ROUTLEDGE HANDBOOKS



The Routledge Handbook of Modality

Edited by Otávio Bueno and Scott A. Shalkowski

THE ROUTLEDGE HANDBOOK OF MODALITY

Modality – the question of what is possible and what is necessary – is a fundamental area of philosophy and philosophical research. *The Routledge Handbook of Modality* is an outstanding reference source to the key topics, problems and debates in this exciting subject and is the first collection of its kind. Comprising thirty-five chapters by a team of international contributors, the Handbook is divided into seven clear parts:

- worlds and modality
- essentialism, ontological dependence and modality
- modal anti-realism
- epistemology of modality
- modality in science
- modality in logic and mathematics
- modality in the history of philosophy

Within these sections the central issues, debates and problems are examined, including possible worlds, essentialism, counterfactuals, ontological dependence, modal fictionalism, deflationism, the integration challenge, conceivability, *a priori* knowledge, laws of nature, natural kinds and logical necessity.

The Routledge Handbook of Modality is essential reading for students and researchers in epistemology, metaphysics and philosophy of language. It will also be very useful for those in related fields in philosophy such as philosophy of mathematics, logic and philosophy of science.

Otávio Bueno is Professor of Philosophy and Cooper Senior Scholar in Arts and Sciences at the University of Miami, USA. He is co-author of *Applying Mathematics: Immersion, Inference, Interpretation* (2018), and Editor-in-Chief of *Synthese* and of the *Synthese Library* book series.

Scott A. Shalkowski is Senior Lecturer in Philosophy at the University of Leeds, UK. With Otávio Bueno, he is an author of the forthcoming Routledge book *Epistemology of Modality*.

ROUTLEDGE HANDBOOKS IN PHILOSOPHY

Routledge Handbooks in Philosophy are state-of-the-art surveys of emerging, newly refreshed, and important fields in philosophy, providing accessible yet thorough assessments of key problems, themes, thinkers, and recent developments in research.

All chapters for each volume are specially commissioned and written by leading scholars in the field. Carefully edited and organized, *Routledge Handbooks in Philosophy* provide indispensable reference tools for students and researchers seeking a comprehensive overview of new and exciting topics in philosophy. They are also valuable teaching resources as accompaniments to textbooks, anthologies, and research-orientated publications.

Also available:

THE ROUTLEDGE HANDBOOK OF FEMINIST PHILOSOPHY OF SCIENCE
Edited by Sharon Crasnow and Kristen Intemann

THE ROUTLEDGE HANDBOOK OF LINGUISTIC REFERENCE
Edited by Stephen Biggs and Heimir Geirsson

THE ROUTLEDGE HANDBOOK OF DEHUMANIZATION
Edited by Maria Kronfeldner

THE ROUTLEDGE HANDBOOK OF ANARCHY AND ANARCHIST THOUGHT
Edited by Gary Chartier and Chad Van Schoelandt

THE ROUTLEDGE HANDBOOK OF THE PHILOSOPHY OF ENGINEERING
Edited by Diane P. Michelfelder and Neelke Doorn

THE ROUTLEDGE HANDBOOK OF MODALITY
Edited by Otávio Bueno and Scott A. Shalkowski

THE ROUTLEDGE HANDBOOK OF PRACTICAL REASON
Edited by Kurt Sylvan and Ruth Chang

For more information about this series, please visit: https://www.routledge.com/Routledge-Handbooks-in-Philosophy/book-series/RHP

THE ROUTLEDGE HANDBOOK OF MODALITY

Edited By Otávio Bueno and Scott A. Shalkowski

Routledge
Taylor & Francis Group
LONDON AND NEW YORK

First published 2021
by Routledge
2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

and by Routledge
52 Vanderbilt Avenue, New York, NY 10017

Routledge is an imprint of the Taylor & Francis Group, an informa business

© 2021 selection and editorial matter, Otávio Bueno and Scott A. Shalkowski; individual chapters, the contributors

The right of Otávio Bueno and Scott A. Shalkowski to be identified as the authors of the editorial material, and of the authors for their individual chapters, has been asserted in accordance with sections 77 and 78 of the Copyright, Designs and Patents Act 1988.

All rights reserved. No part of this book may be reprinted or reproduced or utilised in any form or by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying and recording, or in any information storage or retrieval system, without permission in writing from the publishers.

Trademark notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Cataloging-in-Publication Data
Names: Bueno, Otávio, editor. | Shalkowski, Scott A., editor.
Title: The Routledge handbook of modality / edited by OtávioBueno and Scott A. Shalkowski.
Description: Abingdon, Oxon ; New York, NY: Routledge, 2021. | Includes bibliographical references and index.
Identifiers: LCCN 2020037342 (print) | LCCN 2020037343 (ebook) |
Subjects: LCSH: Modality (Logic) | Modality (Theory of knowledge)
Classification: LCC BC199.M6 R68 2021 (print) | LCC BC199.M6 (ebook)DDC 160--dc23 LC record available at https://lccn.loc.gov/2020037342
LC ebook record available at https://lccn.loc.gov/2020037343

ISBN: 978-1-138-82331-0 (hbk)
ISBN: 978-1-315-74214-4 (ebk)

Typeset in Bembo
SPi Technologies India Private Limited

Dedicated to the memory of Bob Hale


Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

CONTENTS

Notes on Contributors xi

Introduction: Modal matters: philosophical significance 1
Otávio Bueno and Scott A. Shalkowski

PART 1
Worlds and modality

1 Possible worlds 11
Michael De

2 Actualism 21
Karen Bennett

3 Counterfactual conditionals 30
Dorothy Edgington

4 Impossibility and impossible worlds 40
Daniel Nolan

5 The origins of logical space 49
Brian Leftow

PART 2
Essentialism, ontological dependence, and modality

6 Essentialism and modality 61
Penelope Mackie

7 *De re* modality 70
Boris Kment

8 Relativized metaphysical modality: index and context 82
Benj Hellie, Adam Russell Murray, and Jessica M. Wilson

9 Ontological dependence, Grounding and Modality 100
Fabrice Correia

10 Modalism 114
Scott A. Shalkowski

PART 3
Modal anti-realism

11 Modal anti-realism 125
John Divers

12 Modal conventionalism 136
Ross P. Cameron

13 Norms and modality 146
Amie L. Thomasson

PART 4
Epistemology of modality

14 The integration challenge 157
Sonia Roca-Royes

15 The epistemic idleness of conceivability 167
M. Oreste Fiocco

16 Epistemology, the constitutive, and the principle-based account of modality 180
Christopher Peacocke

17 The counterfactual-based approach to modal epistemology 188
Timothy Williamson

18 Modality and a priori knowledge 198
Albert Casullo

19 Intuition and modality: a disjunctive-social account of intuition-based justification for the epistemology of modality 208
Anand Jayprakash Vaidya

PART 5
Modality and the metaphysics of science

20 Modality and scientific structuralism 221
Steven French

21 Laws of nature, natural necessity, and counterfactual conditionals 230
Marc Lange

22 Natural kinds and modality 239
Alexander Bird

23 Modality in physics 251
Samuel C. Fletcher

24 Physical and metaphysical modality 265
Ned Hall

PART 6
Modality in logic and mathematics

25 Modality in mathematics 281
Øystein Linnebo and Stewart Shapiro

26 Modal set theory 292
Christopher Menzel

27 The logic of metaphysical modality 308
Bob Hale

28 Modality and the plurality of logics 319
Otávio Bueno

PART 7
Modality in the history of philosophy

29 Ancient Greek modal logic 331
Robin Smith

30 Modality in medieval philosophy 344
Stephen Read

31 Modality in Descartes's philosophy 355
Alan Nelson

32 Hume on modality 364
Peter Millican

33 Kant on real possibility 378
Nicholas Stang

34 Quine on modality 390
Roberta Ballarin

35 Kripke on modality 400
John P. Burgess

Index 409

NOTES ON CONTRIBUTORS

Roberta Ballarin is Associate Professor of Philosophy at the University of British Columbia. She is the author of various articles on the philosophy and the logic of the modalities, including the *Stanford Encyclopedia of Philosophy* entry on the Modern Origins of Modal Logic.

Karen Bennett is Professor of Philosophy at Rutgers University, though she wrote this piece while at the Sage School of Philosophy at Cornell University. For more information on her work, see www.karenbennett.org.

Alexander Bird is the Bertrand Russell Professor of Philosophy at the University of Cambridge, having previously held the Peter Sowerby Professorship of Philosophy and Medicine at King's College London. His work is primarily in the metaphysics and epistemology of science and medicine, with a particular focus on natural kinds.

Otávio Bueno is Professor of Philosophy and Cooper Senior Scholar in Arts and Sciences at the University of Miami. His research concentrates on the philosophies of science, mathematics, and logic, as well as on epistemology, metaphysics, and philosophy of art. He is the author of numerous papers and several books, including *Applying Mathematics: Immersion, Inference, Interpretation* (with Steven French; Oxford University Press, 2018). He is editor in chief of *Synthese* and of the Synthese Library book series.

John P. Burgess is John N. Woodhull Professor of Philosophy at Princeton University, where he has taught for forty years. He has published eight books and numerous articles in logic, philosophy of mathematics and language, and history of analytic philosophy.

Ross P. Cameron is Professor of Philosophy at the University of Virginia. His recent book *The Moving Spotlight: An Essay on Time and Ontology* (Oxford University Press) is available now from all, or at least some, good bookstores.

Albert Casullo is Professor of Philosophy at the University of Nebraska-Lincoln. He is the author of *A Priori Justification* (Oxford University Press, 2003), *Essays on A Priori Knowledge and*

Justification (Oxford University Press, 2012), and the co-editor of *The A Priori in Philosophy* (Oxford University Press, 2013).

Fabrice Correia is Professor of Analytic Philosophy at the University of Geneva, Switzerland. He is the author of *Existential Dependence and Cognate Notions* (Munich: Philosophia, 2005) a co-author (with Benjamin Schnieder) of *Grounding: Understanding the Structure of Reality* (New York: Cambridge University Press, 2012), and has published extensively on grounding, ontological dependence and modality.

Michael De completed his PhD in 2011 at the University of St Andrews. His main research interests lie in metaphysics, logic, and epistemology. He is particularly interested in systematic metaphysics, the metaphysics and semantics of negation, and analyses of knowledge.

John Divers is Professor of Philosophy at the University of Leeds. His published work on modality includes *Possible Worlds* (Routledge, 2002) and some thirty journal articles. He recently served as President of the Mind Association, Director of the Leverhulme Trust *Thinking Counterfactually* research project and remains an editor of *Thought: A Journal of Philosophy*. He is currently finalizing a monograph, *Necessity after Quine*, for Oxford University Press.

Dorothy Edgington studied at Oxford. Most of her career was spent at Birkbeck College, University of London, interrupted by five years (1996–2001) and then three years (2003–6) as a professor at Oxford, the latter period holding the Waynflete Chair of Metaphysics. She is now an emeritus professor of both institutions. She has worked mainly on conditionals and vagueness, making use of the notion of probability in both these areas.

M. Oreste Fiocco is Associate Professor in the Department of Philosophy at the University of California, Irvine. His research interests include metaphysics, epistemology, and the philosophy of mind. His work has been published in a number of journals and collections.

Samuel C. Fletcher is Assistant Professor in the Department of Philosophy at the University of Minnesota, Twin Cities, Resident Fellow of the Minnesota Center for Philosophy of Science, and an external member of the Munich Center for Mathematical Philosophy, Ludwig-Maximilians-Universität.

Steven French is Professor of the Philosophy of Science at the University of Leeds. His research interests include the metaphysics of science, science and aesthetics, and the nature of theories. His most recent publication is *There Are No Such Things as Theories* (Oxford University Press, 2020). He is Co-Editor-in-Chief of *The British Journal for the Philosophy of Science* and also Editor-in-Chief of the Palgrave-Macmillan monograph series *New Directions in Philosophy of Science*.

Bob Hale (1945–2017) was Emeritus Professor at the University of Sheffield. His main research interests were in the foundations of mathematics and in philosophy of logic and language. He published many papers and the books *Abstract Objects* (Blackwell, 1987), *The Reason's Proper Study* (with Crispin Wright; Oxford University Press, 2001), and *Necessary Beings: An Essay on Ontology, Modality, and the Relations between Them* (Oxford University Press, 2013). He was elected to the Royal Society of Edinburgh in 2000.

