See Humans as Material Objects?" *Cognition*, 94(1): 95–103.
- Leibniz, Gottfried W. 1953 [1686]. *Discourse on Metaphysics*, translated by P. Lucas and L. Grint. Manchester: Manchester University Press.
- Luper-Foy, Steven. 2016. "Death." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. Accessed June 25, 2017. <https://plato.stanford.edu/entries/death/>.
- Moreland, J. P. 2009. *The Recalcitrant Imago Dei: Human Persons and the Failure of Naturalism*. London: SCM.
- Murdock, Paul. 1945. "The Common Denominator of Cultures." In *The Science of Man in the World in Crisis*, edited by R. Linton, 123–142. New York: Columbia University Press.
- Noonan, Harold W. 2003. *Personal Identity*. 2nd edn. Abingdon, UK: Routledge.
- Nozick, Robert. 1981. *Philosophical Explanations*. Cambridge, MA: Harvard University Press.
- Parfit, Derek. 1982. "Personal Identity and Rationality." *Synthese*, 53(2): 227–241.
- Price, H. H. 1964 [1953]. "Personal Survival and the Idea of another World." In *Classical and Contemporary Readings in the Philosophy of Religion*, edited by John Hick, 280–301. Upper Saddle River, NJ: Prentice Hall.
- Reid, Thomas. 1941 [1785]. *Essays on the Intellectual Powers of Man*, edited by A. D. Woozley. London: Macmillan.
- Salmon, Nathan U. 2005. *Reference and Essence*. Amherst, NY: Prometheus Books.
- Shoemaker, Sydney. 2012. "Against Simplicity." In *Personal Identity: Complex or Simple?* edited by Georg Gasser and Matthias Stefan, 123–136. Cambridge: Cambridge University Press.
- Swinburne, Richard. 2012. "How to Determine Which Is the True Theory of Personal Identity." In *Personal Identity: Complex or Simple?* edited by Georg Gasser and Matthias Stefan, 105–122. Cambridge: Cambridge University Press.
- Taliaferro, Charles. 1994. *Consciousness and the Mind of God*. Cambridge: Cambridge University Press.
- van Inwagen, Peter. 1978. "The Possibility of Resurrection." *International Journal for the Philosophy of Religion*, 9(2): 114–121.
- van Inwagen, Peter. 1990. *Material Beings*. Ithaca, NY: Cornell University Press.
- van Inwagen, Peter. 1995. "Dualism and Materialism: Athens and Jerusalem?" *Faith and Philosophy* 12(4): 475–488. DOI: 10.5840/faithphil199512444.
- van Inwagen, Peter. 1998. "Resurrection." In *Routledge Encyclopedia of Philosophy*, edited by Tim Crane. Abingdon, UK: Taylor & Francis. DOI: 10.4324/9780415249126-K088-1.
- van Inwagen, Peter. 2009. "The Possibility of Resurrection." In *The Oxford Handbook of Philosophical Theology*, edited by Michael C. Rea, vol. 2, 321–327. Oxford: Oxford University Press.
- van Inwagen, Peter. n.d. "I Look for the Resurrection of the Dead and the Life of the World to Come." Accessed: January 4, 2016. <http://andrewmbailey.com/pvi/Resurrection.doc>.
- Wiggins, David. 1968. "On Being in the Same Place at the Same Time." *Philosophical Review*, 77(1): 90–95. DOI: 10.2307/2183184.
- Williams, Bernard A. O. 1956. "Personal Identity and Individuation." *Proceedings of the Aristotelian Society*, 57: 229–252.
- Wright, N T. 2003. *The Resurrection of the Son of God*. London: SPCK.
- Wright, N. T. 2011. "Mind, Spirit, Soul and Body: All for One and One for All Reflections on Paul's Anthropology in His Complex Contexts." Society of Christian Philosophers: Regional Meeting, Fordham University, March 18. Accessed: January 4, 2016. [http://ntwrightpage.com/Wright\\_SCP\\_MindSpiritSoulBody.htm](http://ntwrightpage.com/Wright_SCP_MindSpiritSoulBody.htm).
- Zimmerman, Dean W. 1999. "Materialism and Survival: The Falling Elevator Model." *Faith and Philosophy*, 16(2): 194–212.
- Zimmerman, Dean W. 2010. "Bodily Resurrection: The Falling Elevator Model Revisited." In *Personal Identity and Resurrection: How Do We Survive Our Death?* edited by Georg Gasser, 33–50. Burlington, VT: Ashgate.

# I Look for the Resurrection of the Dead and the Life of the World to Come

PETER VAN INWAGEN

Most people in most cultures believe in a life beyond the grave. They tell stories about it. But not all cultures tell the same story. Some cultures tell stories of reincarnation or metempsychosis. In our Western culture there is a tendency to tell stories of the sort we see in the movie *Ghost* (you may remember it: Whoopi Goldberg, Patrick Swayze, and Demi Moore).

In this movie, newly dead people rise from their corpses, and have a kind of diaphanous existence. They look like human beings (to anyone who can see them at all), but they are able to pass through living people and walls and other solid things. (Why don't they fall through the floor, then? You may well ask.) And, of course, they are for the most part invisible to the living. Eventually, bright beings summon them to ascend a beam of light to heaven, or dark, gibbering creatures drag them screaming off to hell. This is, I am afraid, exactly the picture of the afterlife that is current among undergraduates at Notre Dame, although they might be willing to admit that the visual representation of disembodied souls in the movie was either symbolical or what might be called cinematic license. Most of them, every Sunday and major feast day, say the words, "I await the resurrection of the dead and the life of the world to come." And every time they are present at the baptism of a child, they promise to help the parents and godparents of the newly baptized bring the child up in a faith one of whose tenets is (they say these words), "I believe in the resurrection of the body, and the life everlasting." But these words mean nothing to them. They say them, but they are getting no more meaning out of them than a famous six-year-old did from another well-known text; reciting the Lord's Prayer, he said, "And lead us not into Penn Station." I once heard a speech by the former President of Notre Dame about the difficulties of teaching theology to Notre Dame undergraduates. President Malloy remarked sententiously that we cannot presuppose, as we once could, that our students will bring some degree of catechetical formation to the study of theology. I don't think he knows the half of it.

This picture of death and immortality, the Hollywood-and-Notre-Dame-undergraduate picture, is, I believe, very far from the biblical picture of death and immortality. According to the Bible, God formed us out of the dust of the earth and breathed life into us. When, in punishment for our rebellion against him we die and return to the dust out of which he raised our first parents, we're just, well, dead.

I am aware that in saying this I am taking a position denied vehemently by both Roman Catholics and Calvinists during the Reformation and the counter-Reformation. (Martin Luther seems at least sometimes to have accepted it, and it was the position of the Anabaptists.) But the psalmist says *this* of temporal rulers, and then extends it to every "child of earth" (Ps 146:3)

When they breathe their last, they return to earth  
And in that day their thoughts perish.

That's the rendition in the *Book of Common Prayer*.<sup>1</sup> Admittedly, the Jerusalem Bible, which you can usually rely on for a deflationary translation of some favorite passage, has "and all their schemes perish," which is no doubt true but is irrelevant to my thesis. On the other hand, the New English Bible, itself no slouch in the deflationary translation department, has "and all their thinking comes to an end," which *I*, naturally, like even better than "their thoughts perish."

Well, enough. If I tease this verse any more, I'll be accused of treating a text that, although inspired, represents the literary form "liturgical prayer" as if it represented the literary form "metaphysical essay," a practice I've complained of when others do it. Still, it can hardly be denied that in the Hebrew Bible we are represented as living dust, dust into which the spirit of God has entered. And when we die – this is the Hebrew picture – this dust becomes once more ordinary dust, and, as the preacher says, "the spirit returns to God who gave it." (A phrase curiously quoted by some as implying the immortality of the soul. But the spirit referred to here is the breath that God breathed into Adam's nostrils, the spirit, that, as the preacher had said a few verses back, comes to the bones in the womb of the woman with child – or that's one translation.) It is the spirit or breath of which the following is said in the second poem of Elihu in the book of Job (34:14, 15):

If [God] should take back his spirit to himself  
And gather to himself his breath,  
All flesh would perish together,  
And man return to dust.

In the New Testament, there is only one change in this picture, a piece of good news. None of the inspired descriptions of the nature of death in the Old Testament was wrong, the New Testament says, but these descriptions were not the whole story. In Christ, death retains its nature but its sting is drawn, for through his saving action, the dead will live again. And not (at least in the case of the saved) with the old Adamic life, but with a new resurrection life. And, as we all know, a part at least of this good news had been believed by many Jews for hundreds of years before Christ.

The belief in a future general resurrection of the dead arose in late second-temple Judaism (see, for example, Daniel 12:2 and John 11:24). (Whether there would be a resurrection of the dead was, of course, one of the main points that divided the Pharisees and the Sadducees.) When the new Christian movement appeared – before it was clearly

something other than a party or sect within Judaism – it centered on the belief that the crucified Jesus of Nazareth had been, in a literal, bodily sense, raised from the dead (*resurrectus*) and that *his* resurrection was, in some way, the means by which the expected general resurrection of the dead would be accomplished. Indeed, resurrection was so pervasive a theme in early Christian preaching that it was apparently sometimes thought that Christians worshiped two gods called “Jesus” and “Resurrection” (*Anastasis*). (Acts 17:18 is probably an allusion to this widespread misconception.) Belief in the resurrection of Jesus and a future general resurrection continue to be central to Christianity – although, given the sorry state of Christian theological education, it seems that if the Holy Spirit wishes to ensure that future Christians are even *aware* of this doctrine, he has his work cut out for him.

Christians have always insisted that resurrection is not a mere restoration of what the resurrected person had before death (as in the story of the raising of Lazarus) but is rather a doorway into a new kind of life. The status of a belief in the general resurrection in rabbinic Judaism is difficult to summarize. It should be noted, however, that a belief in the resurrection of the dead is one of Maimonides’s “thirteen principles,” which some Jews regard as a summary of the essential doctrines of Judaism. A belief in a general resurrection of the dead is one of many Judeo-Christian elements that have been incorporated into Islam.

The concept of the resurrection of the body (or of the dead) is most easily explained by laying out the ways in which it differs from the most important competing picture of the survival of death, the Platonic picture. According to Plato, when one dies (that is, when one’s body dies), one will continue to be what one has been all along, a soul: an immaterial center of consciousness, reason, and action. One’s death is, therefore, an extrinsic change in one: being dead means simply no longer having a body to animate. Since one’s death is an extrinsic change in one, one’s survival of death is something that happens in the natural course of events: one continues to exist after death by the continued exercise of the same powers or capacities that enabled one to exist when one still animated a body. (This inference is natural and plausible, but, as Descartes would later point out, it is not logically valid: for all logic can tell us, animating a body might be essential to the existence of a soul.) Death is, moreover, not a bad thing, as the vulgar believe, but a liberation, for the body is a prison of the soul – or it might be likened to a millstone that drags the soul down into the world of flux and impermanence. The liberation of the soul by death will not, unfortunately, be permanent, for the soul is destined repeatedly to suffer the misfortune of embodiment.