Ned Hall is the Norman E. Vuilleumier Professor of Philosophy and Chair of the Philosophy Department at Harvard University. His research focuses on topics in metaphysics and epistemology that overlap with philosophy of science, such as fundamental laws of nature, causal structure of the world, objective probability, and metaphysical and physical possibility. He co-authored *Causation: A User's Guide* (with Laurie Paul, Oxford University Press, 2013) and co-edited *Causation and Counterfactuals* (with John Collins and Laurie Paul, MIT Press, 2004).

Benj Hellie is Professor of Philosophy at the University of Toronto. His research (philosophies of mind, language, and logic; history of analytic, especially David Lewis) manifests in collections and journals, including *Mind* and *Philosophical Review*, and in talks on five continents. A book manuscript (*Out of This World: Logical Mentalism and the Philosophy of Mind*, contracted to Oxford University Press) replaces truth in logic with endorsement, yielding "expressivist" treatments of self- and other-knowledge, psychophysical explanation, irrationality, and the mind-body problem.

Boris Kment received a BPhil from Oxford and a PhD from Princeton. He has taught at the University of Michigan and is currently Associate Professor at Princeton. His main interests lie in metaphysics and epistemology, and he is the author of numerous papers and a book, *Modality and Explanatory Reasoning* (Oxford University Press, 2014).

Marc Lange is Theda Perdue Distinguished Professor and Chair of the Philosophy Department at the University of North Carolina at Chapel Hill. He works on the philosophy of science and related regions of metaphysics and epistemology.

Brian Leftow is William P. Alston Professor of the Philosophy of Religion, Department of Philosophy, Rutgers University. From 2002 to 2018 he was Nolloth Professor of the Philosophy of the Christian Religion, Oxford University. He is the author of *God and Necessity* (Oxford, 2012), *Time and Eternity* (Cornell, 1991), and over 100 papers in philosophy of religion, metaphysics, and medieval philosophy.

Øystein Linnebo is Professor of Philosophy at the University of Oslo, having previously held positions at Bristol, London, and Oxford. His main research interests lie in the philosophies of logic and mathematics, metaphysics, and early analytic philosophy (especially Frege). He has published more than 60 scientific articles and is the author of three books, most recently *Thin Objects: An Abstractionist Account* (2018) and *The Many and the One* (with Salvatore Florio, forthcoming), both with Oxford University Press.

Penelope Mackie is Associate Professor and Reader in Philosophy at the University of Nottingham. Her principal research interests are in topics in metaphysics, including modality, identity and persistence, counterfactuals, and free will. Her book, *How Things Might Have Been: Individuals, Kinds, and Essential Properties*, was published by Oxford University Press in 2006.

Christopher Menzel is Professor of Philosophy at Texas A&M University. His research interests are rather narrowly focused on the metaphysics and logic of modality, the metaphysical commitments of logic, and the metaphysics of mathematics. He also occasionally explores the connections between these issues and theism. He serves two cats and is an avid road cyclist.

Peter Millican is Gilbert Ryle Fellow and Professor of Philosophy at Hertford College, Oxford. He has edited several books and published over forty papers on early modern philosophy, with particular focus on Hume's philosophical development and his discussions of induction, causation, free will, miracles, and skepticism. Other interests include philosophy of religion, philosophy of language, epistemology, and digital humanities (including the web resource www.davidhume.org). In 2012, he founded the joint degree program in computer science and philosophy at Oxford University.

Adam Russell Murray is Assistant Professor of Philosophy at the University of Manitoba. He works primarily in metaphysics and the philosophy of language and has published papers in *Oxford Studies in Metaphysics*, *The Canadian Journal of Philosophy*, and the *Routledge Handbook of Propositions*.

Alan Nelson is Professor of Philosophy at the University of North Carolina at Chapel Hill. He has also taught at UCLA; University of California, Irvine; Pittsburgh; Stanford; and USC. He has published widely on Descartes and other early modern philosophers.

Daniel Nolan is McMahon-Hank Professor of Philosophy at the University of Notre Dame. He is the author of *Topics in the Philosophy of Possible Worlds* (Routledge 2002) and *David Lewis* (Acumen, 2005), as well as papers on metaphysics, philosophical logic, philosophy of science, and a range of other subjects. He is Fellow of the Australian Academy of the Humanities.

Christopher Peacocke is Johnsonian Professor of Philosophy at Columbia University and Fellow of the Institute of Philosophy, London University. He is the author of *A Study of Concepts*, *Being Known*, and *The Mirror of the World*. He is currently writing a book on metaphysics and the theory of intentional content.

Stephen Read is Professor Emeritus of the History and Philosophy of Logic at the University of St Andrews, Scotland. Recent publications include the *Cambridge Companion to Medieval Logic* (edited with Catarina Dutilh Novaes) and an English translation of John Buridan's *Treatise on Consequences*. He is currently engaged on a Leverhulme-funded project on "Theories of Paradox in Fourteenth-Century Logic" to produce editions and English translations of the treatises on insolubles by Paul of Venice, John Dumbleton, and Walter Segrave.

Sonia Roca-Royes obtained her PhD from the University of Barcelona in 2007 and that year joined the University of Stirling, where she currently is Senior Lecturer. Her area of specialization is metaphysics and epistemology of modality, with broader interests in ontology, philosophy of mathematics and logic, formal logic, and philosophy of language.

Scott A. Shalkowski is Senior Lecturer in Philosophy at the University of Leeds. His principal research interests are metaphysics, the philosophy of religion, and the philosophies of logic and mathematics. His (and co-authored) articles have appeared in *Journal of Philosophy*, *Philosophical Review*, *Noûs*, *Mind*, *Philosophy and Phenomenological Research*, *Philosophical Studies, American Philosophical Quarterly*, and *Faith and Philosophy.* He is currently engaged (with Otávio Bueno) in a project on modal epistemology for Routledge.

Stewart Shapiro is currently the O'Donnell Professor of Philosophy at The Ohio State University, serves as a visiting professor at the University of Connecticut, and is Presidential

Fellow at The Hebrew University of Jerusalem. He specializes in philosophy of mathematics, philosophy of language, logic, and philosophy of logic, recently developing an interest in semantics. Professor Shapiro has taught courses in logic, philosophy of mathematics, philosophy of science, philosophy of religion, Marxism, aesthetics, Jewish philosophy, and medical ethics.

Robin Smith is Professor of Philosophy Emeritus at Texas A&M University. His publications include translations with commentary of Aristotle's *Prior Analytics* and *Topics I and VIII*. He is currently working on a translation with commentary of Aristotle's *On Sophistical Refutations*.

Nicholas Stang is Associate Professor of Philosophy at the University of Toronto. He works at the intersection of classical German philosophy (especially Kant and Hegel) and contemporary analytic philosophy (especially metaphysics). His first book, *Kant's Modal Metaphysics*, was published by Oxford University Press in 2016. He is currently writing a book on what the Kantian (and post-Kantian) critique of metaphysics has to do with contemporary metaphysics, tentatively titled *How is Metaphysics Possible? A Critique of Analytic Reason*.

Amie L. Thomasson is the Daniel P. Stone Professor of Intellectual and Moral Philosophy at Dartmouth College. She is the author of more than 70 articles and of four books: *Norms and Necessity* (Oxford University Press, 2020), *Ontology Made Easy* (Oxford University Press, 2015—winner of the Sanders Book Prize), *Ordinary Objects* (Oxford University Press, 2007), and *Fiction and Metaphysics* (Cambridge University Press, 1999). She also co-edited (with David W. Smith) *Phenomenology and Philosophy of Mind* (Oxford University Press, 2005).

Anand Jayprakash Vaidya is Professor of Philosophy at San Jose State University. His research focuses on the cross-cultural and multi-disciplinary study of mind, knowledge and reality. For over 20 years he has done research on the epistemology of modality, essence, and grounding.

Timothy Williamson is the Wykeham Professor of Logic at Oxford University. His books include *Identity and Discrimination* (1990), *Vagueness* (1994), *Knowledge and Its Limits* (2000), *The Philosophy of Philosophy* (2007), *Modal Logic as Metaphysics* (2013), and *Tetralogue* (2015).

Jessica M. Wilson is Professor of Philosophy at the University of Toronto. Her research focuses on metaphysics, philosophical methodology, and epistemology, with applications to philosophy of mind and science. In 2014 she was a co-recipient of the Lebowitz Prize for Philosophical Achievement and Contribution, and in 2017 she was awarded a five-year Social Sciences and Humanities Research Grant for her project, "How Metaphysical Dependence Works". Her book, *Metaphysical Emergence*, is forthcoming with Oxford University Press.


Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Introduction
MODAL MATTERS
philosophical significance

Otávio Bueno and Scott A. Shalkowski

Life is concerned not only with what is and how it is, but also with how things could or could not have been otherwise. Some of our most practical deliberations about how to act, we think, account for more than how things *are*. The most obvious reason for doing this is that, when we eventually act, things may be different. We are also aware that how things will be is affected by how we, in fact, act. At the very least, mature individuals have the appearance of being able to choose amongst different courses of action, at least some of which sometimes will have different outcomes. Looking backward instead of forward, as it were, we seem also to be aware that had things gone differently, at least some things would now be different. Had we or others behaved differently, the world would be different from how it now is. Thus, we are aware, so we think, of a great deal of contingency.

Our thinking about contingency is done against a background of necessities, things that we cannot change. Millennia of thinking carefully about such matters has yielded many distinctions. Past realities are necessary in one respect, but not in all. They are necessary because they are now a fixed part of reality. No one can change the past. It is what it is. Yet, the past, in another respect, is not necessary. Caesar might not have crossed the Rubicon, and his life might have ended rather differently.

Philosophers have introduced distinctions to help us navigate these matters. The backward-looking temporal necessity of Caesar's crossing the Rubicon is one thing. Caesar's ability to have chosen a different military and political path is another. Forward-looking deliberation might also need to account for some cold, hard physical realities, though. Choose however you like, but given current understanding of the world, you will never travel faster than 299,792,458 meters per second. Yet, the speed of light is also something not amenable to discovery by conceptual or linguistic analysis; one must actually go look to determine its value. That observational requirement lends credence to the judgment that it and other central physical facts are contingent. No background logic or mathematics makes 299,792,458 the unavoidable value of light's speed. No amount of reflection alone renders it inevitable. Thus, it seems that there is some sense in which it is possible that light behaves differently.

At the same time, physical facts like the speed of light also constitute real limits on other physical facts, amongst them the character and consequences of our actions. Thus, there is at minimum the appearance that not only is necessity to be reckoned with, but there are different

kinds of necessity, permitting something to be both necessary and contingent, exhibiting one kind of necessity but not some other(s).

If history and science are concerned with contingencies—even as some scientists claim to have uncovered some (physical, chemical, biological) necessities—then philosophy, at least on a given understanding favored by some philosophers within its recent analytic incarnation, appears to be concerned with unrestricted, genuine necessity—necessity *tout court*. Platonists affirm the existence of abstract *ante rem* universals. They do so on the basis of what they take to be required for understanding the possession of properties or the nature of resemblance, wherein two or more things "share" a property. Nothing in those reasons makes sense of contingency in the matter. If to be human is to bear some relation to a universal—HUMAN—then there is no prospect that it is merely a happy accident that all of us bear this relation to that universal. It is the nature of the beast to be so related. Those rejecting the platonist case do so either because they think that it is possible that such *ante rem* abstract objects do not exist, or for the more modest reason that there is no necessary connection between having characteristics or resembling other things and the existence of universals, each rejection based on something modal. Even if not all problems in metaphysics are problems infused with necessity, many certainly are. Where necessity is not obvious, it lurks beneath. One might maintain that all earthly minds are somehow physical in nature without the commitment that mentality itself is, of its nature, physical, yet an account of the nature of mentality, whether earthly or not, will yield some necessities—if nothing else relative to other states of affairs: given the physical constitution of bacteria, it is just not possible for them to have mental states.

It is no surprise, then, that philosophers turned their attention to modality, originally conceived as modes of being or modes of truth. It is also no surprise that philosophers have proposed rather different answers to whether there is necessity, whether there is necessity to be discovered, how necessity is to be characterized, whether there are more and less restricted kinds of necessity, whether necessity is a fundamental part of reality's architecture, how we might come to know the scope of necessity, and the like.

What is somewhat surprising is that there has been no handbook guiding interested parties through some of the deep and sometimes subtle and difficult philosophical waters of the philosophy of modality. We are delighted to present here some essays from first-rate scholars on some central themes regarding the necessary and the possible. The contributions to this volume cover four major areas of philosophical concern regarding modality: metaphysics, epistemology, applications (in the context of the sciences, logic, and mathematics), and contributions from historically significant authors. Articles in the first three parts of the current volume treat what, until recently, had been the main focus of attention: the nature of modality and other related phenomena, such as the existence of possible worlds, the semantics for counterfactual conditionals, whether individuals have essential properties, whether there is non-causal ontological dependence, and the overall interpretation of modal discourse, whether realistically (broadly construed) or not. As important as these issues are, how we can know any of these metaphysical claims is equally important. These issues are addressed by contributions to the fourth part of this handbook. The fifth and sixth parts address the applications of modality in science as well as in logic and mathematics. The seventh and final section contains contributions providing some historical background, from ancient and medieval sources through to some major modern and contemporary contributions to the subject of modality.

Part 1 is devoted to the topics that received a great deal of attention in the second half of the twentieth century and informed work in the early twenty-first: worlds and modality. There is, of course, a very intuitive but not philosophically rigorous way of mixing our thinking about modality with our thinking about worlds. Intuitions may be pumped by asking whether one can

imagine a scenario in which thus-and-so occurs. A world is just an extended, complete scenario. Possibilities are just alternatives, but alternatives do not exist in isolation. They exist within a totality—a world. Those coming from the study of formal logics are familiar with treating valid inferences as those for which all models of the premises are models of the conclusion. Similar treatments of the logic of modal operators, '$\Box$' or '$\Diamond$', expanded the domain of models so that not only were alternative assignments to sentential, predicate, and individual variable from within the existing stock of sentences, predicates, and individuals permitted, but alternative "worlds" were envisioned, each with its own ontology, to model valid inference involving the formal expressions for necessity and possibility. For a useful presentation of this "possible worlds semantics", see Copeland (1996). David Lewis was instrumental in bringing a serious, realist, treatment of possible worlds to bear on many philosophical problems, including the philosophy of modality. His influential *On the Plurality of Worlds* (Lewis 1986) provided the primary version of this approach.

Michael De begins this opening section with an article focused directly on possible worlds, addressing what roles philosophers have asked possible worlds to play and whether they are suitable for those roles. By examining the analyses of modality, belief, and conditionals, he concludes that worlds are not as suitable as they first appeared to be, requiring either supplanting or supplementing. Karen Bennett follows with a discussion of actualism, the view that only actuality exists, or, in her preferred formulation, "absolutely everything actually exists". No alternate realities exist, which is not to say that possible worlds do not exist. Those thinking that worlds are maximal sets of propositions, or of states of affairs, or complex properties maintain that possible worlds exist, though not all things mentioned in propositions exist, not all states of affairs obtain, and not all properties are exemplified. Actualism stands in contrast to possibilism, the view that merely possible objects do, really, exist, even if not within our own possible world. Though many find actualism natural, it is difficult to make actualism perfectly clear, as she demonstrates by discussing some of its challenges. Dorothy Edgington provides the reader with an introduction to and overview of the nuances required for understanding counterfactual conditionals, those that tend to be about unrealized possibilities. The most common treatments, due to David Lewis and Robert Stalnaker, respectively, have availed themselves of a framework of possible worlds to account for the conditions under which these conditionals are true or false. These conditionals concern possibilities, but only some within a region of logical space where things mostly resemble how things actually are. Our attention to the array of possibilities must be restricted when thinking about counterfactuals. Let the relevant unrestricted modality be metaphysical modality. Daniel Nolan shows how some of the same problems—such as giving accounts of the contents of belief—and some of the same methodological considerations—assessing theories according to their value—lead to taking impossible worlds no less seriously than possible worlds. In the end, the scope of modal realism may turn out to be even broader than initially anticipated. Brian Leftow closes this part of the handbook with a challenge to a nearly universal assumption, i.e., that so-called logical space has no origin. It just is what it is. Continuing with attention to philosophical method and to common Ockhamist tendencies amongst philosophers, Leftow proposes that we rethink the standard no-origins thesis about logical space.

Part 2 moves attention away from a direct focus on worlds to the relations, if any, between modality and issues of whether objects have essential properties—those without which they could not exist—and whether there are relations of ontological dependence amongst objects or phenomena. If there is non-causal, ontological dependence, then some things will be prior to or metaphysically more significant than others, even if all exist necessarily. Penelope Mackie begins this portion of the volume with an article on essentialism and modality, noting that some use

modality to articulate essentialism, while others not, and some even reverse this order of treatment. Boris Kment follows this with a complementary piece on *de re* modality, where his discussion revisits concerns, some made prominent by W. V. Quine, about the legitimacy of any attributions of necessity to individuals and whether quantifying into modal contexts is philosophically well behaved. By way of discussing whether qualitative identity fixes numerical identity, Kment ends by discussing modal contingentism, whether claims of the form $\Box P$ and $\Diamond P$ are themselves necessary. This provides a route into the issues that concern Benj Hellie, Adam Murray, and Jessica Wilson in the essay that follows. They give a formal framework for what can be thought of as a classical treatment of metaphysical modality. After introducing the reader to puzzles that arise on this treatment, they suggest that philosophical sophistication requires relativized metaphysical modality, demanding a significantly more complex semantics to account for plausible judgments about how it is that some things are not possible, but could be. As simple uses of possible worlds may not suffice for all philosophical purposes, so possible worlds themselves may be insufficient for some purposes to which they have been put, without pragmatic considerations. Fabrice Correia surveys relatively recent developments in thinking about metaphysical matters, specifically developments regarding grounding and ontological dependence. Metaphysicians sometimes think that one phenomenon explains another, when that explanatory relation is not causal and may even involve matters that are themselves necessary. Grounding and ontological dependence are to be the relevant explanatory metaphysical relations. The most natural ways of understanding these relations is in terms of modality. Scott A. Shalkowski closes this part of the book with a focus on whether modality is reducible to non-modal items. Since Mackie's essay has covered the challenge from essentialists, such as Kit Fine, E. J. Lowe, and Bob Hale, Shalkowski focuses attention on other reductive programs intended to undermine the thesis that modality is primitive, metaphysically ineliminable. If modalism is correct, then possible worlds cannot be used in service of ontologically serious claims, even if their heuristic and modeling virtues remain intact.

Up to this point, the literature surveyed and critiqued has assumed that modality is something real, something objective, something discoverable. Part 3 contains three contributions questioning this assumption. John Divers helpfully begins with a taxonomy of different kinds of realism about modality—ideological and ontological—and correlated kinds of anti-realism. As one may well have come to expect, the issues are often subtle, and navigating them requires great care. Ross Cameron follows with an entry on modal conventionalism. As Divers distinguished different forms of (anti-)realism about modality, so Cameron guides us through conventionalisms of varying degrees of sophistication. Amie Thomasson closes this part of the book by noting ontological and epistemological problems for modal claims, when interpreted realistically. She argues that typical modally qualified metaphysical claims are not really descriptive claims, but are instead normative. Their function is to convey semantic rules and their consequences, thus avoiding difficulties in the epistemology of modality.

That knowledge of necessity and possibility should be accounted for by those thinking that reality is modally informed has led to the development of various approaches to the epistemology of modality. Part 4 contains essays on some of the most significant attempts to address these matters, without taking anti-realist options. Sonia Roca-Royes kicks off this part of the handbook by bringing to our attention the integration challenge. Intuitively, the challenge is quite general: claims we are inclined to make should be those to which we can have epistemic entitlement. Though the challenge is not limited to the philosophy of modality, Roca-Royes moves past the intuitive formulation of the challenge to note some of the varieties of issues that arise and the options one has to meet the challenge. One of the most natural ways to engage claims of possibility and necessity is to think, to cogitate, to engage in trying to conceive of how things

could be, even if they are not, and of how things go beyond our capacity to conceive of things. M. Oreste Fiocco's chapter discusses the prospect of this armchair route into modal knowledge, concluding that ultimately conceivability is idle, epistemically speaking, and thus not suitable for modal epistemology. Having prompted the current interest in the integration challenge, Christopher Peacocke proposed a principle-based account of metaphysical necessity. Since he introduced this account in 1999, it has been subjected to challenges. In his contribution to this volume, Peacocke addresses those challenges in a defense of his rationalist approach to the epistemology of modality generated from a theory of modal understanding. In contrast, Timothy Williamson has argued that there is a close link between the metaphysical modality and counterfactual conditionals. In his essay here, Williamson defends his account of our knowledge of modality by way of our knowledge of counterfactuals through the development of counterfactual suppositions. Both Peacocke's and Williamson's views deviate in their respective ways from what had been, at least until the 1970s, a rather standard, even if not wholly uncontroversial, claim that knowledge of necessity is a priori—it is not tied, in one way or another, to experience. Albert Casullo's contribution takes readers through important nuances and complexities regarding the nature of a priori knowledge itself, as well as its alleged involvement in knowledge of necessity. Finally, intuition is often used in metaphysical discussions to label some basis for judgment. Anand Vaidya closes this part of the volume by focusing on intuition. After considering the kind of mental state intuition is and whether the mental faculty from which intuition emerges is reliable, he advances an intuition-based account that recognizes the role played by a social dimension in knowledge of modality.

In the foregoing chapters, some attention is given to the relatively common view that if there is modality at all, there are different kinds. Perhaps there is a basic, most general kind and others are restricted by holding some things fixed. If the most fundamental, absolute, necessity can be axiomatized, think of a restricted necessity as that articulated with additional axioms. Different modalities structured in this way permit us to make sense of something being both necessary and contingent while saving ourselves from contradiction. We are unable to violate laws of nature, if such there be. Yet, these laws might themselves be contingent. So, there may be some actions that are both possible and impossible for me. In one, say, metaphysical sense, we can violate laws of nature because those laws are themselves contingent, yet at the same time there is a sense that we cannot violate those laws, since we live in a world governed by those laws.

Part 5 takes us away from metaphysics broadly and focuses attention on the role that modality plays in the articulation and understanding of science. Laws of nature are the most common locus of necessity. Steven French begins this part of the handbook with some direct examination of the status of laws in the context of modern physics. Three major views are discussed: Humean eliminativism of the modal status of laws, its reduction by way of dispositions, and its treatment as primitive, as is articulated in ontic structural realism. When distinguishing laws of nature (i.e., nomic regularities) from accidental regularities, we often advert to counterfactual conditionals, those expressing that had reality been different in one respect, it would have also been different in some other respect. Marc Lange sketches an account of the interrelations of laws of nature, their attending nomic necessity, and counterfactual conditionals. If there are laws of nature that science sometimes discovers, in what terms are those laws formulated? They report nomic facts about what? One natural answer is that as well as uncovering nature's constraints—the laws—science also uncovers its natural objects that fall into natural kinds. Alexander Bird revisits the reintroduction of natural kinds into philosophical discussions via concerns about theories of reference by Saul Kripke and Hilary Putnam. This discussion was instrumental in bringing back into fashion essentialism, specifically but not exclusively, essentialism regarding natural kinds. Science, on their views, uncovers not the mere fact that water is H_2O, for example, but that water

is essentially H_2O. It could have no other structure. Sam Fletcher follows with a broad survey of how modality is thought to figure in physics, charting issues regarding not only the status of laws, but also specific issues that arise for thermodynamics and statistical mechanics, space-time theory, and quantum theory. Ned Hall closes this part of the volume by arguing for the unorthodox idea that the real struggle in the philosophy of modality is not to find room for a modality appropriate for the sciences. Rather, the struggle is to find a place for so-called metaphysical necessity. Necessity's role is primarily in explanation, and that seems to be the domain of the physical modality.

On the received view about these matters, mathematics and logic are common sources of instances of what are supposed to be necessary truths: 7 + 5 does not just happen to be 12, it must be. Similarly, assuming classical logic, there seems to be no way around Q following from $P \rightarrow Q$ and P. Part 6 is devoted to the role that modality plays in these disciplines. Øystein Linnebo and Stewart Shapiro begin by showing how modality has entered into accounts of mathematics, first in accommodating the objectivity of mathematics when platonic objects are rejected and also in making room for the potentiality of infinity. Modality is useful, though, even when an objective mathematical ontology is granted. How does that ontology behave when we account for multiple possibilities and the relations between concrete objects and the sets to which they belong? Those are issues discussed by Chris Menzel in his chapter on modal set theory. Suppose, with Linnebo, Shapiro, and Menzel, that modality is required for a proper understanding of mathematics. What exactly is the import of that modality? What is its logic? There are, of course, many different formal systematizations of the logic of modalities. Bob Hale takes us through issues bearing on the logic of metaphysical modality. As there are different claimants to metaphysical necessity, there is a plurality of non-modal logical systems. Otávio Bueno sets out the thesis not just that there is a plurality of logics, but that one should be a pluralist in one's embrace of a multiplicity of logics as in good, useful standing, arguing that modalism permits this embrace in a way that other approaches do not.