Christians and Jews (and Muslims) who believe in the resurrection of the dead will accept two of Plato’s theses about death: that the person does survive death, and that dead persons will not be forever disembodied. But everyone who believes in resurrection will dispute the following elements of Plato’s metaphysic of body, soul, and death: that the body is a prison; that the embodied soul has been disembodied in the past and will experience a large, perhaps infinite, number of “reincarnations” in the future. Christians, moreover, will insist that the new bodily life that awaits the soul (the saved soul, at least; perhaps this is not true of the damned) will not be of the same sort as its earlier life. The doctrine of resurrection, however, is no more than a doctrine. It is not a worked-out metaphysic of body, soul, and death. (The primary biblical data concerning the metaphysics of resurrection are found in I Cor. 15:35–55. The highly poetical language of this passage, however, is open to a variety of interpretations.) There are several competing philosophical theories of the metaphysics of resurrection. Some who accept the doctrine of resurrection deny the

existence of a separable, immaterial soul (Tertullian, for example, or, in the present day, the Scottish computer scientist D. M. MacKay and the English physicist J. C. Polkinghorne). Others accept the existence of an immaterial soul, but differ on the question whether the *person*, the “I,” is the immaterial soul. Aquinas, for example, sees the human person as essentially a composite of a human soul and a human body. According to the “composite” theory, a person cannot exist without a body: for one to exist is for one’s soul (always numerically the same) to animate some human body or other. (In the interval between one’s death and one’s receiving a new body at the time of the general resurrection, one’s soul exists and thinks and has experiences, but *one* does not, strictly speaking, exist.) Others who believe in a separable soul, however – most of the Fathers of the Church, and, probably, most Christians who have not given the matter much thought – accept a metaphysics of soul and body that is deceptively similar to Plato’s: one *is* an immaterial soul, and one will exist and think and have experiences throughout the interval during which one is without a body. But even the members of this party – the theologically well instructed among them, at any rate – would accept the following anti-Platonic theses: that the death of one’s “first” body is not a natural consequence of the impermanence of material things, but is rather a result of the Fall; that existing without a body is not a good thing for the soul, an essential part of the *telos* of which is to animate a body; that the life of the “spiritual” or “glorified” body that the saved soul will be given at the general resurrection will be qualitatively different from (and superior to) the life of the soul’s first or “natural” body. (It must be emphasized that, whatever “spiritual body” may mean, it does not mean “immaterial body.”)

According to this picture we have from the Fathers, the soul of a newly dead human being goes, or is sent, to heaven or hell. I think it is important to point out that there is no mention in the New Testament of “heaven” – not as a place one might go to upon death, or even, more abstractly, as a spiritual condition one might enter into upon death. In the New Testament, the word “heaven” (and it often occurs in the plural: “the kingdom of the heavens”; “Our father in the heavens”; my impression is that the singular is used only in opposition to “the earth”) seems to me mainly to mean – apart from its literal meaning, “the skies” – that part of creation where God is perfectly obeyed: at the present time, the abode of the blessed unfallen angels. (Of course, whatever Christians of the first century may have thought, today we must concede that an angel can have an “abode” only in some very abstract sense of the word.) When we have been raised imperishable, heaven in this sense will have been extended to comprehend the earth – even, we are told, the *kosmos* – and those who have been raised, who will then be in some respects like angels and who will be given authority over angels, can certainly say they are in heaven. But this idea is not the idea of a place to which bright beings can lead one *via* a shaft of light upon one’s death; nor is it the idea of a spiritual condition entry into which might be symbolized by a picture of bright beings leading one up a shaft of light upon one’s death.

And neither is hell a currently existing place or condition, according to the New Testament – unless, perhaps we use the word to refer to the abode of the rebel angels. Not that the word *hades* is ever used in this sense in the New Testament, but the plain sense of the New Testament is that there is such a person as Satan and that he is the prince of devils; so there is room for the concept. Those human beings who refuse salvation will, after the general resurrection, be thrust into a condition that the New Testament describes by a mixture of three images: exclusion (the “outer darkness” parables), pain, and refuse. The last two of these images are associated with pictures of fire, since burns are very painful and burning is one very common way to dispose of trash. We learn from Revelation, if nowhere



else, that the damned will come to share the condition of the rebel angels, so, if we use “hell” in the sense I have suggested, the damned will be able to say “We are in hell.” But this idea is not the idea of a place to which dark, gibbering creatures can drag one upon one’s death; nor is it the idea of a spiritual condition the commencement of which could be symbolized by a picture of dark, gibbering creatures dragging the newly dead person screaming away.

I will remark in passing that I have discussed the two New Testament texts that raise the greatest difficulty for the view I am defending – the parable of Dives and Lazarus, and the words of Christ from the Cross to the penitent thief – elsewhere (van Inwagen 1995, 283–284), and I’m not going to return to them here.

The picture I have presented strikes me as the biblical and creedal picture. As I have said elsewhere, if you have a more Platonic picture of the Christian afterlife, a picture of a disembodied soul facing the particular judgment immediately after death, I do not call you a heretic. I ask you only not to call me one either.

I now return to metaphysics. I must address a difficult metaphysical question. Is resurrection possible, given materialism? It can be plausibly argued that the doctrine of the resurrection of the dead presupposes some form of dualism. For if human persons are not immaterial souls, if they are living animals, then it would seem that death must be the end of them. A living animal is a material object. A material object is composed, at any given moment, of certain atoms. But if one is composed of certain atoms today, it is clear from what we know about the metabolisms of living things that one was not composed of those same atoms a year ago: one must then have been composed of a set of atoms that hardly overlaps the set of atoms that composes one today – and so for any living organism. This fact, the fact that the atoms of which a living organism is composed are in continuous flux, confronts the materialist who believes in resurrection with a grave metaphysical difficulty.

This difficulty is embodied in the following well-known argument. Suppose that God proposes to raise Socrates from the dead. How shall he accomplish this? How shall even omnipotence bring back a particular person who lived long ago and has returned to the dust? – whose former atoms have been, for millennia, spread pretty evenly throughout the biosphere? This question does not confront the dualist, who will say either that there is no need to bring Socrates back (because, so to speak, Socrates has never left), or else that Socrates can be brought back simply by providing his soul (which still exists) with a newly created human body. But what shall the materialist say? From the point of view of the materialist, it looks as if asking God to bring Socrates back is like asking him to bring back the snows of yesteryear or the light of other days. For what can even omnipotence do but *reassemble*? What else is there *to* do? And reassembly is not enough, for Socrates was composed of different atoms at different times. If someone says, “If God now reassembles the atoms that composed Socrates at the moment of his death, those reassembled atoms will once more compose Socrates,” there is an obvious objection to his thesis. If God can reassemble the atoms that composed Socrates at the moment of his death in 399 BCE – and no doubt he can – he can also reassemble the atoms that composed Socrates at some particular instant in 409 BCE. In fact, if there is no overlap between the two sets of atoms, God could do both these things, and set the two resulting men side by side. And which would be Socrates? Neither or both, it would seem, and, since not both, neither.

It might be objected that God would not do such a frivolous thing, and this may indeed be so. Nevertheless, if God were to reassemble either set of atoms, the resulting man would be who he was, and it is absurd, it is utterly incoherent, to suppose that his identity could depend on what might happen to some atoms other than the atoms that composed him. In

the end, there would seem to be no way round the following requirement: if Socrates was a material thing, a living organism, then, if a man who lives at some time after Socrates's death and physical dissolution is to be *Socrates*, there will have to be some sort of material and causal continuity between the matter that composed Socrates at the moment of his death and the matter that at any time composes that man. (St Paul seems to suggest, in the passage from 1 Corinthians cited above, that this will indeed be the case.) But "physical dissolution" and "material and causal continuity" are hard to reconcile. To show how the continuity requirement can be satisfied, despite appearances – or else to show that the continuity requirement is illusory – is a problem that must be solved if a philosophically satisfactory "materialist" theory of resurrection is to be devised.

Before I was a Christian, or a theist of any sort, when I was a sort of fellow traveler, I proposed a solution to this problem that has, let us say, not won wide assent. (This was in an essay called "The Possibility of Resurrection"; van Inwagen 1978, 1992.<sup>2</sup>) I suggested that God could accomplish the resurrection of, say, Socrates, in the following way. He could, in 399 BCE, have miraculously translated Socrates's fresh corpse to some distant place for safe-keeping (at the same time removing the hemlock and undoing the physiological damage it had done) and have replaced it with a simulacrum, a perfect physical duplicate of Socrates's corpse; later, on the day of resurrection, he could reanimate Socrates's corpse, and the reanimated corpse, no longer a corpse but once more a living organism, would *be* Socrates. Or, I suggested, he might do this with some *part* of the corpse, its brain or brain-stem or left cerebral hemisphere or cerebral cortex – something whose presence in a newly whole human organism would insure that that organism *be* Socrates.<sup>3</sup>

No one, as I say, was convinced. Some said, in effect, that the suggestion must be wrong because it was a very silly suggestion. I'm inclined to grant that very silly suggestions are at best very rarely correct. But I've never seen a defense of the *premise*. Some have made a more interesting objection: that my story represents God as a deceiver. But a deceiver about what? It's true that Socrates's friends believed that the corpse before them was *Socrates's* corpse, and that this belief was based upon a correct use of the sensory and cognitive apparatuses that God gave them. And if my story is true, their belief was false. Still, it must be remarked, we often do form false beliefs under just such optimal conditions, and I'm not, like Descartes, willing to say that each such case represents an abuse of free will. If God does replace each fresh corpse with a simulacrum, he does thereby show us an important truth: what death means, or what it would mean if he had not gone beyond justice, beyond mercy, and drawn death's sting in Christ. He shows us this, so to speak, counterfactually, but it is a counterfactual situation he's showing us. He's showing us what would have been if he were no more than a God of justice and had left us to the situation we had earned for ourselves by our rebellion against our creator. And that is a very important thing to be shown.

Some protest that this suggestion confuses resurrection with resuscitation, that it neglects the fundamental theological fact that the resurrection is a doorway into a new kind of life, and not the restoration of the old, Adamic life. But when I speak of God restoring a corpse to life, I mean to imply only that he causes a dead organism to become a living organism. I do not say that the once-more-living organism lives with the old kind of life. The resurrection life, as the post-resurrection stories of Jesus show, is a physical life, the life of an organism. After the resurrection, we shall still be composed of up-quarks, down-quarks, and electrons; but they will be organized somewhat differently. We shall be like a badly functioning mechanism that has been repaired and now functions perfectly – that is, functions, on a physical level, as its perfect designer intended it to function. I should

mention that it is very hard to separate the idea of the resurrection life from the idea of the Beatific Vision. There will be, I believe, a physical substrate to the life we shall live after the resurrection (this is why I used the words, “functions, *on a physical level*, as its perfect designer intended it to function”), but the most important feature of that life will be the fact that we who live with that life enjoy a direct vision of the divine life. The “physical substrate” is, I would suppose, a necessary condition for the enjoyment by the organism of this vision of God, but that is to say that it is of extrinsic rather than intrinsic importance; it is a necessary condition for something more important than itself.