Finally, this handbook treats the reader to some important historical antecedents to the present currents in the philosophy of modality. Robin Smith introduces Part 7 by presenting the developments of the logic of modality in Ancient Greek philosophy. Stephen Read notes the continued influence of Aristotle in the medieval period as he presents how thinking about modality developed amongst influential medieval thinkers. Earlier, Brian Leftow invited us to consider the prospect of the space of possibilities having an origin. Descartes is (in-)famous for having maintained that seeming necessary truths are subject to divine will or creation. Alan Nelson's contribution examines what Descartes actually maintained about modality, divine attributes, clear and distinct ideas, and the like, to determine both what Descartes's understanding of modality was and whether Descartes held the view some have attributed to him. David Hume was a well-known critic of many aspects of Descartes's philosophy. Famously, he questioned whether we are entitled to think that causal relations involve any kind of necessity at all, and he maintained that necessities were relations of ideas rather than matters of fact. Peter Millican explores Hume's treatment of modality and whether Hume had the resources to unify causal and conceptual modalities. Kant's difficult but influential philosophy can hardly be articulated without modality. Nick Stang provides an opinionated introduction to the role of modality in Kant's critical philosophy, which until relatively recently has received little attention. This handbook concludes with entries on two very influential contemporary philosophers: W. V. Quine and Saul Kripke. Quine has the reputation of being a severe critic of modal discourse, particularly of any that seems to commit users to what he took to be disreputable Aristotelian essentialism, while Kripke is credited with rehabilitating that essentialism. Roberta Ballarin sets Quine's critiques in their historical context to clarify those critiques and to argue that, in the

end, Quine made his peace with modal discourse. John Burgess closes the handbook with attention to Kripke's work, taking readers through Kripke's route to essentialism, his epistemology of modality as well as his more technical work on modal logic and its model theory.

We are grateful to all of the contributors, who have made this handbook a valuable resource to those interested in the philosophy of modality with their significant and philosophically illuminating work. The chapters themselves, along with their accompanying references, will orient readers in shaping their own understandings of modality and its importance.

References

Copeland, J., ed. (1996) *Logic and Reality: Essays on the Legacy of Arthur Prior*. Oxford: Clarendon Press.
Lewis, D. (1986) *On the Plurality of Worlds*. Oxford: Blackwell.


Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

PART 1

Worlds and modality


Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Chapter 1
POSSIBLE WORLDS

Michael De

1.1 Introduction

Modality concerns what might or must be the case. We use modal expressions such as 'might', 'must', and cognates in a variety of ways, however. One can truthfully say, for example, that the neither yet proved nor disproved Goldbach's conjecture (that every natural number greater than two is the sum of two primes) *might* be true, even if the conjecture is false and hence, in a *metaphysical* sense, necessarily so. What one means here is that the conjecture is compatible with our evidence, since we have neither proven nor disproven it.

Modal expressions are typically *intensional* (with an 's'). An expression α is intensional just in case the substitution of extensionally equivalent expressions under the scope of α need not preserve truth. For instance, even though '8' and 'the number of planets' have the same extension, 'Necessarily, 8 is greater than 1' and 'Necessarily, the number of planets is greater than 1' do not share their extension (i.e. truth-value). However, while many modal notions are intensional, not all are. Consider e.g. the *de re* modal relation expressed by 'it is necessary of x that it be such that A', where A is to be replaced by a declarative statement. Thus, even though the expressions 'modal' and 'intensional' are sometimes used interchangeably, modality and intensionality are not equivalent.

How best are we to analyze modality and intensionality? The most popular strategy since the mid-twentieth century is to employ *possible worlds*. Reference to possible worlds dates back at least to Leibniz (see Mates, 1968), but they do not assume their familiar role until the 1940s when Rudolf Carnap (Carnap, 1946) gave an analysis of modal operators resembling a modern treatment in terms of quantification over what we would now call possible worlds. Such an analysis, often called *possible worlds semantics*, was later generalized throughout the 1940s and 1950s independently by a number of logicians, including Saul Kripke and Jaakko Hintikka, and is now the standard treatment for a wide variety of intensional notions including modality. An impressive number of intensional notions have been given possible worlds analyses, only some of which include: conditionality, causation, knowledge, *de se* belief, intrinsicality, dispositionality, aboutness or subject matter, supervenience and dependence, truthmaking, the laws of nature, essence, property, propositional and intentional content, fictional worlds, and truth in fiction. The use of possible worlds in linguistics, logic, and computer science has also seen enormous success. The fact that possible worlds talk has become common parlance in many areas of

contemporary analytic philosophy, and other fields, raises important questions concerning their ontological status and the explanatory value they afford.

Rather than answering the question *What are possible worlds?* I wish to discuss what I think is a more tractable question, namely, *What what theoretical roles are possible worlds supposed to play, and are they cut out to play those roles?*[1]

The question is tractable because we can simply look and see to what purposes possible worlds have been put and whether possible worlds analyses have survived the test of time, or whether they have been succeeded by superior analyses which either do away with worlds altogether or else demote them to a lesser role.

In giving a partial answer to the question, we will begin by looking at traditional possible worlds analyses of intensional and modal concepts (Section 1.2). We will then look at three possible worlds analyses that have played an important role in their perceived success, viz. the analyses of (i) modality and possibilities in counterpart theory (Section 1.3), (ii) belief contents (Section 1.4), and (iii) conditionals (Section 1.5).

1.2 Possible worlds semantics

Let us have a more careful look at possible world semantics and the reasons for its success. Let us denote 'Necessarily' by '$\Box$', 'Possibly' by '$\Diamond$', and 'It is not the case that' by '$\neg$'. Then, according to the simplest quantificational analysis of the broadest sort of necessity, often called *metaphysical*,

($\Box^\star$) $\Box A$ is true at a world iff A is true at every world.

Given the equivalence of $\Diamond A$ with $\neg\Box\neg A$, we have:

($\Diamond$) $\Diamond A$ is true at a world iff A is true at some world.

The quantificational analysis is simple and has the virtue of providing a way of determining whether complex modal sentences (or sentence forms), such as those containing a large number of iterated modalities (e.g. 'It is possibly necessarily possible that A only if it is necessary that A'), are true (valid) or not. Before the quantificational analysis, determining which modal inferences were valid rested mainly on potentially shaky intuitions concerning the plausibility of individual axioms or rules.[2] Possible worlds semantics provides a translation of an obscure intensional language into the pristine clarity of an extensional (meta)language.[3]

More restricted versions of necessity can be given a similar analysis by making the notion of possibility a relative matter:

($\Box$) $\Box A$ is true at a world w iff A is true at all worlds possible relative (or accessible) to w.

For doxastic modality, for instance, worlds represent possible states of belief of an agent, and one state of belief w' is possible relative to another w just in case w' cannot be ruled out by what the agent believes, as determined by w. ($\Box^\star$) is equivalent to the special case of ($\Box$) when every world is possible relative to every other, so ($\Box$) provides a semantics for a broader range of modalities. I will call ($\Box$) the *simple quantificational analysis* of modality.

What makes the simple quantificational analysis so attractive is that it provides, at a schematic level, simple and uniform semantics for a very broad class of modalities. In addition, there is a natural correspondence between properties of accessibility—the relation that holds between one world and another when one is possible relative to the other—and modal validities.

For example, if accessibility is reflexive (i.e. if every world is possible relative to itself), then $\Box A$ semantically entails A and *vice versa*. Thus, once we settle the properties accessibility has, we settle the modal validities along with them.

Despite its advantages, the simple quantificational analysis has serious drawbacks. Perhaps the most widely discussed of these is that both possibility and necessity are closed under strict implication, which means that the following holds:

(K) $\Box(A \rightarrow B) \rightarrow (\Box A \rightarrow \Box B)$.

This property is sometimes referred to as *logical omniscience*.[4] The problem is that one may know or believe a proposition without knowing everything that necessarily follows from it. Just because you know the axioms of Peano Arithmetic does not mean that you know everything that follows from them. Moreover, since necessary truths are strictly implied by everything, the simple quantificational analysis yields that every necessary truth is known. While this may hold for the most idealized form of knowledge, it fails for any interesting notion that is the object of philosophical analysis. Relatedly, what would it mean to be morally obligated to see to it that Justin Trudeau be human? And can't one have inconsistent beliefs? If so, the simple quantificational analysis falls short as an analysis of knowledge, morality, agency, belief, and other notions for which closure under strict implication is highly implausible.

However, the fact that it falls short for certain intensional notions does not mean that an analysis in terms of possible worlds is unworkable. One way to achieve such an analysis is to associate to each world a set of propositions that are necessary relative to that world. Clearly assigning to each world a set of worlds that are possible relative to it suffices to assign a family of propositions necessary there (namely, the set of propositions true at each relatively possible world), but as we have seen, this closes the family under potentially undesirable properties. Let N be a function that takes a world and yields the set of propositions necessary there; thus, $N(w)$ is the set of worlds necessary relative to w. Let $\|A\|$ denote the set of worlds at which A is true, i.e. the proposition expressed by A. Then,

(NS) '$\Box A$' is true at a world w iff $\|A\|$ is a member of $N(w)$.

We can see that necessity is no longer closed under strict implication, since it can be the case that both $\|A \rightarrow B\|$ and $\|A\|$ are in $N(w)$ while $\|B\|$ is not. Indeed, necessity is not closed under much at all until some conditions are placed on N.

The main drawback of this sort of 'analysis' is that it provides a poor explanation of when a sentence is necessary—it says it is necessary just in case the proposition it expresses is, which is true but trivial. We could, of course, provide a philosophical interpretation of N and corresponding analysis, but that will not make (NS) any less trivial. Moreover, determining which properties N has will depend on determining in advance which sentences are to be valid. So, unlike ($\Box$), the theorems are not derived from intuitive semantical properties, but instead the axioms determine what the semantical properties are.[5]

It is standard to assume that possible worlds satisfy the following two properties:

Consistency: a world w is *consistent* iff there is no collection Γ of sentences true at w such that Γ entails (in some given sense) a contradiction.

Maximality: a world w is *maximal* iff for every sentence A, either A or its negation is true at w.

It is important to keep in mind that the notion of entailment that figures in consistency need not be logical. For instance, a world that makes true both 'x has mass m kg' and 'x has mass n kg'

for distinct *m* and *n* will, given the relevant sense of entailment, entail a contradiction. At a certain level of abstraction, consistency and maximality are the only properties we need care about. In providing a semantics for certain bits of language, for example, it will not matter whether worlds are concrete or abstract entities, or whether they are built up from properties or propositions, or something else entirely. For instance, we could let a world be a class of atomic facts and say that an atomic fact is true at a world just in case that fact is a member of the world. The truth of complex facts is given as usual; a conjunctive fact is true (at a world) just in case each conjunct is, a negative fact is true just in case the negand isn't, and a universal fact is true just in case each instance is. Then, on the essential assumption that any collection of atomic facts is consistent (which one may reasonably deny), it is easy to show that a world as a set of atomic facts is maximal consistent. A variety of other constructions will similarly yield the same result.

In what follows, I illustrate, using three central examples, why worlds taken as maximal consistent objects cannot play the role they were initially assigned to play. That role includes giving an analysis, even a mere truth conditional semantics, for (i) modality, (ii) belief, and (iii) conditionals. We could list other examples that give compelling reasons for supplanting possible worlds with something better suited to the task. But I think these three examples, which figure prominently in the literature, do well to illustrate the limitation of possible worlds traditionally construed and their role in philosophical theorizing.

1.3 Counterpart theory and possibilities

Modal realism is, roughly, the view that modal propositions are grounded in the existence of concrete, non-actual individuals and worlds. It is given its fullest defense by David Lewis (Lewis, 1986a). According to modal realism,

> **Plenitude:** for any way the world (or a part of it) could possibly be, some world (or a part of it) is.

A world, as Lewis defines it, is a mereological sum of spatiotemporally (ST-) related individuals, satisfying the condition that if *w* is a world, *x* is part of *w*, and *x* and *y* are ST-related, then *y* is part of *w*. If ST-relations are non-modal, as is very plausible, then modal realism provides a reductive analysis of modality, a feature that is touted as one of its main virtues.