Another objection to my story, and a serious one, one I had not considered (after all, I was not a Christian when I made this suggestion) is that its truth would make nonsense of the Pauline principle that a corpse is a temple of the Holy Spirit. If the suggestion is right, we never actually encounter any of these temples, and we might just as well (if we knew the truth) treat corpses with the same indifference as that which Socrates said was all *his* corpse would merit: dispose of them, literally without ceremony, in some sanitary and ecologically sound manner, and forget about them. I wonder, though, if this objection cannot be met by the version of my story according to which God has removed not the whole corpse, but only some small, identity-bearing part of it. After all, we still treat corpses with reverence, and properly so, when the undertaker has removed much larger parts of them than God removes according to this version of the suggestion.

I should like to remind you that, in any case, I did not say that the story I told about how God could achieve the resurrection was the true story, the story of what he actually did. My words were, the words “He could do it this way *if no other*.” I did not really emphasize this point, however. In fact, I said, speaking of the corpse-removal-and-simulacrum story, “I think this is the *only* way [God] could accomplish [the resurrection].” This was probably because I was not a Christian and was not discussing the doctrine of the resurrection with the same sense of attendant intellectual responsibility that would be felt by someone who actually accepted the doctrine. Now that I do feel this sense of intellectual responsibility in respect of what I say about the resurrection, I will explicitly say something a little more nuanced about the way I regard the story I told. The way I regard the story is best understood by considering the familiar distinction, familiar to students of the problem of evil, between a theodicy and a defense – these terms being used in the senses that Plantinga has given them. The story I have told is analogous to a “defense,” not to a theodicy.

Speaking on the metaphysics of resurrection today, as a believing Christian, I should not make any such definite statement as “I think this – the story of the corpse and the simulacrum – is the *only* way God could accomplish the resurrection.” My goal in “The Possibility of Resurrection,” was to argue for the metaphysical possibility of the resurrection of the dead. My method was to tell a story, a story I hoped my readers would grant was a metaphysically possible story, in which God accomplished the resurrection of the dead. But I was, I now see, far too ready to identify the reality of the Resurrection with what happens in the story I told to establish its possibility. I am now inclined to think that there are almost certainly other ways in which an omnipotent and omniscient being could accomplish the resurrection of the dead than the way that was described in the story I told, ways I am unable even to form an idea of because I lack the conceptual resources to do so. An analogy would be this: a medieval philosopher, or even a nineteenth-century physicist, could have formed no idea of the mechanisms by which the sun shines, not because these mechanisms are a mystery that surpasses human understanding, but simply because some of the concepts needed to describe them were not available before the twentieth century.

This analogy can be pressed a bit. Despite overwhelming evidence (provided by the fossil record) that there had been life on the earth for hundreds of millions of years, the great nineteenth-century physicist Lord Kelvin insisted that the sun had been shining for at most twenty million years. He maintained that the only conceivable mechanism of solar radiation was this: the sun is undergoing very gradual gravitational contraction, and solar radiation is due to the resulting gradual transformation of gravitational potential energy into radiant energy. When you plug the sun's mass, radius, and surface temperature into the appropriate equations (Kelvin contended), you will find that the sun cannot have been putting out radiant energy at anything like its current level for more than twenty million years. So (he concluded) the geologists and paleontologists have, demonstrably, drawn a false conclusion from their fossils and sedimentary layers.

Lord Kelvin's calculations were (I understand) correct: given his premise about the mechanism of solar radiation, his conclusion follows. Twentieth-century nuclear physics, however, has supplied the real mechanism of solar radiation, and we now know that Kelvin's premise and conclusion were both wrong and that the conclusion the despised geologists and paleontologists drew from the fossil record was right. Even in the nineteenth century, however, it would have been possible to show that Kelvin's premise and conclusion were not indisputable. Even within the confines of classical physics, it would have been possible to tell "just-so stories" according to which the sun had been shining for hundreds of millions of years. Here is the beginning of one: The sun is made up of rapidly *spinning* atoms; continual collisions between these atoms result in their kinetic energy of rotation being gradually transformed into radiant energy.

If one continues the story by specifying (for some particular moment in the past) the right average rotational kinetic energy for the solar atoms, and the right average linear velocity and mean free path of the atoms between collisions, and the right average loss of rotational kinetic energy in each collision, the resulting filled-out story will have the consequence that the sun has been producing light and heat at its present level for hundreds of millions of years – or for any period one likes.

This is, of course, a "just-so story": although it serves to establish a possibility, it isn't *true*. In fact – as Kelvin would certainly have been quick to point out – it is, miracles apart, a preposterous story, for no imaginable physical mechanism could have produced the initial conditions (the enormous rotational kinetic energy of the solar atoms) the story postulates. And yet, in a way, the story *is* true. There is one very abstract – and very important – feature that the sun-in-the-story shares with the real sun: most of the energy that the sun gives off in the form of light and heat was not stored before it was radiated as gravitational potential energy, but rather in the inner dynamics of the atoms of which the sun is composed. (In the story as kinetic energy of rotation; in the real world as nuclear binding energy.)

I am inclined now to think of the description that I gave in "The Possibility of Resurrection" of how an omnipotent being could accomplish the resurrection of the dead as a "just-so story": although it serves to establish a possibility, it probably isn't true. (And it is easy to see why someone might think it was preposterous, although it might be questioned whether any of us is in an epistemic position to make a judgment of this sort.) But I am also inclined to think that even if the story is not true, even if it gets the "mechanism" of resurrection wrong, it nevertheless *is* true – in a way. That is, I am inclined to think that, even if the story is wrong about the specifics of the Resurrection, the Resurrection-in-the-story, like the sun-in-the-story, nevertheless shares some important but very abstract feature of the real thing. My inclination is to believe that God will

somehow – in the way I have imagined or in some way I lack the conceptual resources to imagine, “in this way or some other,” – preserve a remnant of each person, a *gumnos kókkos* (a naked kernel: 1 Cor 15:37), which will be sown in corruption and raised in incorruption.

Are there any alternatives for the materialist? Is there any other way for a materialist to make sense of the doctrine of the resurrection of the dead? I want to examine one alternative, an alternative that is due to Lynne Rudder Baker (2001). Baker’s account of the resurrection depends on her general metaphysics of material objects, which she calls “constitution theory.” I will give a brief exposition of Baker’s constitution theory, as it is presented in her book *Persons and Bodies: A Constitution View* (Baker 2000).

Baker is a materialist as regards human persons: in her book, you and I and everyone we know is a material thing. But then how are we material persons related to our bodies, which are also material things? Unlike many materialists, including myself, she rejects the following answer to this question: We are *identical with* our bodies (or with some part of them, such as the brain). What, then is her answer to what she calls the “person-body question”? I will summarize her answer. My summary will use language very different from hers, since it will rely heavily on the language of parthood, and she is extremely hostile to any attempt to use the concept “part” in connection with her theory. Nevertheless, the use I make of this concept is innocuous, and my representation of her answer to the person-body question is accurate (as far as it goes: much is left out).

I begin by defining the notion of “temporally relativized mereological summation”:  $x$  is a mereological sum of the  $y$ s at  $t$  just in the case that all the  $y$ s are parts of  $x$  at  $t$ , and, at  $t$ , every part of  $x$  shares a part with at least one of the  $y$ s. The population of the natural world consists of fundamental particles and such mereological sums as any of the subsets of the set of all particles may have. For any particles whatever, there is a unique object that has the following modal properties: necessarily, it exists when and only when all those particles exist, and, necessarily, whenever it exists it is then a sum of those same particles. Such objects are called *aggregates* (of particles). Consider those particles – the  $P$ s – that are my ultimate parts at the present moment. It follows from the definition of “sum” that at this moment I am a sum of the  $P$ s. But I am not an aggregate, for my temporal extent and my modal properties are not those of any aggregate. And yet, if Baker is right, the  $P$ s now have a sum that is an aggregate, and have at every moment at which they all existed had that very same aggregate as a sum. (A year ago that aggregate was no doubt “smeared” across the terrestrial biosphere.) So, if Baker is right, *right now* the  $P$ s have at least two sums: one of them is I, an object that began to exist in 1942, and which has only very recently (perhaps within the last fraction of a second) become a sum of the  $P$ s; one of them is an aggregate, which has, no doubt, existed for millions of years, has always been a sum of the  $P$ s, and which, until recently – when it began the process, completed only a few seconds ago, of shrinking and congealing into a man-shaped, man-sized, solid object – has been an attenuated spherical shell of particles about eight thousand miles across and maybe a few hundred yards thick. What is the (present) relation between me and this aggregate? Well, the two of us are now composed of the same particles (and are therefore now spatially coincident). Can more be said? A great deal, according to Baker. The aggregate and I, she says, are now related by the ancestral of a relation (irreflexive, asymmetrical, and intransitive) called *constitution*. The aggregate now “constitutes” an object we have not yet mentioned, a *third* current sum of the  $P$ s, a living animal: my body. And my body, in its turn, now constitutes *me*. Each of these three current sums of the  $P$ s has different persistence conditions: the aggregate antedated both me and my body by millions of years

and will outlast us; my body antedates me (it was once a fetus and I have never been a fetus: for I am *essentially* a person, and a human body, when it is a fetus, does not yet have the right causal capacities to be – or to constitute – a person) and may outlast me (it will if it becomes a corpse or a “vegetable”).

What is this “constitution”? According to Baker, it is not a relation that holds only between bodies and human beings or aggregates and bodies. It is, she says, a relation that is *ubiquitous* in the material world. (Its relata are always mereological sums of particles. Baker appears to deny this vehemently in the section of her book called *Constitution and Mereology*. Her and my apparent disagreement on this point is only verbal. It arises because she builds much more into the meaning of “mereological sum” than actually follows from the above definition.) A certain piece of marble bears it to Michelangelo’s *David*; a certain piece of cloth bears it to the US flag that flies on the courthouse; a certain piece of colored paper bears it to this ten-dollar bill. (All these items are, at every moment at which they exist, then sums of certain fundamental particles.) When certain appropriately structured sums of particles are placed in appropriate circumstances, another sum of those same particles can be brought into existence, and the “new” sum will then be constituted by the preexistent and continuing sum. Depending on how the law is written, for example, it may be that at the moment I hand a legally mandated fee to a clerk at the Bureau of Motor Vehicles, a piece of colored plastic will begin to constitute a driver’s license; that is, there will begin to exist an object, a driver’s license, that is made of the same particles as the preexistent piece of plastic. But this statement requires qualification, for it suggests that the preexistent piece of plastic is *not* a driver’s license. Baker’s position is this: at the moment it comes to constitute a driver’s license, the piece of plastic will *become* a driver’s license. But it will differ from the “new” object it constitutes in this way: the new object will be a driver’s license *nonderderivatively*, and the preexistent object will be a driver’s license only *derivatively*: it will be a driver’s license in virtue of constituting something that is nonderivatively a driver’s license. The new object, moreover, will be *essentially* a driver’s license, and the preexistent object, the piece of colored plastic, will be a driver’s license only *accidentally* (for the first few minutes of its existence, it wasn’t one, and it might never have become one). Moreover, the nonderivative driver’s license, the new object, *is* a piece of plastic – but derivatively, only in virtue of being constituted by something that is nonderivatively a piece of plastic.

It is important to realize that, for Baker, being derivatively F is a species of being F. It’s *not* that, after the “real” driver’s license has come into existence, we apply the term “driver’s license” to the preexistent piece of plastic “vicariously” or “only in the loose and popular sense,” as a shorthand way of saying that something intimately related to the initial piece of plastic is a driver’s license. No, the preexistent piece of plastic now *really is* a driver’s license, and the new thing that is essentially a driver’s license *really is* a piece of plastic. The two objects are pieces/licenses *for different reasons*, but they are both literally and strictly pieces and licenses. A similar point applies to me and my body: I am a person nonderivatively and it derivatively, but it *really is* a person – and I *really am* a body, although I’m not the body we call “my body,” for this phrase denotes the thing that is nonderivatively a body. (Are there then two co-located persons? Well, there are two numerically distinct and co-located objects each of which is a person, but the two are not, Baker says, *separate* persons – or, indeed, separate objects. I doubt whether any sense can be made of this, but I will not press the point.)

Baker devotes the body of her book to explaining constitution (there is an elaborate Chisholm-style definition) and to meeting objections to “the constitution view” (both as a

general metaphysic of the material world and in its particular application to the person-body problem). I have only one major objection to “constitution theory”: I can’t bring myself to take seriously the idea that constitution is real. It seems to me as obvious as anything can be that if a piece of plastic becomes a driver’s license, that’s like a man’s becoming a husband: *entirely* a matter of a preexistent thing’s acquiring a new legal status. It seems equally obvious to me that there is nothing numerically distinct from me that is spatially coextensive with me. And Baker’s strenuous, extended, and very intelligent efforts to convince her readers that there are good reasons to believe these things move me not at all; I retain a complacent, unworried conviction that these things that seem obvious to me deserve to seem obvious to anyone who considers them. Well, that’s philosophy. Let us, however, set aside any worries we may have about the constitution theory, and examine Baker’s application of it to the problems with which the doctrine of the resurrection presents the materialist.

How does the “constitutionalist” view identity across time for human persons? First, a material thing is a human person if (i) it is a mereological sum of fundamental particles that have a member of the species *Homo sapiens* as one of their mereological sums, and (ii) it has (nonderivatively) a “first-person perspective.” Now consider me and my body. We are both human persons, but I am a human person nonderivatively, and it is a human person only derivatively; I have a first-person perspective nonderivatively and it has a first-person perspective only derivatively. I will describe this state of affairs by saying that I am a nonderivative human person – and the *only* nonderivative human person that currently shares my proper parts. In the sequel, I will use “person” to mean “nonderivative person” and “body” to mean “nonderivative body.” Suppose I die and am dissolved into my elements. What can God do to bring it about that some person, some person he brings into existence in the future, is I? (Baker agrees with me that if an organism, I mean a thing that is nonderivatively an organism, is totally destroyed, even God can’t bring it back into existence. That is why she thinks that constitution theory, according to which a human person is not, nonderivatively, an organism, enjoys a decisive advantage in the project of providing a materialist account of the metaphysics of resurrection over those materialist theories according to which a human person is, simply and without qualification, an organism.) Baker’s answer to the question “What can God do to bring it about that some person, some person he brings into existence in the future, is I?” is this: He must bring it about that this future person and I have the *same* first-person perspective. And this, Baker thinks, is something God can do without fooling about with corpses and simulacra, for he can cause the future person and me to have the same first-person perspective without there being any physical continuity between us. Identity of organisms, nonderivative organisms, would require physical continuity; identity of derivative organisms need not require physical continuity, and identity of those special derivative organisms that are human persons does not require physical continuity.

But what is it for *x* and *y* to have the same first-person perspective? Baker insists that no criterion of sameness of first-person perspective is possible, and that it would be a mistake to demand one. Perhaps that is so, but I am not asking for a criterion – whatever that means – but for a definition. I am asking what the words “*x* and *y* have the same first-person perspective” *mean*. Baker makes it plain that, in her view, the familiar distinction between descriptive and numerical identity applies to first-person perspectives. My *Doppelgänger* on Twin Earth and I have minds with identical content at each moment and can therefore be said to have *descriptively* identical first-person perspectives. But we have *numerically*

*distinct* first-person perspectives because, if for no other reason, our first-person perspectives are “directed at” different human bodies. In this language, the language of numerical identity, what God must do to make a post-resurrection person *me* is to provide him with a first-person perspective numerically identical with mine. (Merely descriptively identical first-person perspectives wouldn’t do, for two or even a hundred post-resurrection persons could have descriptively identical first-person perspectives.) But what *is* the numerical identity of first-person perspectives? “Well, you understand what it is for *x* and *y* to have first-person perspectives, and you understand numerical identity, so you must understand the sentence ‘the first-person perspective of *x* is identical with the first-person perspective of *y*.’” To paraphrase Wittgenstein, you understand “it’s five o’clock” and you understand the adverbial phrase “on the sun,” so you must understand “It’s five o’clock on the sun.” No, my antecedent understanding of “first-person perspective” and “is identical with” are not sufficient for my understanding “the first-person perspective of *x* is identical with the first-person perspective of *y*.” I need some sort of definition, some explicit statement of meaning. And, unfortunately, the only definition I can think of (Baker gives none, and would probably not agree with me that one was needed) is this:

The first-person perspective of *x* is identical with the first-person perspective of *y*  
 = df  
*x* has a first-person perspective and *y* has a first-person perspective and *x* is identical with *y*.

But if this is what identity of first-person perspectives *means*, then it’s hard to see how being told that God can make a post-resurrection person *me* by giving that person a first-person perspective numerically identical with mine explains anything – for an essential part of giving a person a first-person perspective identical with mine is making that person identical with me. And how God might do *that* is just what identity of first-person perspectives was supposed to help us to understand. I conclude that even if I know that I am a thing that has, nonderivatively, a first-person perspective and is only derivatively a living organism, and is spatially coincident with a nonderivative living organism that has a first-person perspective only derivatively, this rather astounding and certainly very important piece of metaphysical information will be of no help whatever to me if I want to reconcile a materialist theory of the human person with the Christian doctrine of the general resurrection. It seems to me that the materialist who believes in the general resurrection is, so to speak, stuck with saying that there must somehow be some sort of physical continuity between the person who dies in the present age of the world and the person who is raised on the day of resurrection. If human persons are physical substances, nothing but physical continuity can ground the identity of human persons across time. The problem for the “Christian materialist,” therefore, is to try to present a plausible theory according to which such physical continuity exists.

## Notes

1. The Standard and Revised Standard versions use similar wording, but place this text as verse 4 – eds.
2. The reprinted version (van Inwagen 1992) is slightly revised, and the revisions are improvements.
3. A variant on this suggestion can be found in Dean Zimmerman (1999). In Zimmerman’s “model,” an instant before Socrates’s death, God causes him to fission into two objects, *X* and *Y*. One of them, let us say *X*, dies and becomes what Phaedo *et al.* regard as Socrates’s corpse. God transports *Y* into the future, to the day of



Resurrection, and resurrects *Y*. I do not see what is accomplished by this elaboration of the story, however. *Socrates* is resurrected only if *Y* is *Socrates*. And if *Y* is *Socrates*, *X* is not *Socrates*. And if *X* is not *Socrates*, *X*'s corpse has no better claim to be *Socrates*'s corpse than would a simulacrum of *Socrates*'s corpse created *ex nihilo* or formed from the rearrangement of ambient nonbiological matter.

## References

- Baker, Lynne Rudder. 2000. *Persons and Bodies: A Constitution View*. Cambridge: Cambridge University Press.
- Baker, Lynne Rudder. 2001. "Material Persons and the Doctrine of Resurrection." *Faith and Philosophy*, 18(2): 151–167. DOI: 10.5840/faithphil20011821.
- van Inwagen, Peter. 1978. "The Possibility of Resurrection." *International Journal for Philosophy of Religion*, 9(2): 114–121.
- van Inwagen, Peter. 1992. "The Possibility of Resurrection." In *Immortality*, edited by Paul Edwards, repr. with rev., 242–246. New York: Macmillan.
- van Inwagen, Peter. 1995. "Dualism and Materialism: Athens and Jerusalem?" *Faith and Philosophy*, 12(4): 475–488. DOI: 10.5840/faithphil199512444.
- Zimmerman, Dean. 1999. "The Compatibility of Materialism and Survival: The 'Falling Elevator' Model." *Faith and Philosophy*, 16: 194–212.