According to a standard possible worlds analysis, a *de re* modal statement such as 'Hillary Clinton could have won the election' is true just in case there is a world where Clinton—she herself, and not some simulacrum—wins the election. According to Lewisian modal realism, things exist in precisely one world, so if they are not to have all their properties necessarily, *de re* modal claims cannot be given the standard analysis.[6] Lewis proposes instead what he calls *counterpart theory*, according to which the statement 'Clinton could have won the election' is true just in case there is a world in which a counterpart of our actual Clinton wins the election. What is a counterpart of Clinton? It is someone who sufficiently resembles her in the relevant (i.e. contextually determined) respects, e.g. someone with a similar history to Clinton's, someone with the same or similar origin to Clinton, and so on. The counterpart relation therefore serves as a more flexible substitute for the usual relation of (transworld) identity, more flexible because it need not be transitive, symmetric, or one-to-one. In particular, an individual may have multiple counterparts within a single world, and it may even have different counterparts relative to different ways of being named.[7]

Whether a world represents of an individual that it has such and such properties therefore depends not just on the counterpart relation determined by the context, but also on which

counterpart of the individual we choose to do the representing. For instance, suppose a world w contains two identical twins, each of whom is a counterpart of Jane. Then if one is born at eleven o'clock and the other at twelve o'clock, it cannot be possible that Jane both is and is not born at eleven o'clock. If the counterpart relation is purely qualitative, as Lewis argues, then one can even have multiple counterparts within their own world.[8] This is important because it shows that counterpart theory (as the qualitative counterpart theorist conceives it) cannot identify possibilities, not even maximal ones, with possible worlds. Rather, a possibility must be identified with something finer-grained, such as a pair consisting of a world and a counterpart function f mapping individuals to at most one counterpart per world. For example, in the case of Jane, if we call the younger twin Molly and the elder Holly, and f and g are counterpart functions mapping Jane to Molly and to Holly respectively, then $\langle w,f\rangle$ and $\langle w,g\rangle$ represent two distinct possibilities for the same individual in the same world.

Why does this matter? First, and as Lewis himself notes, this marks a break from 'established theory' according to which (maximal) possibilities are just possible worlds. Lewis later makes the break by dropping the assumption that no two worldmates can be counterparts, but that break was made at the inception of counterpart theory by allowing things to have multiple counterparts in the same world.[9] Second, it shows that worlds by themselves cannot do the work the theory needs them to do. Worlds alone cannot serve as possibilities nor can sets of them serve as the contents of propositions, which are two of their primary roles.

The fact that possibilities cannot be worlds has important implications. By way of example, consider the doctrine known as *haecceitism*, which, as Lewis puts it, is the claim that two *possibilities* may differ in what they represent *de re* concerning an individual without thereby differing qualitatively. Lewis rejects haecceitism for a number of reasons, all of which rely on the crucial assumption that what does the representational work on a standard possible worlds account are worlds *and worlds alone*.[10] But given the fact that possibilities need to be something such as world-counterpart-function pairs, Lewis turns out committed to the doctrine. Very briefly, suppose our world, call it '@', is a world of one-way eternal recurrence. Then we have intrinsic duplicates and hence counterparts in each epoch, even though we ourselves inhabit exactly one of them.[11] Let $\langle @,c\rangle$ and $\langle @,c'\rangle$ be exactly alike except that c maps me to myself and c' is just like c except that it maps me to a qualitatively indiscernible other-epoch worldmate. Then these possibilities represent things as being qualitatively the same even though they differ in what they represent concerning me: one represents that I inhabit one epoch, and the other represents of me that I inhabit another. It follows that a qualitative counterpart theorist such as Lewis is committed to genuine haecceitism (rather than what he calls the 'cheap substitute') precisely because possibilities need to be played by entities having a richer structure than worlds.

1.4 Belief and centered worlds

Belief has posed a problem for the simple quantificational analysis for a variety of reasons, including the following main ones:

1. one can have inconsistent beliefs;
2. belief is not closed under strict implication;
3. first-personal or *de se* belief poses a unique challenge.

We have discussed the first two problems and have seen one way of at least formally dealing with them, i.e. in terms of (NS), but even given the immense flexibility such a framework provides, many are convinced that it is still unable to capture the distinctive feature of *de se* belief.[12]

Indexical expressions are those that are context-sensitive. This includes expressions such as 'I', 'you', 'here', and 'now'. A sentence containing an indexical expression cannot be assigned a truth-value until a context supplies a value for the indexical. Once those values are supplied, the sentence is assigned a content, that is, the proposition expressed by the sentence relative to the context. Thus, 'Jane is here' expresses different propositions depending on the context because 'here' can refer to different places. If 'here' refers to her house, then the proposition expressed is <Jane is at Jane's house>, and if it refers to the skating rink, the proposition expressed is <Jane is at the skating rink>.

Now suppose Jane and John have gone for a hike in the woods and are confronted by a bear who begins to chase Jane, and suppose she yells 'I am being chased by a bear!' Then, according to the traditional account, the proposition expressed by her utterance is <Jane is being chased by a bear>. If the contents of beliefs are propositions, as is commonly assumed, then John and Jane share the same belief concerning the situation in question, for they both believe that Jane is being chased by a bear. However, if what explains their behavioral differences—as Jane is the only one climbing a tree!—is a difference in belief, then John and Jane do not after all share the same belief, viz., the one that is expressed by Jane when she yells 'I am being chased by a bear'. Many take the difference in behavior to be explained by Jane's distinctively *de se* belief that she herself is being chased by a bear, versus John's merely *de re* belief that Jane is being chased by a bear.[13]

What Lewis proposes is to treat belief as the self-ascription of a *property*, rather than the belief in a content understood as a set of worlds. Call a *centered world* a pair consisting of a world and an individual on which the world is centered. To simplify matters, let us assume with Lewis that individuals are worldbound (i.e. exist at precisely one world), so that a centered world can be represented by a single individual, i.e. its center. (Centered worlds can be thought of as world-indexed individuals.) Then call a doxastic alternative of an agent an individual whom the agent cannot rule out as being herself, and say that an agent believes that she has property φ iff all her alternatives have φ. Instead of treating *de se* differently from *de re* or *de dicto* belief, we can say, for any *A* whatever, that an agent believes that *A* just in case each of her alternatives inhabits an *A*-world. Finally, to return to our example involving Jane and John, it is clear that they no longer share a belief, for only Jane's alternatives are being chased by a bear, and this difference in belief can be used to explain their difference in behavior.

To unify our account of propositions, a proposition as a set of worlds can be represented as a set of individuals, so the centered worlds analysis is not committed to two types of contents. For every set of worlds there corresponds the set of individuals each of which inhabits one of those worlds: say that such a set of individuals is true at a world just in case it contains an inhabitant of that world. On the other hand, not every *de se* content can be represented as a set of worlds. So, what to say about the truth of a *de se* content that corresponds to no set of worlds? Is singleton {*Jane*}, for instance, *true* or not? The question needs answering because we need to know when someone has a true belief about oneself, and not just when it is true that one believes something about oneself. The question does have an answer, but it can be given only once we supply an individual relative to which the content can be evaluated. {*Jane*} is true only relative to Jane, since only she can have the true, first-personal belief <I am Jane>. *De se* contents, then, do not stand alone in the same way sets of worlds do, which is a sign of their irreducibly indexical nature.

The role that worlds traditionally played as maximal consistent doxastic states has been assumed by individuals in order to capture the distinctively *de se*. There simply is no way to account for the difference in Jane and John's behavior in terms of a proposition understood as a set of worlds. There are other ways of accounting for the difference that allows us to hang onto the traditional view about propositions, but we will have to leave matters here.

1.5 Conditionals and impossible worlds

Consider the following pair of conditionals:

1. If Oswald didn't shoot Kennedy, then someone else did (true);
2. If Oswald hadn't shot Kennedy, then someone else would have (false).

Given our knowledge of the facts, viz. that Kennedy was shot, the first is true. But the second seems false, for we know of no other possible shooters besides Oswald. This difference in truth-value between the two conditionals implies a difference in meaning of their respective conditionals, the first being referred to as the *indicative* and the second as the *counterfactual*.

Both kinds of conditionals are in need of analysis, as it is clear that classical material implication fails to capture the meaning of either.[14] We will focus here on the counterfactual. The most natural intensional conditional we could define in a basic modal language is the strict conditional, but there is good reason for thinking that the counterfactual, which is also intensional, is not a strict conditional. In particular, the strict conditional satisfies antecedent strengthening:

(AS): $\Box(A \rightarrow B) \rightarrow \Box((A \wedge C) \rightarrow B)$,

Since if B is true in all A-worlds, it must also be true in all A-worlds that are also C-worlds. The counterfactual, on the other hand, seems not to satisfy (AS). Consider the following sequence:

1. if Otto had come, it would have been a lively party, but
2. if both Otto and Anna had come, it would have been a dreary party, but
3. if Waldo had come as well, it would have been lively, but . . .

The first proposition is true, the second false, the third true, and so on. Sequences sharing this pattern are called *Sobel sequences* and have attracted considerable attention.

What Stalnaker and Lewis proposed independently and at around the same time was roughly the following analysis:

(>): $A > C$ is true at a world w just in case C is true at all the A-worlds *closest* to w.[15]

This analysis has a number of important features. First, it is easy to see that antecedent strengthening fails, as do a number of other properties of the strict conditional that are intuitively invalid for the counterfactual. Second, and what will be the most important to us, any counterfactual with an impossible antecedent—a *counterpossible*—is necessarily true, a property (>) shares with the strict analysis of the conditional. However, it is intuitively false, e.g. that if Anaxagoras had squared the circle, nobody would have squared the circle. For this reason, many have rejected the Stalnaker-Lewis analysis, at least without some further amendment.

One obvious way to circumvent this problem is by admitting impossible worlds, so that the antecedent of a counterpossible can be true at an impossible world without the consequent also being true there.[16] This amendment requires no change to (>), only a broadening of the class of worlds. The toughest challenge facing such accounts concerns what to say about closeness now that impossible worlds are in the picture: for example, are possible worlds always closer than impossible ones to possible worlds?[17]

What sort of entity would an impossible world be? Most construct them from fairly uncontroversial entities (e.g. sentences, propositions, or states of affairs).[18] Some believe them to be real

or concrete worlds not unlike our own (ontologically speaking).[19] Others who accept the importance of impossible worlds reject the view that possible and impossible worlds ought to be the same sort of entity, ontologically speaking.[20]

It would be odd to believe that the Stalnaker-Lewis account gets it right for counterfactuals with possible antecedents (and possibly false consequents, which pose a similar problem), but that it fails terribly otherwise. That is, it would be strange to believe that the truth conditions for counterpossibles from those for counterfactuals with possible antecedents. Thus, any broadly Stalnaker-Lewisian account of the counterfactual (i.e. any variably strict analysis) will need to employ impossible worlds.[21] Since the Stalnaker-Lewis account is the best on offer, there is good reason for thinking that impossible worlds are as central as possible ones in our understanding of counterfactual reasoning.[22]

1.6 Conclusion

We have looked at three important areas where possible worlds have been supplanted, or at least supplemented, by entities better suited to the task originally assigned to worlds. There are at least two other important areas we have not discussed, which include the incomplete situations of situation semantics, and the propositions of two-dimensional semantics.[23]

The conclusion to draw from this is not that possible worlds have no important role to play in the analysis of modality, belief, and so on, but that the simple analyses of these notions that made possible worlds semantics initially attractive required, and continues to require, further refinement in light of the complexities exhibited by the phenomena under analysis.

Notes

1 There is a vast literature on the ontology of possible worlds. Some excellent sources include Lewis (1986a), Armstrong (1989), and Divers (2002).
2 The modal logic of the simplest quantificational analysis is called **S5**.
3 See Routley and Meyer (1977) for an argument against extensional reduction.
4 $\Box(A \rightarrow B)$ expresses that A strictly implies B.
5 The semantics given by (NS) is called *neighborhood* or *Scott-Montague* semantics. It can be seen as a generalization of the simple quantificational analysis.
6 Since Lewis accepts unrestricted mereological summation, there are individuals that do not wholly exist in one world but have parts from different worlds. We can ignore such individuals since they play no role in Lewis's analysis of modality.
7 Allowing for multiple counterparts within a world invalidates the necessity of identities: that if two things are identical, they are necessarily identical. See the translation scheme of Lewis (1968) for details, and for Lewis's original presentation of the theory.
8 Consider a world consisting of only two qualitative duplicates.
9 See Lewis (1986a) for his motivation for dropping the assumption. The break from established theory is made as early as Lewis (1968) and not in the much later Lewis (1986a), as Lewis suggests. It is also made by allowing multiple counterpart relations relative to a single context, a strategy already employed in Lewis (1971) for dealing with puzzles of coincidence. Consider a statue and the coinciding lump of clay that constitutes it and suppose, as Lewis does, that they are identical. If the expressions 'statue' and 'lump of clay' evoke different counterpart relations even relative to the same context, then one and the same world can represent differently concerning one and the same individual via multiple counterpart relations.
10 See Lewis (1986a) for his attack on haecceitism.
11 Lewis uses the example of duplicate worlds of one-way eternal recurrence to show that qualitatively identical, overlapping worlds can exhibit haecceitistic differences when the counterpart relation is identity. See Lewis (1986a).