# Index

- acquaintance, 33, 52, 198, 175  
aesthetic argument, 15, 287  
Alexander, Samuel, 383  
angel, 88, 89, 124, 235, 297, 421, 491, 125, 126, 128, 130  
animal, 16, 33, 64, 92–107, 125, 164, 170, 177, 194–196, 286, 297–305, 310–313, 318, 321, 356–359, 363, 364, 411, 428, 469, 492  
    human. *See* human, animal  
    kind. *See* kind, sensory  
    nonhuman, 13, 18, 43, 48, 64, 65, 94, 96, 97, 124, 129, 130, 164, 290, 298, 349, 360–362  
    rational, 9, 13, 88, 91, 92, 96, 418  
animalism, 9, 11, 16, 194, 196, 287, 297–299, 303–305, 307, 311–313, 357, 482, 484, 497  
animism, 417  
anthropology  
    biblical, 11, 19, 413–417, 423, 424, 427  
    Christian, 16, 318, 325, 394  
    dualistic, 419, 421, 475  
    holistic, 1, 19  
    medical, 429, 431, 432  
    monistic, 413, 418, 419, 424  
    New Testament, 420, 428–430  
    Pauline, 429  
    philosophical, 19, 96, 99, 100  
    physicalist, 395, 400, 404, 408  
    theological, 427, 429, 431, 434, 437  
Apollinarianism, 447, 460, 464  
appetite, 92, 96, 173, 321, 326  
Aquinas, Thomas, 4, 9, 45, 76, 88–102, 123–130, 298, 320, 345, 346, 348, 377, 399, 462, 491  
Aristotle, 9, 14, 97, 120, 298, 300, 320, 345, 346, 377, 379, 386, 399, 402, 418  
Armstrong, D. M., 23, 24, 25, 31, 56, 372  
artificial intelligence, 6  
aspects, 51  
associative identity disorder, 83  
atomism and consciousness, 188, 198–200, 210, 217–219  
Augustine, 9, 44, 76, 415  
Augustine, Keith, 230–234  
Averill, E., 27  
awareness, 18, 79, 109, 110, 129, 185, 189, 190, 202, 218, 219, 312, 313, 319, 333, 369  
    and first-person perspective. *See* first-person perspective, and awareness  
    conscious, 15, 81, 134, 227, 248, 252, 255, 256, 258, 259, 261  
    direct, 44, 155. *See also* acquaintance  
    fundamental phenomenological, 192  
    self. *See* self, awareness  
    unified. *See* unity of consciousness  
Ayer, A. J., 50

- Bailey, Andrew, 18, 359, 373  
Baker, Lynne Rudder, 4, 8, 17, 18, 20, 56, 287, 343, 347, 351–364, 394, 395, 397, 442, 447, 464, 471, 482, 483, 496–499  
Barnett, David, 14, 76, 194, 208, 213–215, 396, 399  
basic mentality thesis, 17, 328, 335–337  
beatific vision, 494  
behaviorism, 4, 5, 6, 23, 29, 31, 214  
Bennett, Karen, 30  
Bible. *See* Scriptures  
biblical anthropology. *See* anthropology, biblical  
biblical dualism, 414, 415, 421  
biblical holism. *See* holism, biblical  
binding problem, 82, 185, 321  
biological argument, 15  
biology, 1, 64, 105, 106, 116–118, 126, 170, 187, 205, 279, 318, 321–323, 399  
Bitbol, Michel, 55  
Blackmore, Susan, 234, 235, 237, 242, 321  
blindsight, 288–290  
Block, Ned, 6, 24, 214, 397  
body, 1, 2, 4, 8, 9, 13–20, 26, 29, 43–118, 128, 133, 137, 138, 144, 146, 150, 154–163, 168–174, 209, 210, 215, 299, 301, 309, 337, 341–351, 379–387, 398, 399, 414, 418, 431, 436, 441, 453, 454, 461, 471  
    and Thomism, 88, 89, 95–98, 107, 123–126  
    resurrection, 71, 344, 347, 422, 423  
brain, 6, 8, 12, 27, 29, 31–33, 46, 49, 56, 62, 65, 67, 69, 70, 75, 77–79, 107, 109, 110, 117, 144, 157, 163, 174–176, 188, 190, 198, 200, 201, 208, 214, 216, 228, 258, 262, 268, 272, 274, 275, 290, 321, 337, 342, 395–397, 404, 405  
damage, 6, 64, 196, 233, 277, 278, 289, 321, 352  
dead, 15, 138, 240, 241, 252, 256  
split. *See* commissurotomy  
temporal lobe, 218  
transplant, 14, 16, 222, 303, 305  
unity of, 82  
upper, 277, 233  
Burke, Michael, 444  
Cambridge changes, 481  
Cambridge Platonists, 47  
Campbell, Keith, 24, 30  
capacity, 92, 116  
    first-order, 362, 363  
    higher-order, 362, 363  
    in-hand, 362  
    remote, 362  
    second-order, 17, 342  
Cartesian substance dualism, 1, 9, 13, 14, 25, 29, 30, 33, 125, 153, 154, 168, 309, 377–379, 381, 445–447. *See also* substance dualism  
causal argument for materialism, 159, 162  
causal closure, 7, 35, 62, 67, 147, 276, 325, 336, 337, 406  
causal powers, 4, 8, 9, 25, 32, 91, 92, 124, 159, 193, 274, 337, 343, 371, 377, 380, 384, 387, 389, 390, 402, 403, 407, 445  
    active, 161  
    emergent, 18, 66–68, 74, 75, 81, 82  
    immanent, 388  
    passive, 161, 382  
causal reductionism. *See* reductionism, causal  
causal relations, 30, 48, 160, 162, 165, 201, 268, 275, 380, 389, 453, 454  
causation, 4, 12, 15, 27, 28, 43, 48, 49, 57, 113, 116, 154, 178, 337, 369, 377, 379  
    and explanation, 9, 10, 23, 62, 177, 146, 160, 161, 275, 381  
    and interaction, 4, 13, 22, 25–27, 29, 31, 32, 33, 48, 63, 65, 83, 115, 146, 154, 157–159, 161–163, 268, 275, 336, 337, 380, 381, 383, 388, 394  
    and the completeness of physics, 25, 28, 380, 390  
antirealism, 26  
counterfactual accounts of, 26, 48  
downward, 18, 318, 322, 324  
efficient, 82, 324, 384, 387–390  
formal, 387, 391, 399  
Humean, 26, 48  
immanent, 92, 399, 482  
mental, 13, 18, 43, 48, 49, 157–159, 377, 380, 405–407  
physical, 12, 27, 43, 48, 57, 160, 161, 406  
psychophysical, 7, 14, 170  
substance, 4  
transeunt, 92  
whole-part, 322  
Chalcedon, the council of, 358, 460  
Chalmers, David, 14, 30, 31, 32, 80, 81, 108, 140, 142, 185, 186, 188, 208, 212, 223, 268, 272, 400  
cheshire cat objection, 385, 386  
Chisholm, Roderick, 79, 309, 313, 497  
Chomsky, Noam, 49  
Christian physicalism, 15, 16, 394, 408

- Christology  
   and animalism, 441  
   and constitutionalism, 442  
   and materialism, 440–443, 445–447  
   and substance dualism, 446  
 Churchland, Paul, 26, 28, 29, 31, 32, 50–52, 270, 272  
 cladism, 354  
 cognition, 29, 280, 369, 395  
 combination problem, 80, 81, 83, 191  
 commissurotomy, 75, 83, 210, 211, 221  
 composition, 78, 285–287, 313, 357, 370, 387, 390, 476, 478  
 computational theory of mind, 5  
 conceivability, 13, 55, 156, 157, 209, 330, 472, 474  
   ideal, 54  
   prima facie, 54  
   secunda facie, 54  
 concept  
   causal, 24  
   formation, 401  
   transcendent, 403, 404  
 conceptual framework, 52  
 concretism, 445  
 consciousness, 3, 10, 15, 23, 29, 49, 51, 56, 64, 65, 69, 70–81, 111, 156, 184, 185, 199, 208, 214–216, 231, 243, 247, 256, 259, 285, 287, 288, 290, 291, 397, 399  
   cosmic, 248, 260  
   evolution of, 63  
   hard problem of, 15, 16, 93, 126, 292, 294, 400, 404  
   phenomenal, 3, 16, 191, 193, 202, 272, 286, 292–294  
   stream of, 194–197, 209–211, 395  
   unity of. *See* unity of consciousness  
 conservation, 65  
   of matter and energy, 25, 27, 28, 32, 148, 275, 380  
 constituent realism, 103, 201  
 constitution, 343, 347, 352, 354, 357, 385, 421, 482, 497, 498.  
   and the human person, 351, 361, 434, 483, 496. *See also* constitutionalism  
   and Thomism, 98, 103, 106, 115, 118  
 constitutionalism (constitution view), 17, 18, 341, 343, 344, 347–349, 351–353, 355, 356, 358–360, 363, 364, 482  
 contingency, 7, 8, 10, 11, 24, 55, 56, 77, 78, 140, 150, 192, 193, 287, 302, 333, 343, 403, 480  
 Corcoran, Kevin, 4, 8, 15, 286, 394, 395, 397–399, 403, 405, 442, 474  
 correlation, 7, 57, 91, 113, 176, 251, 257, 335, 336  
 corruptionism, 98, 99, 100  
 counting argument, 17, 329, 330–332, 334–337  
 Crane, Tim, 55  
 creation ex nihilo, 63, 77  
 creationism, 12, 62–64, 65, 71, 130  
 Cross, Richard, 441, 456  
 cryogenic freezing, 447  
 cybernetics, 323  
  
 dappled world picture, 389  
 Darwinian view of origins, 120, 205, 277, 279, 402  
 Davidson, Donald, 7, 269, 355  
 death, 9, 13, 43, 45, 95, 98, 150, 241, 477, 478, 493  
   annihilation, 95, 260, 479  
   clinical, 238. *See also* brain, dead  
   life after, 16, 55, 99, 100, 112, 123, 125, 150, 152, 165, 175, 241, 267, 273, 301, 302, 344–348, 358, 385, 387, 413–434, 446, 470–479, 489, 490, 492  
 decoherence, 390  
 de-escalation, 18, 384, 385  
 Dennett, Daniel, 2, 46, 47, 49–51, 173, 272, 308, 395, 396, 454  
 dependence, 15, 27, 31, 63–65, 76, 81, 83, 130, 289, 290, 328, 332, 357, 474  
   causal, 47, 116, 288, 382  
   diachronic, 379  
   metaphysical, 379  
   neural, 29, 31, 288  
   synchronic, 388  
 Des Chene, Dennis, 102, 104, 105  
 Descartes, René, 63, 79, 130, 138, 158, 169, 170, 415  
 design argument, 12, 278–280  
 designator  
   informative, 13, 141–143  
   rigid, 53, 139, 140  
   uninformative, 140–142  
 desires  
   first-order, 269  
   second-order, 269  
 determinism, 27, 318, 322, 324  
 DeWeese, Garrett, 445  
 dichotomism, 415, 417, 418  
 Dion/Theon objection, 385, 386

- disembodiment, 4, 13, 18, 19, 20, 31, 54–56, 77,  
     98–100, 123, 133, 156, 163, 196, 209, 311,  
     420, 422, 423, 447, 311, 345–348, 358, 382,  
     385, 413, 415–417, 419, 420, 421, 423, 446,  
     447, 454, 456, 457, 459, 470, 472, 490  
     and NDE, 15, 112, 235  
 Docetism, 460  
 doctrine of arbitrary undetached parts, 386,  
     443  
 downward causation. *See* causation, downward  
 dream, 15, 48, 136, 228, 247–249, 259  
 dreamless, 137  
 dual-aspect monism, 317  
 dualism. *See* dualism, biblical; emergent subject  
     dualism; holism, dualistic; neo-thomistic  
     dualism; property dualism; substance  
     dualism; thomistic dualism  
 duplication problem (argument), 356  
 dying brain hypothesis, 241  
  
 ego, 1, 2, 22, 26, 28, 30–32, 116, 169, 199,  
     483  
 eliminativism, 111, 269–271, 278  
 Elisabeth of Bohemia, 2, 158, 159, 275  
 embodiment, 57, 58, 347, 408, 413, 416, 420,  
     423, 447, 452–458, 461–464, 490  
     and Jesus, 19, 446, 462. *See also* incarnation,  
     Christology  
     primitive, 459  
     the virtues of, 12, 44, 45  
 embryo, 287, 288, 298, 361, 362, 441  
 emergence, 3, 4, 5, 8, 32, 65–67, 74, 77, 78, 80,  
     113, 118, 119, 130, 177, 187, 188, 273, 274,  
     370, 382, 384, 385, 387, 388, 395, 398, 400,  
     404, 445  
     of agency, 322  
 emergent dualism, 12, 67, 68, 70, 71, 73–83,  
     130  
 emergent individualism, 18, 369, 377–379, 381,  
     382, 391  
 emergent monism, 8, 317  
 emergent physicalism, 8–11, 407, 416  
 emergent subject dualism, 1, 3  
 emergent substance dualism, 12, 73  
 emergentism, 3, 7, 12, 13, 62, 65, 71, 74, 78, 274,  
     385, 406, 440  
 EMP (Emergent Material Persons theory), 67,  
     68, 69, 70  
 Epicureanism, 430, 435  
 epiphenomenalism, 7, 8, 25, 28, 66, 380  
 eschatology, 19, 413  
  
 New Testament, 415, 420, 422, 424  
 Old Testament, 419  
 two-stage, 419, 421, 423  
 essence, 74, 103–107, 114, 115, 117–119, 128,  
     129, 133, 137, 140, 143, 154, 155, 381, 402,  
     435  
     arguments from, 156, 157  
 essentialism, 13, 119, 130  
 Aristotelian, 103, 106, 118  
 kind, 19, 462, 463  
 mereological, 53, 111  
 origins, 379  
 species, 127  
 ethnomedicine, 431  
 Evans, C. Stephen, 57, 116  
 events, 2, 3, 4, 74, 143, 160, 195, 310, 312, 398,  
     465, 477, 481  
     emergent mental, 75, 143  
     life. *See* life, as an event  
     mental, 23, 26–29, 66–68, 106, 142–144,  
     146–149, 164, 204, 268, 276  
     physical, 7, 27, 32, 43, 46, 143, 144, 146–149,  
     268, 276, 337, 406  
 evolution, 28, 62–64, 67, 178, 297  
     and emergentism, 65, 113, 130  
     and property dualism, 31  
     and substance dualism, 31  
     and Thomistic dualism, 120, 125, 127, 129  
 exclusion problem (argument), 7, 8, 380, 406  
 experience, 3, 49, 50, 69, 110, 186, 190, 201  
     auditory, 217, 220, 221  
     conscious, 6, 15, 64, 67, 69, 80, 81, 130, 186,  
     200, 202, 256, 288–291, 294, 356  
     sensory, 109, 127, 128, 189, 456. *See also*  
     knowledge, sensory experiential  
     tactile, 200, 276, 291  
     unified conscious, 15  
     visual, 154, 200, 216–218, 220  
 explanation, 7, 8, 12, 23, 26, 29, 81, 146, 184, 186,  
     187, 278, 292, 293  
     and causation. *See* causation, and explanation  
 explanatory power, 12, 77, 78, 116, 117, 244  
     argument from, 12, 276, 278, 279  
 extended mind hypothesis, 268  
  
 faculty, 95, 113, 114, 116, 169, 209  
 falling elevator model, 20, 479, 480  
 fetus, 125, 126, 128, 130, 146, 148, 497  
 first-person perspective, 4, 8, 199, 222, 269, 271,  
     286, 287, 292, 293, 342, 347, 348, 352, 353,  
     359, 360, 394, 395, 400, 483, 498, 499

- and awareness, 30  
 robust, 17, 342, 360–363  
 rudimentary, 342 360, 361
- fission, 196, 479
- Flint, Thomas, 443
- folk psychology, 66, 210, 270
- form, 88–90, 97, 98, 118, 123–125, 380, 386, 387, 418  
   accidental, 90, 91, 123, 124  
   subsistent, 124–129, 415  
   substantial, 9, 88, 90, 91, 94, 96–98, 100, 123–126, 385
- free will, 18, 269–271, 273, 369, 377
- full body amputation, 13, 95
- Fumerton, Richard, 52, 53
- functionalism, 6, 214
- fundamentality, 322, 352, 370–372, 391
- ghost in the machine, 22, 76
- gnosticism, 415, 420
- Goethe, 49
- Goldbach's conjecture, 156
- grounding, 81, 111, 118, 189, 192, 332–334, 354, 357, 370  
   metaphysical, 193, 378–380, 390  
   partial, 374
- grounding problem, 352, 353, 355
- haecceity. *See* thisness
- hard problem of consciousness. *See* consciousness, hard problem of
- Hasker, William, 3, 12–14, 30, 54, 57, 63, 67, 71–83, 108, 117, 188–208, 216, 217, 219, 274, 407, 415, 446, 477, 480–482
- hell, 249, 421, 488, 491, 492
- hermeneutics, 19, 414, 415  
   neuro-hermeneutics, 414
- Hill, Jonathan, 441, 446
- holism, 106, 118, 199, 200, 384, 417, 418, 424  
   and consciousness, 109, 188, 189, 195, 199, 200, 211  
   biblical, 19, 413  
   dualistic, 414  
   phenomenal. *See* phenomenal holism
- holistic dualism. *See* holism, dualistic
- homunculus, 202  
   Penfield, 291
- human,  
   animal, 287, 298, 304, 305, 351–361  
   being, 13, 88, 91–100, 114, 124–126, 150, 170, 179, 286, 319, 278, 379, 381, 382, 385–388, 415, 417, 430, 443, 445, 447, 462  
   kind, 92  
   nature. *See* human nature  
   organism, 305, 343, 463, 464, 482  
   person. *See* person, human
- human nature, 17, 89, 96, 114, 317, 320, 326  
   aspective, 319, 320  
   partitive, 319, 320
- Hume, David, 4, 30, 63, 267, 275, 278
- hylomorphism, 104, 377, 385  
   Aristotelian, 104  
   Thomistic, 18, 377–379, 381, 382, 389
- idealism, 370, 380
- identity, 7, 20, 46, 58  
   closest-continuer account of, 480, 481, 482, 483  
   diachronic, 174  
   necessity of, 480  
   numerical, 312, 470, 498, 499  
   personal. *See* personal identity  
   relative, 53, 386, 440  
   vagueness. *See* of vagueness, identity
- identity theory, 5, 6, 22, 274. *See also* token  
   physicalism; type physicalism
- illusion, 14, 63, 64, 130, 142, 143, 147, 215, 216, 252, 255, 299
- imagination, 50, 54, 55, 95, 171, 277, 280
- immaterial, 1, 2, 13–16, 19, 26, 27, 30–32, 67, 75, 89, 124, 126, 130, 153, 154, 158, 160, 162, 163, 170
- impenetrability of matter, 162, 163
- incarnation, 17, 19, 384, 440, 452, 454, 457  
   and materialism, 19, 440, 443, 445, 447  
   and physicalism, 440, 461–464  
   and substance dualism, 17, 19, 58, 440, 446
- incorporeal. *See* immaterial
- indexicals, 116, 195  
   and constitutionalism, 363, 364
- indiscernibility of identicals, 5, 46, 49, 53
- indubitability  
   argument from, 155
- inseparable part. *See* parthood, inseparable part
- integrative dualism, 44, 76
- intellect, 89, 92, 94–96, 99, 105
- intellection, 13, 110  
   and immateriality, 13, 89
- intentionality, 6, 7, 10, 18, 77, 126, 203, 342, 369, 400

- interactionism. *See* causation, and interaction  
intermediate state, 77, 319, 341, 345, 348, 358, 385, 417, 420, 422, 424  
internal relations, 83, 106, 114, 115, 119, 188, 191, 200, 202  
introspection, 16, 24, 33, 111, 169, 210, 270, 271, 277, 300  
intuition, 56, 196, 221, 222  
    modal, 56, 112, 196  
irreflexivity of proper parthood, 386
- Jackson, Frank, 5, 28, 49, 187
- Kant, Immanuel, 9  
Keating, B., 27  
Kenny, Anthony, 55  
kenosis, 457  
Kim, Jaegwon, 4, 7–9, 14, 26, 66, 147, 148, 187, 322, 406, 407  
kind  
    essentialism. *See* essentialism, kind  
    human; human, kind  
    sensory, 92  
    stable, 402  
    vegetative, 13, 92  
Klein, Stan, 47  
knowledge, 16, 25, 47  
    by acquaintance. *See* acquaintance  
    first-person, 408, 473  
    immediate, 455  
    infallible, 135, 136  
    introspective. *See* introspection  
    of external reality, 400  
    of God, 403, 404  
    of other minds, 29  
    propositional, 456  
    self, 400  
    sensory experiential, 456  
knowledge argument, 49, 50, 57, 272  
Kripke, Saul, 139, 140, 142
- laws  
    causal, 119  
    emergent, 74, 79  
    psychophysical, 7, 76–78, 333, 336  
    teleological, 119  
Lazarus, 422, 490, 492  
Le Poidevin, Robin, 441  
Leftow, Brian, 442, 444, 445, 462  
Leibniz, Gottfried Wilhelm, 68, 154, 171, 208, 212, 213, 289  
Levine, Joseph, 212, 292  
Lewis, David, 5, 23–27, 56, 57, 222, 269  
libertarianism, 18, 74, 116, 193  
life, 32, 310, 312, 313  
    as an event, 16, 310, 312, 398, 477, 478  
    and persistence, 16, 312, 398, 476–478  
life after death. *See* death, life after  
lived life, 269, 271  
localized property instantiation, principle of, 69  
location problem, 187, 188  
Locke, John, 49, 398, 472  
logic, 3, 11, 48, 93, 490  
Lorber, John, 79  
Lowe, E. J., 4, 14, 57, 171, 210, 357, 371  
Lund, David, 50, 53–56  
Lyons, William, 204
- Marmodoro, Anna, 446  
materialism, 20, 22, 25, 28, 31, 43, 45, 47, 51, 66, 58, 73, 153, 208, 211–213, 223, 285–288, 416, 440. *See also* physicalism  
matter, 46, 49, 55, 67, 71, 80, 90, 91, 98, 107, 162, 163, 187, 345, 346, 351, 418, 427  
    prime. *See* prime matter  
McCord Adams, Marilyn, 261  
McGinn, Colin, 56, 184, 294  
Mellor, D. H., 55  
mental  
    causation. *See* causation, mental  
    disorder, 277, 278  
    events. *See* events, mental  
    powers, 48, 304  
    primacy of the, 43, 48, 49, 50, 51  
    properties. *See* properties, mental  
    qualities, 3  
mental causation. *See* causation, mental  
mereological essentialism. *See* essentialism, mereological  
metamorphosis, 478  
metaphysical aristotelianism, 104  
mini-entailment, 134, 135, 141–143  
modal argument for substance dualism, 12, 53–57, 112, 113  
modality, 139, 142, 143  
modes, 45, 168, 169, 189, 371  
monism, 44, 153, 413, 414, 416, 421  
    dual-aspect. *See* dual-aspect monism  
    emergent. *See* emergent monism  
    metaphysical, 416  
    multi-aspect, 325