12 Concerning inconsistent beliefs, (NS) clearly allows for them, but every outright contradiction is still represented by the empty set. The centered worlds analysis of *de se* content discussed further on in this chapter has the advantage that one can have inconsistent beliefs in numerously distinct ways: e.g. one can believe they are someone else, or that they lack a property they necessarily they have.
13 One might think that some other difference in belief is responsible for the difference in behavior, but we could assume that Jane and John share *all* of their beliefs, if none of them are distinctively *de se*. Moreover, we need not assume more generally that behavior is to be explained in terms of belief-desire psychology. See Ninan (2016) for a defense of the *de se* from those who are skeptical that there is a distinctive kind of first-personal content.
14 For arguments to the contrary, see e.g. Grice (1989), Jackson (1979), and Lewis (1986b).
15 See Stalnaker (1968) and Lewis (1973). For discussion concerning the relation of closeness between worlds, see Lewis (1979).
16 See e.g. Goodman (2004), Nolan (2013), Nolan (2017), and Brogaard and Salerno (2013) on impossible worlds amendments to the Stalnaker-Lewis account.
17 See Nolan (1997) for discussion concerning closeness between possible and impossible worlds.
18 See e.g. Vander Laan (1997).
19 See e.g. Yagisawa (2010).
20 See e.g. Berto (2009).
21 Of course, one could also provide a pragmatic defense of vacuously true counterpossibles similar in spirit to the Gricean defense of the material conditional.
22 The use of impossible worlds is nothing new in the semantics of conditionals. They are employed, e.g. in the semantics of relevance logic, where a counterexample to the irrelevant $(A \wedge \neg A) \rightarrow B$ is needed. See Priest (2008).
23 See Barwise and Perry (1983) on situation semantics, and García-Carpintero and Macia (2006) on two-dimensional semantics.

References

Armstrong, D. M. (1989). *Universals: An Opinionated Introduction*. *Focus Series*. Boulder, CO: Westview Press.

Barwise, J. and Perry, J. (1983). *Situations and Attitudes*. Bradford Books, Cambridge, MA: MIT Press.

Berto, F. (2009). Impossible worlds and propositions: against the parity thesis. *The Philosophical Quarterly*, 60(240): 471–486.

Brogaard, B. and Salerno, J. (2013). Remarks on counterpossibles. *Synthese*, 190: 639–660.

Carnap, R. (1946). Modalities and quantification. *Journal of Symbolic Logic*, 11: 33–64.

Divers, J. (2002). *Possible Worlds*. London: Routledge.

García-Carpintero, M. and Macia, J. (eds.) (2006). *Two-Dimensional Semantics*. Oxford: Clarendon Press..

Goodman, J. (2004). An extended Lewis/Stalnaker semantics and the new problem of counterpossibles. *Philosophical Papers*, 33(1): 35–66.

Grice, P. (1989). *Studies in the Way of Words*. Cambridge, MA: Harvard University Press.

Jackson, F. (1979). On assertion and indicative conditionals. *The Philosophical Review*, 88(4): 565–589.

Lewis, D. (1968). Counterpart theory and quantified modal logic. *The Journal of Philosophy*, 65(5): 113–126.

Lewis, D. (1971). Counterparts of persons and their bodies. *The Journal of Philosophy*, 68(7): 203–211.

Lewis, D. (1973). *Counterfactuals*. Cambridge, MA: Harvard University Press.

Lewis, D. (1979). Counterfactual dependence and time's arrow. *Noûs*, 13(4): 455–476.

Lewis, D. (1986a). *On the Plurality of Worlds*. Oxford: Blackwell Publishing.

Lewis, D. (1986b). Postscript to "Probabilities of conditionals and conditional probabilities". In *Philosophical Papers*, vol. 2, pp. 152–156. Oxford: Oxford University Press.

Mates, B. (1968). Leibniz on possible worlds. In van Rootselaar, B. and Staal, J. (eds.), *Logic, Methodology and Philosophy of S1cience*, vol. 3, pp. 507–529. Amsterdam: North-Holland Publishing Company.

Ninan, D. (2016). What is the problem of *de se* attitudes? In García-Carpintero, M. and Torre, S. (eds.), *About Oneself: De Se Attitudes and Communication*, pp. 86–120. Oxford: Oxford University Press.

Nolan, D. (1997). Impossible worlds: a modest approach. *Notre Dame Journal of Formal Logic*, 38(4): 535–572.

Nolan, D. (2013). Impossible worlds. *Philosophy Compass*, 8(4): 360–372.
Nolan, D. (2017). Causal counterfactuals and impossible worlds. In Beebee, H., Hitchcock, C., and Price, H. (eds.), *Making a Difference*. Oxford: Oxford University Press.
Priest, G. (2008). *Introduction to Non-Classical Logics: From Ifs to Is*, 2nd edn. Cambridge University Press.
Routley, R. and Meyer, R. K. (1977). Extensional reduction I. *Monist*, 60(3): 355–369.
Stalnaker, R. C. (1968). A theory of conditionals. In Rescher, N. (ed.), *Studies in Logical Theory*, pp. 98–112. Oxford: Blackwell Publishing..
Vander Laan, D. (1997). The ontology of impossible worlds. *Notre Dame Journal of Formal Logic*, 38(4): 597–620.
Yagisawa, T. (2010). *Worlds and Individuals, Possible and Otherwise*. Oxford: Oxford University Press.

Chapter 2
ACTUALISM[1]

Karen Bennett

2.1 Slogans and refinements

Let's begin with some handwavy sloganeering. Actualism is the view that everything actually exists, that there are no mere possibilia, that *this* is all there is. It is the opposite of possibilism, the view that there are mere possibilia—things that exist but do not actually exist.

Now let's try harder. What exactly does 'everything actually exists' mean? Here are a couple of bad readings. One bad reading involves taking the quantifier to be restricted to the actual world. This renders the slogan trivial; it just says that 'every actual thing actually exists,' which everyone accepts, including those who self-identify as possibilists. So that isn't what the slogan is intended to mean. The quantifier must be read unrestrictedly, as ranging over absolutely everything there is.

That thought can prompt a second bad reading. If the quantifier ranges over literally everything, the misguided thought goes, it must range over other possible worlds, and what exists in them. So a second bad reading would take the slogan to say that anything that possibly exists also actually exists—alternatively, that it's necessary that everything is actual. I call this 'domain inclusion actualism' in Bennett (2005). The main problem with reading the slogan this way is that it commits actualists to a claim that most of them do not want to endorse—namely, that it is not even *possible* for there to be anything that doesn't actually exist. If there actually are no *F*s, *F*s are not even possible. Now, some actualists—notably Bernard Linsky, Edward Zalta, and Timothy Williamson[2]—do endorse that claim; I shall return to them later. My point for now is that such a view goes beyond the central core of actualism. The claim that nothing could exist that doesn't actually exist—i.e., that 'aliens' are impossible—is an extra commitment and ought not be baked into the very characterization of actualism. Again, most self-identifying actualists do think that there *could be* things that do not actually exist (see, e.g., Stalnaker 2012; Menzel 2016), so charity requires interpreting their own slogan in a way that permits that.[3]

My point here is not that the domain inclusion understanding of actualism is not really actualist (though I return to this question in Section 2.6).[4] My point is rather that it is not the *only* way to be an actualist, and thus we need a reading of the slogan that is compatible with but does not entail it. And one is readily available.

The right reading is one that takes the quantifier to be unrestricted and otherwise takes the slogan at face value. Everything, absolutely everything, actually exists. The domain inclusion mistake

was to take this to mean that other possible worlds and their contents exist, too. But this is not what (typical) actualists think; not exactly. Most actualists will distinguish between other possible worlds and entities as—I reach for a general word here—devices, and as the situations and entities that would obtain if things were as that device suggests. For example, an actualist like Robert Adams (1974) who thinks that worlds are "world stories", or maximal sets of propositions, thinks that the *propositions* actually exist, even though various things they mention do not. An actualist like Robert Stalnaker (1976, 2012) who thinks that worlds are properties thinks that the properties actually exist, but that various things that would exist if they were exemplified do not. And an actualist like Alvin Plantinga (1974) who thinks that worlds are maximal possible states of affairs thinks that the states of affairs exist, but various things that would exist if they obtained do not.

The central strategy, common to all three of those views, is to draw a certain kind of distinction. As Stalnaker puts it,

> an actualist needs the distinction between existing and being exemplified in order to be able to explain the sense in which a merely possible world exists (a property the world might have had exists) and the sense in which it does not (no world that is that way exists). (2012: 8–9)

However, this passage is specific to Stalnaker's property-based account of worlds and isn't quite true in full generality. Not all actualists need the distinction between existing and being exemplified. First, some actualists rely on related but different distinctions. The Adams-style view (what David Lewis calls 'linguistic ersatzism' (1986: §3.2)) instead needs the distinction between existing and being true; the Plantinga-style view instead needs the distinction between existing and obtaining. Second, actualists who endorse domain inclusion—or a stronger view called necessitism, on which more in Section 2.3—need no such distinction at all. Neither do actualists who don't believe in worlds. So Stalnaker's claim needs to be both restricted and generalized. The right claim is this: actualists who utilize possible worlds and deny domain inclusion need a distinction between existing and existing *according to a world*. To exist is, well, to exist. To exist according to a world is to be such that it would exist if that world were actual—if the set of sentences were true, or if the maximal state of affairs were to obtain, or if the world-property were exemplified.

Overall, then, the point is this: actualism says that absolutely everything exists. On the non-domain inclusion approach, all the representations or properties that are other possible worlds do actually exist, though various things that exist according to them do not.[5] But that is no violation of actualism, because existence-according-to-a-world is not existence. Suppose I have a book that has pictures of dragons in it. Actualism commits me to the existence of the book, and to the existence of paper and ink, but not to the existence of dragons. Dragons exist according to the book, but they do not exist.

Actualism, then, is the view that absolutely everything actually exists. There are no mere possibilia. This is compatible with both domain inclusion (or necessitist) versions of actualism, and, as we have just seen, with more typical versions of actualism according to which, possibly, there are things that do not actually exist (see Bennett 2005 for an extended discussion of the actualist slogan).

2.2 Challenges to actualism

A number of challenges to actualism have been raised. In this section, I sketch three. All of them are primarily challenges for actualists who believe that possibly, there are things that don't actually exist—or, more generally, who allow that different things can exist in different worlds.

First, it is claimed that actualism plus the claim that there could be things that do not actually exist turns out to be incompatible with Kripke semantics. Let me be clear about the dialectic here: Kripke (1963) initially presented his semantics precisely as a way to *allow* the possibility of things that do not actually exist. The entire point was to avoid validating the Barcan formula ($\Diamond\exists x\phi \rightarrow \exists x\Diamond\phi$, or, equivalently, $\forall x\Box\phi \rightarrow \Box\forall x\phi$), the converse Barcan formula ($\exists x\Diamond\phi \rightarrow \Diamond\exists x\phi$, or $\Box\forall x\phi \rightarrow \forall x\Box\phi$), and the claim that everything exists necessarily ($\forall x\Box\exists y(y = x)$). Kripke's technical innovation was to do the formal semantics in a way that allows different worlds to have different domains, an advance (and complication) over the kind of constant domain system that Linsky and Zalta aptly call the "simplest quantified modal logic" (1994). (Hughes and Cresswell's classic modal logic textbook (1996) initially presents the simpler semantics before complicating things à la Kripke to allow varying domains.) So what is the problem supposed to be?

The problem is that it would appear that Kripke's metalanguage is possibilist. The formal details and options for repair have been presented excellently elsewhere (e.g. Linsky and Zalta 1994; Bennett 2005; Menzel 2016); I here confine myself to the intuitive gist. The key point is that a Kripke model for a language includes a set D of individuals and a function that assigns its members to worlds, yielding a D_w for each world w—w's domain, or what exists according to it. Let $D_@$ be the domain of the actual world. If things that do not actually exist are possible, there are things that exist in other worlds but not in this one—there are things in various D_ws that are not contained in $D_@$. Now, because the quantifiers are defined to be world restricted, one cannot say in the object language that there are things not in the actual world. But if one lifts the lid and looks in the metalanguage, there they are!

The second and third problems both bring out the fact that actualists have trouble accommodating possibility claims about specific nonactual individuals. Take the sentence 'Possibly, Santa Claus exists'. Actualists will, of course, deny that there is a thing, Santa Claus, who has the property of possible existence. They will instead say that it is possible for there to be a jolly fat man who lives at the North Pole, delivers presents, and so forth—that there is a certain general *de dicto* possibility, but no truly *de re* possibility about the existence of a specific person. That is a perfectly reasonable thing to say; no objection to actualism here. But the next two problems are ways of trying to press this issue.

The second problem, then, twists this line of thought by considering *de re* possibility claims about *actual* things. The trick is to start with the fact that many actual things are contingent; it is possible for them not to exist. In particular, 'possibly Karen Bennett does not exist' is true. Let w be a world according to which I do not exist. Is it true in w that I could exist? That is, is 'possibly Karen Bennett exists' true at w? If actualism is true at w, that sentence looks a lot like 'possibly Santa Claus exists' looks here in the actual world. So perhaps the actualist must say that all that is true is some general descriptive possibility? But in this case, there is greater pressure to say that 'possibly Karen Bennett exists' is true and *de re* in w—I'm right here typing, after all! In Bennett (2005), I call this an 'out-and-back world-hopping argument'; two classic discussions of issues in this ballpark are Adams (1981: 28–32) and Plantinga (1983).

A third problem, usually called "the McMichael problem" (McMichael 1983), has to do with the actualist's ability to accommodate iterated *de re* modal claims. I don't have a sister. But it seems that I could have had one, and indeed that I could have had a sister who was a teacher. And isn't it possible for her, that very teacher, to have instead been a lawyer? After all, it seems that I could have had a sister who was torn between two career choices, who took the LSAT and applied to law school, but who ultimately chose to be a teacher instead. So isn't it possible that I have a teacher sister who—that very woman—could have been a lawyer? However, it is hard to see how an actualist can handle claims like this. They take the following form: $\Diamond\exists x(Fx \,\&\, \Diamond Gx)$.

Making sense of the embedded possibility apparently requires tracking a specific nonactual individual across worlds, which is not easy for most versions of actualism.[6]

There are solutions to these puzzles, of course. However, providing a full taxonomy of the various possible moves goes beyond the scope of this brief essay. For more discussion, I recommend Bennett (2005), Menzel (2016), and the many classic papers cited therein. All I shall say here is that two distinctions are important to have on the table when assessing the challenges: that between what I have called 'proxy' and 'nonproxy' actualism (Bennett 2006), and that between necessitism and contingentism. I shall introduce these in reverse order in the next two sections.

2.3 Should we scrap the actualism/possibilism dispute in favor of the necessitism/contingentism dispute?

The labels 'necessitism' and 'contingentism' were introduced by Timothy Williamson (2010, 2013).[7] Necessitism is the view that everything necessarily exists—that necessarily, everything is necessarily something. Contingentism is the negation of this; it says that it is contingent what exists. And Williamson argues that the actualism-possibilism distinction should simply be abandoned in favor of this alternative distinction:

> the use of [the terms 'possibilism' and 'actualism'] has become badly confused . . . It is better to make a fresh start with fresh terminology and clearer distinctions. Thus the proposal is to abandon that debate as hopelessly muddled, and to get on to the clearer necessitism-contingentism debate.
>
> (Williamson 2013: 23–24; reproduced from Williamson 2010: 662–663)

Why does he think this? He says that on standard formal treatments of the 'actually' operator, it makes a difference only to the truth value of a sentence when it is within the scope of a modal operator. So, he claims, 'everything actually exists' is equivalent to 'everything exists,' and actualism is thus trivially true. Similarly, 'there exist things that do not actually exist' is equivalent to 'there exist things that do not exist,' rendering possibilism trivially false. Williamson goes on to say that "the actualist needs another reading of 'actual' than the one well understood in modal logic," a reading on which "being actual had better be . . . something harder than just being" (Williamson 2013: 23; see also Williamson 2010: 662–663; Williamson 1998: 259 for similar discussion). He does not think such a reading is easily available and thus proposes to consign discussions of actualism and possibilism to the rubbish bin.

However, this line of thought fails for dialectical reasons. I will bring this out in two related ways. First, let's grant for the sake of argument that standard formal treatments of 'actually' treat it as inert when not in the scope of a modal operator.[8] What follows is not that it is trivial that everything is actual, but rather that standard formal treatments *assume actualism*. David Lewis, the most well-known contemporary possibilist, certainly does not agree that 'actually' adds nothing to sentences without a modal operator.[9] Speaking unrestrictedly, he thinks it true that there are talking donkeys even though there actually are no talking donkeys. That's because he does have the requested reading of 'actual' on which being actual is harder than merely existing—for Lewis, to be actual is to be spatio-temporally connected to the speaker. For him, then, the earlier sentence amounts to the claim that while there are (unrestrictedly) talking donkeys, none of them are spatio-temporally connected to me. The upshot is that it is not legitimate to rule out possibilism on the basis of formal treatments of the 'actually' operator that possibilists would not accept. (Here is another way to put the point. If actualism is true, then maybe the correct modal

logic will render the actualist slogan trivial. But even if so, it doesn't follow that the *dispute* is trivial, because it is unclear whether actualism *is* true.)

Second, consider Williamson's claim that the actualist needs a reading of 'actual' on which being actual is harder than merely existing. I can see three interpretations of this claim, but none are dialectically acceptable. They differ on what exactly it is that requires that being actual be harder than merely existing. The first interpretation—and least likely to be the intended one—says that it is the *truth of actualism* that requires that being actual be harder than merely existing. But actualism is precisely the view that there is no difference between existing and being actual! So the first interpretation is incoherent (and, again, unlikely what Williamson means). The second interpretation says that it is the *non-triviality of actualism* that requires that being actual be harder than merely existing. This is not much of an improvement over the first interpretation; it amounts to saying that if actualism is a substantive position, it is false. The third interpretation says that it is the *worthwhileness of the actualism-possibilism debate* that requires that being actual be harder than merely existing. But this cannot be correct either. After all, that is the very question at issue. Quite generally, it cannot be the case that the tenability or worthwhileness of a dispute about q requires that both sides agree that q! Again, we are led back to the thought that there can be a substantive debate about the truth of a thesis that might be such that *if* it is true, it is trivial.

2.4 On the relations between actualism, possibilism, necessitism, and contingentism

So we have not been given adequate reason for abandoning the actualism-possibilism debate in favor of the necessitism-contingentism debate. But, of course, I have not said anything to suggest that we ought not to think about the necessitism-contingentism debate *also*. They are just two distinct debates. That means that we face the question of what the relationship is between the two debates, or at least between the two pairs of positions. There are two main connections—or one, depending on how you count; the 'two' are contrapositives. First:

- Necessitism entails actualism, in letter if not in spirit.

I shall explain what I mean by 'the letter' and 'the spirit' of actualism later. For now, just note the basic point: if everything exists necessarily, everything exists actually, too. In terms of worlds, the idea is that if everything is in every world, *a fortiori* everything is in the actual world.[10] So necessitism entails the domain inclusion version of actualism. (The converse does not hold; domain inclusion does not entail necessitism. A domain inclusion actualist could think that some actual things exist contingently, even though she denies that anything could exist that doesn't actually exist. In terms of worlds: she could think the actual world is the biggest world.) Second:

- Possibilism entails contingentism.

If there exist things that do not actually exist, then there are things that do not necessarily exist.[11]

With these claims in hand, it is useful now to revisit the challenges to actualism that I laid out in Section 2.2: they can be understood as providing reason for actualists to be necessitists. Certainly Linsky and Zalta (1994, 1996) deploy them that way. (Williamson, in contrast, rejects the label 'actualist', and the considerations in Section 2.2 do not play a large role in his defense of necessitism.) If everything necessarily exists, there is no need to worry about possibilia hiding in Kripke's metasemantics; every world has the same domain, namely D. Further, there is no

need to worry about contingent existence claims entailing otherworldly *de re* possibilities about actual entities; there are no true contingent existence claims. And there is no difficulty with iterated modal claims that ascribe *de re* possibilities to merely possible things. If everything exists necessarily, everything is available in every world. Now, of course those arguments provide at best defeasible reason for actualists to be necessitists; contingentist strategies are available, and necessitist actualism faces objections too (e.g. Hayaki 2006).

2.5 Yet another distinction

At this point it is necessary to introduce yet another distinction—between proxy and non-proxy actualism, views I briefly mentioned earlier. The proxy/non-proxy actualism distinction crosscuts the necessitist/contingentist distinction in that proxy actualists can be either necessitists or contingentists (though all necessitists are proxy actualists as I characterize the view).

What is proxy actualism? It is the view that anything that possibly exists has a proxy or trace in every world. (For a more careful characterization, see Bennett 2006: especially 272.) Some proxy actualists take the proxy relation to be identity—i.e., they take the proxy for an entity *E* to be *E* itself. Such proxy actualists—notably Linsky and Zalta (1994, 1996)—are necessitists. Indeed, every necessitist is a proxy actualist. But not every proxy actualist is a necessitist; contingentist proxy actualism is possible too, and indeed, so is the view naturally suggested by the word 'proxy'. A contingentist proxy actualist takes the proxy relation to be something other than identity, and thus takes the proxy for an entity *E* to be something other than *E*. Plantinga (1974), for example, thinks that entities like me exist contingently, but that I have an individual essence that exists necessarily, and which—in my terms, not his—serves as my proxy in every world. So proxy actualism crosscuts the necessitism-contingentism distinction, but it should be clear from even this quick sketch that contingentist proxy actualism will share many of the virtues and problems of necessitist actualism. It is, after all, structurally isomorphic to it. See Bennett (2006) for extended discussion, though I do not use the label 'necessitist'.

A lot of distinctions are on the table at this point; a recap of the various versions of actualism is in order. Here are intuitive versions; more careful versions that avoid the apparent quantification over possibilia are in the footnote.

Actualism: everything actually exists.
Domain inclusion actualism: every possible thing actually exists.
Proxy actualism: every possible thing either itself necessarily exists, or else has a distinct proxy that necessarily exists.
Necessitism: every possible thing necessarily exists.[12]

With one exception, each thesis entails the one above it, but not below it. (The exception is that only the first, necessitist disjunct of proxy actualism entails domain inclusion actualism.) It is worth reiterating that this means that actualism itself does not entail any of the further claims; those are versions of actualism, but not the only versions. I myself strongly prefer contingentist non-proxy non-domain-inclusion actualism, and I am pleased to have Robert Stalnaker (2012) as an ally. But while I recognize that a serious gauntlet has been thrown by my necessitist and proxy actualist opponents, this essay is not the place to fully take it up.

Still, I would like to conclude by returning to my cryptic reference to the 'spirit' and 'letter' of actualism. Earlier, I said that necessitist actualism obeys the letter but not the spirit of actualism. I will now say the same about proxy actualism more generally. But what is that supposed to mean?

2.6 The spirit vs. the letter of actualism

In Bennett (2006), I declared that proxy actualist views—a category that includes necessitist views—were not in fact actualist at all (Section 2.7). I probably should not have said that. The prevailing usage of the term 'actualist' is such that acceptance of the slogan 'everything is actual' is necessary and sufficient for counting as an actualist, and it should be clear from this essay that proxy actualists accept the slogan. Hence I now claim that proxy actualism does obey the letter of actualism. I stand by most of my earlier discussion; it is just that I now only take it to show that proxy actualism is not in the *spirit* of actualism.

I do not intend to argue anew for this claim. But the basic idea starts with the fact that the proxy actualist postulates an awful lot of particular entities and then engages in various bits of fancy footwork to make them sound acceptable. Both Williamson and Linsky and Zalta, for example, think that there actually exist a plethora of nonconcrete entities that only instantiate modal properties (this is a slight simplification). That one is possibly a dragon, possibly spatially located, possibly nonconcrete, and so forth. Strictly speaking, it exists. But it isn't a dragon.