- neutral. *See* neutral monism  
 priority, 352  
 psychosomatic, 413  
 monophysitism, 460  
 Moore, G. E., 48  
 morality, 11, 96  
     argument from the existence of, 273  
 Moreland, J.P., 9, 10, 13, 14, 57, 102, 103,  
     127–129, 198, 201, 203, 243, 313, 396, 415,  
     483  
 multiple realization, 5, 6, 407  
 mystery, 10, 161, 184, 310, 397, 463, 494  
  
 Nagel, Thomas, 5, 10, 49, 120, 184, 205,  
     271, 272  
 naturalism, 4, 10, 11, 55, 186, 187, 204,  
     317, 416  
 NCC (Neural Correlates of Consciousness), 218,  
     220, 291  
 NDE (Near Death Experience), 15, 112, 113,  
     227–240, 248, 255, 256  
     and blindness, 15, 112, 231, 235, 237–241, 243  
     and distant vision, 235  
     and toxic delirium, 248  
 necessary being, 370, 441  
 necessity, 104, 140, 142, 194, 197, 332, 370, 480  
 neo-thomistic dualism, 11, 102, 103, 107, 114,  
     115, 117, 125, 127  
 Nestorianism, 19, 444–446  
 neural plasticity, 290, 291, 324  
 neuropsychological argument, 15  
 neuroscience, 1, 17, 23, 26, 28, 57, 106, 116, 149,  
     219, 220, 270, 271, 277, 290, 291  
     cognitive, 17, 318, 320, 325, 335–337, 395  
 neutral monism, 416  
 New Testament, 19, 70, 250, 344, 414, 415,  
     417–422, 427–431, 433–435, 489, 491, 492  
     1 Corinthians, 344, 345, 347, 423, 490, 493  
     1 Peter, 434–436  
     1 Thessalonians, 421, 423, 427, 429  
     1 Timothy, 433  
     2 Corinthians, 422, 423  
     Acts, 420–423, 490  
     Galatians, 433  
     Hebrews, 260, 421  
     James, 433  
     John, 422, 424, 431, 489  
     Luke, 421–424, 431–433  
     Mark, 420, 422, 424  
     Matthew, 421, 422, 465  
     Philippians, 422, 423  
     Revelation, 421, 424, 491  
     Romans, 433  
 non-Cartesian substance dualism, 168, 309  
 non-reductive physicalism, 17, 66, 68,  
     317–320, 322, 325, 335, 377–379, 381,  
     413, 416  
  
 OBE (Out of Body Experience), 15, 228,  
     247–263  
 objects  
     basic, 55, 370–374  
     fundamental, 48, 371  
 Ockham's Razor, 53, 269, 278  
 Oderberg, David, 98, 99, 354  
 Old Testament, 417–419, 421, 489  
     afterlife, 417–419  
     Daniel, 419, 420, 489  
     Ezekiel, 419  
     Genesis, 298, 417–419  
     Isaiah, 419  
     Job, 489  
     Leviticus, 432  
     Psalms, 319  
     Samuel, 418, 419  
 Olson, Eric, 8, 16, 287, 299–313, 352, 355,  
     356, 442  
 “only X and Y” principle, 480  
 organicism, 4, 6, 11, 13, 22–24, 30  
 organism, 13, 104–107, 113, 114–119, 129  
 overdetermination  
     and emergentism, 398  
     causal, 25, 28, 325, 406, 407  
 ownership. *See* self, and ownership  
  
 pain, 3, 5, 6, 9, 24, 25, 52, 127, 128, 146, 155, 157,  
     184, 213, 251, 273, 304, 305, 309, 310, 356,  
     359, 447, 460  
 pairing problem, 9, 14, 26, 30, 113, 159,  
     161–163, 165, 391  
 Panglossianism, 28  
 panpsychism, 74, 80–82, 119, 188, 191  
 Papineau, David, 25, 272  
 paradox of increase, 442, 443  
 parsimony, 23, 25  
 parthood  
     inseparable part, 12, 14, 83, 102, 103, 106, 109,  
         113, 114, 117–119, 186–189, 199, 201–203,  
         397  
     separable part, 12, 16, 76, 81, 82, 102–104,  
         109, 111, 115, 186–188, 195, 199, 200–202,  
         308–310, 312, 313, 396–398, 400, 405



- particular
  - bare, 103, 373
  - thick, 103, 118, 119
  - thin, 103–105, 108, 118, 119
- parts,
  - substantial, 171–174, 210–212
- Pasnau, Robert, 102–104, 105
- patterns, 322
- PCC (Principle of Causal Closure). *See* causal closure
- Penfield, Wilder, 291
- Peoples, Glenn Andrew, 445
- perception, 33, 68, 110, 154, 157, 164, 172, 174–176, 190, 209, 212, 248, 257, 276–278, 288, 289, 309
- persistence, 8, 9, 16, 111, 123, 169, 170, 174, 312, 313, 352–354, 381, 385, 395, 398, 402, 403, 405, 471, 472, 478, 483
  - of living bodies, 16, 477, 478. *See also* life, and persistence
  - of persons, 170, 473, 476, 479, 484. *See also* personal identity
- person, 16, 45, 53, 68, 73, 75, 100, 103, 168–171, 188, 196, 285–287, 297, 342, 352, 355, 424, 428, 444, 462, 491
  - as mode, 45
  - as phase, 45
  - corporate, 173, 174
  - divine, 454–456, 458, 460
  - functional unity of the, 44, 45
  - human, 154, 158, 179, 287, 297, 298, 304, 326, 341, 342, 351–353, 356, 361, 364, 394, 395, 429, 444, 453, 464, 483, 498
- personal identity, 1, 20, 203, 221, 222, 345, 348, 420, 428, 433, 471–474, 476, 479, 481, 483, 484
  - and constitutionalism, 347, 482
  - the complex view of, 20, 221, 472–474, 482, 483
  - the psychological view of, 88, 203, 428, 473
  - the simple view of, 20, 472–474, 476, 479
- personhood, 75, 103, 287, 297, 360–362, 413, 435, 444–446
- phantasms, 388
- phantom limb, 15, 228, 290, 291
- Pharisees, 344, 415, 419–422, 424, 489
- phenomenal holism, 109, 189, 217
- phenomenal unity
  - objectual, 110, 185, 203
  - subject, 185, 203
  - subsumptive, 108, 185, 186, 188, 193, 203
- phenomenalism
  - naïve, 194–198
  - substrate, 194, 197
  - virtual, 14, 194, 197, 198
- Philo, 420, 430, 434–436
- philosophy of mind, 1–6, 11, 17, 18, 45, 73, 93, 106, 115, 164, 168, 184, 204, 217, 269, 274, 322, 329, 378, 407, 428, 484
- phlogiston, 2, 117
- physical, 5, 9, 26, 27, 43, 46–49, 55, 63, 66, 110, 143–145, 171, 286, 328, 329, 370
- physicalism, 3, 5, 7–10, 15, 17, 43, 66, 153, 204, 268, 273, 285, 286, 317, 328, 335, 378, 394, 400, 440, 461. *See also* Christian physicalism; emergent physicalism; non-reductive; reductive physicalism; token physicalism; type physicalism
- pineal gland, 29, 157–159, 163, 211
- Place, Ullin, 22
- Plantinga, Alvin, 289, 460, 462, 494
- Plato, 19, 44, 76, 95, 120, 307, 330, 345, 402, 413, 418, 427, 430, 436, 490
- platonic dualism, 125
- platonism, 96, 100, 415, 420, 421, 435
- pluralism, 44
- possibility, 13, 31, 54, 55, 113, 116, 156, 177, 474
  - and imagination, 54
  - epistemic, 133
  - logical, 13, 139, 149
  - metaphysical, 13, 31, 54, 76, 113, 139, 156, 157, 308, 475, 476, 479, 484, 494
- possible worlds, 7, 56, 139, 153, 154, 156, 480
- potentiality, 90
- powerful qualities. *See* qualities, powerful
- powers
  - causal. *See* causal powers
  - immaterial, 75, 93
  - intellectual, 93
  - mental. *See* mental powers
  - pure, 371
- pre-established harmony, 25, 380, 384
- Price, H. H., 48, 472, 473
- prime matter, 88, 90, 91, 103, 123, 125, 126, 128
- primitive, 371
  - body-soul union, 128, 159, 163, 164
  - conceptually, 214
  - embodiment. *See* embodiment, primitive
  - persistence, 174, 175
  - psychophysical law, 32
- principle of credulity, 148, 474

- principle of individuation, 162
- principle of uniformity, 329, 332, 333
- priority principle, 18, 359, 360
- private language, 2, 57, 115
- problem of causal interaction, 4. *See* causation, and interaction
- problem of complexity, 336
- problem of interaction. *See* causation, and interaction
- problem of the many, 78
- proper function, 114
- properly basic belief, 111, 116, 190
- properties, 201
  - additive, 109
  - basic, 68, 330
  - behavioral, 66, 301
  - emergent, 77, 109, 119, 187, 189, 274, 374
  - emergent mental, 68
  - essential, 137, 144
  - functional, 66
  - fundamental, 374
  - immaterial, 30–33
  - impure, 483, 486
  - mental, 7, 17, 18, 22, 30, 31, 63, 66–68, 75, 78, 81, 144, 145, 154, 164, 165, 195, 199, 211, 220, 268, 274, 303–305, 329–335, 359, 360, 388, 406, 407, 453, 462, 464, 465, 483
  - modal, 17, 195, 298, 351, 353, 354, 402, 482, 496
  - natural, 371, 372, 374, 375
  - neurophysiological, 66, 180
  - nonphysical, 30, 32, 268, 272, 329, 332, 378
  - physical, 4, 7, 14, 16, 17, 27, 32, 56, 63, 66, 75, 144, 164, 165, 191, 192, 268, 272, 274, 275, 329–335, 353, 378, 394, 407, 442, 453, 461–463, 472
  - relational, 18, 160, 324, 353–355
  - structural, 68, 187
- property dualism, 12, 30, 31–33, 67, 106, 115, 164, 165, 208, 212, 268, 272–274, 462
- propositional attitudes, 93, 210, 270
- proprioception, 44, 276, 310
- protopanpsychism, 74, 75, 80, 81
- psychology, 1, 5, 11, 66, 170, 173, 176, 204, 210, 270, 277, 399, 429, 431, 432
- purgatory, 341, 348
- PUT (Phenomenal Unity of consciousness Thesis), 191, 192
- Putnam, Hilary, 5, 139, 142, 176
- Pythagoras, 44
- qualia, 6, 8, 10, 30, 51, 77, 93, 377
  - inverted, 7
- qualities, 3, 162, 272, 324, 371, 382
  - powerful, 372
- quantifier shift fallacy, 169
- quantum indeterminacy, 324
- rational animal. *See* animal, rational
- rationality, 10, 49, 92, 95, 96, 120, 126
- Rea, Michael, 127, 128, 357, 402, 440
- realism
  - ontological, 377
  - scientific, 371
- reason, 29, 45, 92, 93, 126, 276, 278, 404, 405
  - practical, 404
  - theoretical, 404
- reduction, 17, 62, 66, 335, 348
  - explanatory, 335
  - microphysical, 335
  - ontological, 335
- reductionism, 12, 17, 62, 63, 65, 71, 204, 205
  - biological, 178, 318
  - causal, 318, 325
  - neurobiological, 321, 322
  - ontological, 62
- reductive physicalism, 5, 10, 17, 164, 165, 269, 274, 378
- Reid, Thomas, 43, 49, 313, 473
- reincarnation, 273, 488
- relations
  - basic, 371
  - causal. *See* causal relations
  - external, 8, 12, 82, 83, 187, 191–200, 217, 323
  - fundamental, 371
  - internal, 202
- resurrection, 4, 11, 19, 70, 71, 229, 358, 407, 414, 420–424, 470, 471, 477, 490
  - and animalism, 479, 482, 484
  - and constitutionalism, 17, 341, 344, 345, 347, 348, 483
  - and emergent substance dualism, 71, 76
  - and materialism, 20, 475–499
  - and reassembly, 470, 471, 492
  - and substance dualism, 150, 415, 460, 474
  - and Thomism, 98, 123, 125, 345, 346
  - body. *See* body, resurrection
- resuscitation, 231–240, 256, 477, 479, 482, 493
- Robinson, Daniel, 57
- Robinson, Howard, 48, 57, 269, 270
- Ross, James, 93

- Russell, Bertrand, 49, 275  
Ryle, Gilbert, 2, 4, 22, 24, 26, 29, 33, 44, 57
- Sabom, Michael, 112, 228, 234, 236, 240, 241, 248, 250, 262  
Sadducees, 344, 420, 421, 489  
salvation, 14, 18  
scientism, 25, 205, 269, 271, 278  
scriptural mandate, 394, 401  
Scriptures, 18, 19, 70, 317, 319, 341, 414, 417, 427, 428, 430, 431, 435–437, 489  
Searle, John, 4, 6, 78, 81, 193, 205, 329, 335  
Second Temple Judaism, 420  
self, 9, 16, 50, 81, 108, 109, 168, 171–180, 185, 189, 194, 197–199, 202, 204, 212, 213, 221, 395, 398, 423, 430, 432  
    and ownership, 203  
    and perspective, 193  
    and reference, 194, 197–199, 202, 203  
    and stream of consciousness, 14, 194, 195, 197. *See also* consciousness, stream of  
    awareness, 53, 116, 198, 202  
    emergent, 67, 75. *See also* emergent dualism  
    immaterial, 1, 67, 75, 108, 213, 237, 243  
    simplicity of. *See* simplicity, of self  
    substantial, 1, 108, 197, 199, 202–204  
    unity of, 174  
self-knowledge. *See* knowledge, self  
Sellars, Wilfrid, 24, 32  
sensation, 24, 45, 63, 92, 95, 96, 157, 164, 291  
sense data, 32, 234  
sensus communis, 321  
separable part. *See* parthood, separable part  
sex, 13, 95, 96  
Shrader, Warren, 190, 193  
simple argument, 76, 112, 278  
simplicity  
    and theory choice, 26, 67, 278  
    epistemic, 107  
    of self, 14, 173, 174, 203, 213, 223, 224  
    of the soul, 75, 76, 103, 104, 116, 188, 200, 309, 312, 313, 387, 404, 470, 483  
simulacrum, 20, 474–479, 481, 493, 494  
sleep  
    deep, 259  
    narcolepsy, 259  
    paralysis, 259  
    REM, 247, 249, 258, 259  
Smart, J. J. C., 5, 23, 26, 188, 274  
Snowdon, Paul, 133, 303, 363  
socio-cultural inquiry, 431  
sorites, 80, 83, 113  
soul. *See also* simplicity, of the soul  
    animal, 125  
    as form, 9, 76, 89, 91, 96–100, 118, 123–128, 345, 346, 380, 386, 387, 418, 462  
    as immaterial, 1, 2, 13, 16, 20, 28, 43, 67, 68, 71, 75, 76, 89, 100, 106, 108, 124–130, 157, 158, 161–169, 291, 341, 345, 346, 349, 397, 416, 428, 429, 441, 470, 471, 490, 491  
    Cartesian, 75, 445  
    emergent, 74–76  
    location of the. *See* soul, nonspatiality  
    nonspatiality, 14, 24, 26–33, 115, 158, 161–163, 442  
    nutritive, 399  
    rational, 103, 128, 321, 345, 346, 358, 418, 446, 447, 460, 464  
    sensitive, 124, 126, 128  
    sleep, 319  
    substantial, 13, 73, 110, 190, 197, 417  
    unity of the, 82, 83  
    vegetative, 124, 321  
species, 103, 204, 286, 297, 354, 355, 378  
    essentialism, 127, 130  
    of substance, 103, 123  
spiritual being, 341, 342, 408  
Sprigge, T. L. S., 49, 50  
statue, 8, 157, 179, 201, 343, 353, 442, 482  
stewardship, 16, 394, 400, 401, 404  
Stich, Stephen, 270, 277  
Stoicism, 430, 435  
Strawson, Galen, 30, 32, 49  
structural facts, 81, 286  
structural properties. *See* properties, structural  
Stump, Eleonore, 91, 123–126, 130, 131, 415  
Suarez, Francisco, 103  
subject, 1, 3, 4, 8, 10, 51, 67, 75, 155, 185, 189, 197, 203, 288, 363, 364  
    conscious, 2, 51, 212, 218, 397  
    mental, 4, 8, 397, 398, 404, 405  
    unity, 191, 395, 404  
subject dualism, 1, 208  
subjectivity, 3, 6, 7, 8, 10, 12, 47, 77, 202, 285, 286, 292, 293, 404  
    argument from, 155  
subsist, 9, 97, 98, 345, 346, 418, 419. *See also* form, subsistent  
substance, 75, 88–91, 96, 98, 100, 103, 119, 136, 138–140, 145, 169, 174, 202  
    biological, 168, 170  
    composite, 18, 385

- ul style="list-style-type: none; padding-left: 0;">
- emergence of, 68
- emergent, 73
- immaterial, 26, 32, 130, 170, 196, 202, 223, 441, 470, 471, 473, 483
- material, 89, 97
- mental, 4, 9, 17, 110, 144, 162–164, 190, 221, 328, 336, 337
- physical, 144, 169, 268, 272
- psychological, 168, 169, 170
- spiritual. *See* immaterial; soul, as immaterial; substance, immaterial
- substance dualism,
  - cartesian. *See* cartesian substance dualism
  - emergent. *See* emergent substance dualism
  - integrative. *See* integrative dualism
  - non-Cartesian. *See* non-Cartesian substance dualism
  - organicism. *See* organicism
  - platonic. *See* platonic dualism
- substantial union, 170
- substratum-attribute theory, 18, 370, 372, 373
- supervenience, 7, 109, 176, 177, 192, 193, 269, 274, 332, 378
- survivalism, 98, 99, 100
- SUT (Subject Unity Thesis), 191, 192
- Swinburne, Richard, 4, 7, 9, 13, 14, 28, 57, 196–222, 415, 453, 455, 460, 464, 472, 474, 479
- systems theory, 322–325
- Talbott, Stephen, 118, 119
- Tallis, Raymond, 51
- teleology, 106, 114–118, 127, 398, 399, 408
- theism, 4, 10, 11, 12, 58
- third-person perspective, 24, 27, 271
- Thiselton, Anthony, 429
- thisness, 9, 13, 14, 145, 148–150, 372
- Thomism, 13, 386
- Thomistic dualism, 11, 13, 130
- thought, 3, 62, 63, 65, 67, 68, 74, 94, 95, 99, 103, 137, 176, 199, 209, 216, 298, 303, 395, 401, 408
  - I-thoughts, 193, 194, 198
  - rational, 3, 181, 298, 355
- thought experiment, 14, 115, 149, 194, 196, 197, 221, 353–355, 397
- token physicalism, 5, 6, 23, 181, 273
- Toner, Patrick, 98
- too many thinkers argument, 356–359. *See also* problem of the many
- total phenomenal unity, 108, 186, 188
- transformative experience. *See* experience, transformative
- transitivity argument, the, 273
- transparency, 210
  - argument from, 155
- traveling-forms model, 389
- trichotomism, 421, 427
- Tye, Michael, 4, 292
- type physicalism, 5, 6, 164
- Unger, Peter, 57, 78
- union argument, 396, 397, 399, 405
- unique and exclusive attributes, doctrine of, 169, 170
- unity,
  - functional, 44, 45, 76, 82, 194, 415
  - metaphysical, 76, 82
  - of consciousness. *See* unity of consciousness
  - phenomenal. *See* phenomenal unity
- unity of consciousness, 4, 10, 12, 14, 73, 82, 107–110, 173, 184–192, 200, 202, 204, 208, 212, 216, 217, 219, 220, 222, 223, 397
  - argument from the, 69, 70, 73, 82, 127, 134, 209, 212, 219–221
- vagueness, 137, 398
  - of brains, 79, 397, 404
  - of identity, 175
- Vallicella, William, 52
- van Gulick, Robert, 322, 324
- van Inwagen, Peter, 4, 8, 20, 287, 312, 318, 358, 386, 387, 390, 394, 440, 443, 444, 447, 460, 471, 474–479, 481, 432, 488, 493
- virtual phenomenalism. *See* phenomenalism, virtual
- vitalism, 13, 105, 117, 129
- volition, 74
- waking brain hypothesis, 241, 243, 244
- weak supplementation principle, 99, 100, 386
- Wilkes, Kathleen, 27
- will, 14, 75, 83, 92, 103, 146, 152, 172, 173, 321, 390, 462
  - free. *See* free will
- Wittgenstein, Ludwig, 24, 499
- worldview, 16, 184, 185, 188, 204, 205, 243, 244, 322, 323, 394, 395, 414, 415
- Wright, N. T., 19, 413–424, 427, 471
- Zimmerman, Dean, 20, 30, 36, 57, 73, 78, 79, 352, 356, 397, 399, 475, 479, 480
- zombies, 7
  - and constitutionalism, 18, 355
- Zoroastrianism, 44