Obviously, this kind of view carries heavy ontological commitments. I am not going to use that to argue against the view—I am not going to argue against the view at all—but I do want to use it to support the idea that views like this go against the *spirit* of actualism. I take it that the spirit of actualism is something like this: not only does everything actually exist, but also no special entities are needed to make sense of modal talk. There are no hidden particular actual entities that do duty for particular possible objects. There's just, well, *this*—insert expansive gesture here. There are tables and people and properties and relations and so forth. And those ontological commitments are enough to concoct a theory of modality. That is, it is in the spirit of actualism to rely only upon ontological commitments we have, as it were, already incurred. It is true, as I explained previously, that the contingentist, non-domain inclusion, non-proxy actualist needs the notion of existence according to a world. But existence according to a world is just like existence according to a picture or a story; 'existence according to' is a notion we need *anyway*. It is not a special tool, and certainly not a special ontological commitment.

I recognize that this thought about the spirit of actualism is pretty nebulous. My defense of it will be even more so. It is just that most actualists resist possibilism in part because of a preference for a leaner ontology. Further, two of the main three actualist views from the renaissance of interest in modality in the 1970s and 1980s—Stalnaker's, Adams's, and Plantinga's—are non-proxy views, and the one that is a proxy view—Plantinga's—was, I think, viewed with suspicion on ontological grounds. So the thought is that contingentist, non-domain-inclusion, non-proxy views are more in the spirit of actualism than proxy views, necessitist or not. (Menzel 2016 seems sympathetic to this thought, as evidenced by his calling non-proxy actualism 'strict' actualism.) But at the end of the day, 'actualism' is just a label, and disputes about whether a view counts as actualist or not—or more actualist than another—are just verbal disputes. We can do just as well with the more specific, if more unwieldy, labels I have used throughout this essay.

Notes

1 Thanks to Ted Sider and Michael Bergmann for helpful comments.

2 Williamson does not self-identify as an actualist; he rejects the actualism/possibilism distinction. See Sections 2.3 and 2.4 of this chapter.

3 Here is another way at exactly the same point: this domain-inclusion way of understanding the actualist slogan has it stating precisely the conditions that validate the Barcan formula: $\forall x \Box \phi \rightarrow \Box \forall x \phi$. But most actualists—the exceptions have already been noted—deny the Barcan formula. They think that it is possible for there to be, say, a giant telepathic hedgehog, but do not think there is any particular actual object that could be a giant telepathic hedgehog.

4 I suggest otherwise at the end of my 2005 (320); Ross Cameron is right to criticize me for that (2016: 116 n. 7).
5 Or, more awkwardly but avoiding the apparent quantification over the things that only exist according to other worlds: other possible worlds do actually exist, though according to them there exist various things that do not actually exist.
6 It might be the case that actualists who do away with possible worlds altogether do not have a problem here. Thanks to Ted Sider for discussion.
7 I am not sure that Williamson used the *labels* before 2010, though he certainly discussed the *issues*: e.g. Williamson (1998, 1999, 2000, 2001).
8 I am not entirely sure this is correct. For example, Crossley and Humberstone (1977) explicitly deny that Ap is equivalent to p, saying that the closest equivalency theorem is $A(Ap \leftrightarrow p)$. But I am not a logician and will not challenge the claim here.
9 Williamson seems to think that discussions of the actualism–possibilism distinction ought not mention modal realism at all. I do not understand this. He is right to point out that "most participants in the actualism–possibilism debate reject modal realism, and hold that if there are such other spatiotemporal systems, they are just as actual as our own" (2013: 22), because most participants in the debate are actualists. He is also right to say that "the [actualism-possibilism] debate is not about whether there are other spatiotemporal systems" (2013: 22). This is correct, both because (as he notes) someone might believe in other actually existing spatio-temporal systems, and because a possibilist strictly speaking need not believe in worlds at all. But none of this justifies taking modal realism off the table altogether, as Williamson seems to want. Modal realism—understood on its own terms, not merely as the view that there are worlds spatio-temporally unconnected to ours—is a clear example of a possibilist view. And if one wants to deny modal realism on the grounds that these worlds spatio-temporally unconnected to ours would have to be actual, then one is denying modal realism *because it is possibilist*. And that is dialectically inappropriate when the question on the table is not whether possibilism is true or false, but whether the positions can be formulated coherently. For related discussion, though in the context of defending possibilism rather than the actualism/possibilism debate, see Lewis (1986: 97–101).
10 This means that Williamson is, in fact, an actualist. He does not self-identify as such because of his rejection of the actualism–possibilism distinction.
11 The claim that possibilism entails contingentism, plus the claim that Lewis is a possibilist, entail that Lewis is a contingentist. But both Williamson (2013: 24) and Stalnaker (2012: 1–2) suggest that Lewis is a necessitist. This isn't the best way to classify him. By his own lights, on his own semantics, what is necessary is what is true in all worlds, and he does not think that most existence claims are true in all worlds. While it is true that the pluriverse itself—all of modal space—is fixed, that does not make him a necessitist given his own treatment of the quantifiers and modal operators. Besides, down this road lie problems of advanced modalizing that are beyond the scope of this essay. See Divers (1999) for what I believe is the first statement of those problems.
12 Actualism: everything actually exists.
Domain inclusion actualism: necessarily, everything actually exists.
Proxy actualism: either necessarily, everything is necessarily something, or necessarily, everything has a proxy that is necessarily something.
Necessitism: necessarily, everything is necessarily something.

References

Adams, R. (1974) "Theories of Actuality," *Noûs* 8: 211–231.
————. (1981) "Actualism and Thisness," *Synthèse* 49: 3–41.
Bennett, K. (2005) "Two Axes of Actualism," *The Philosophical Review* 114: 297–326.
————. (2006) "Proxy 'Actualism'," *Philosophical Studies* 129: 263–294.
Cameron, R. (2016) "On Characterizing the Presentism/Eternalism and Actualism/Possibilism Debates," *Analytic Philosophy* 57: 110–140.
Crossley, J., and Humberstone, L. (1977) "The Logic of 'Actually'," *Reports on Mathematical Logic* 8: 11–29.
Divers, J. (1999) "A Genuine Realist Theory of Advanced Modalizing," *Mind* 108: 217–239.
Hayaki, R. (2006) "Contingent Objects and the Barcan Formula," *Erkenntnis* 64: 87–95.
Hughes, G.E., and Cresswell, M.J. (1996) *A New Introduction to Modal Logic*, New York: Routledge.

Kripke, S. (1963) "Semantical Considerations on Modal Logic," *Acta Philosophica Fennica* 16: 83–94.
Lewis, D. (1986) *On the Plurality of Worlds*, Oxford: Blackwell.
Linsky, B., and Zalta, E. (1994) "In Defense of the Simplest Quantified Modal Logic," *Philosophical Perspectives* 8: 431–458.
Linsky, B.–. (1996) "In Defense of the Contingently Nonconcrete," *Philosophical Studies* 84: 283–294.
McMichael, A. (1983) "A Problem for Actualism about Possible Worlds," *The Philosophical Review* 92: 49–66.
Menzel, C. (2016 [2000]) "Actualism," *The Stanford Encyclopedia of Philosophy*. Available at: https://plato.stanford.edu/archives/win2016/entries/actualism/.
Plantinga, A. (1974) *The Nature of Necessity*, Oxford: Clarendon Press.
————. (1983) "On Existentialism," *Philosophical Studies* 44: 1–20.
Stalnaker, R. (1976) "Possible Worlds," *Noûs* 10: 65–75.
————. (2012) *Mere Possibilities*, Princeton, NJ: Princeton University Press.
Williamson, T. (1998) "Bare Possibilia," *Erkenntnis* 48: 257–273.
————. (1999) "Truthmakers and the Converse Barcan Formula," *Dialectica* 53: 253–270.
————. (2000) "Existence and Contingency," *Proceedings of the Aristotelian Society* 100: 117–139.
————. (2001) "Necessary Existents," in A. O'Hear (ed.), *Logic, Thought, and Language*. Cambridge: Cambridge University Press, 233–251.
————. (2010) "Necessitism, Contingentism and Plural Quantification," *Mind* 119: 657–748.
————. (2013) *Modal Logic as Metaphysics*, Oxford: Oxford University Press.

Chapter 3

COUNTERFACTUAL CONDITIONALS

Dorothy Edgington

Introduction

Counterfactuals are conditional judgements about what *would* have happened *if* such-and-such had happened, or more generally, what *would* be the case *if* such-and-such were the case. Thus, they are typically about unrealised possibilities. These judgements are useful to us, for instance in making inferences: "It's not a problem with the liver," says the doctor, "for if it had been a problem with the liver, it would have shown up in the blood test." They are also common in the expression of regret, or relief: "If I had left ten minutes earlier, I would have avoided the accident," "if we had taken the main road, we would have been stuck in traffic for hours."

As well as being an important class of judgements in ordinary life, counterfactuals have played a role in philosophy in the last half-century, in the elucidation of other notions, for instance, dispositional properties: an object's being fragile at a time t when it is not dropped seems to be connected with the fact that if it *had* been dropped at t, it would have broken. There are theories of causation such that "flipping the switch *caused* the lamp to light," is explained as: "if the switch had not been flipped, the lamp would not have lit" (Lewis 1973b); it has been argued that these are the conditionals needed in decision theory (Gibbard and Harper 1981), and that the difference between knowledge and mere true belief that p may be been explained by the claim that in the former case, if p were not true, one would not believe it (Nozick 1981).

Counterfactuals are contrasted with indicative conditionals, the latter being claims that something *is* (was, or will be) the case *if* such-and-such is true. It is controversial how deep the distinction is between the two kinds. Many examples suggest that they are closely aligned. "If they're here by eight, we'll eat at nine" is rephrased hungrily at ten, "If they had been here by eight, we would have eaten at nine." "If you go in you'll get hurt," I say. You look sceptical but stay outside, when there is a large crash as the ceiling collapses. "You see, if you had gone in you would have got hurt—just as I said." Strawson (1986: 230), considering pairs like these, says "The least attractive thing to say about the difference between [the members of such a pair] is that 'if' has a different meaning in one from the meaning it has in the other." But other examples highlight the contrast between them:

(1) If Oswald didn't kill Kennedy, someone else did (indicative).
(2) If Oswald hadn't killed Kennedy, someone else would have (counterfactual).

One can accept (1) yet reject (2) (see Adams 1970).

For those who accept the classical truth-functional account of the indicative conditional, e.g. Jackson (1987) and Lewis (1976: 152–6), counterfactuals are a different story. Those who advocate an alternative account of indicatives are more likely to aim at a broadly unified account of the two kinds of conditionals, e.g. Stalnaker (1968), Adams (1975) and Lycan (2001). But some who adopt an alternative account of indicatives persist in taking counterfactuals to be radically different, e.g. Gibbard, who writes, "the apparent similarity between these two 'if' constructions hides a profound semantic difference" (Gibbard 1981: 211), and Bennett (2003: 256).

Counterfactuals are also called "subjunctive conditionals." Neither name is perfectly apt, all agree. Stalnaker (2005) speaks of "the combination of tense, aspect and mood that we have gotten in the habit of calling 'subjunctive'." As for "counterfactual": a counterfactual is not just a conditional with a false antecedent, as plenty indicative conditionals have false antecedents. Moreover, it may be quite correct even when the antecedent is true: "If you had dropped it, it would have broken." "You're right—I did drop it, and it broke, but I did such a marvellous repair job, you never could tell." Nevertheless, it might be said, counterfactuals are *intended* to be about unrealised possibilities. Often they are, but not always. There is also this use, highlighted by Anderson (1951): a patient is brought to hospital in a coma and the doctor says, "If he had taken arsenic, he would be showing just these symptoms" (those which he in fact shows). Or consider this exchange: "A bus is coming." "How do you know?" (for we can't see the oncoming traffic). "People in line are picking up their bags and moving forward—and *that's what they would be doing if a bus were coming*." In these cases, "counterfactuals" are used in inference to the best explanation of what we observe. All that can be said in justification of the label is that, while indicatives are standardly used when one takes it to be an open question whether the antecedent is true—"If it rains . . .," "If John is at home . . ."—counterfactuals are standardly used when we take ourselves to know that the antecedent is false—"If it had rained . . .," "If John had been at home . . ."

Early theories

The problem of counterfactuals was first raised in the philosophy of science, in connection with dispositional properties, and with laws of nature. What is the difference between a generalization which is a law, and a generalization which, though true, is accidentally true? A popular answer was: the former has counterfactual implications, the latter does not. "All copper conducts electricity" supports the claim that if this plate were copper, it would conduct electricity. "All the coins in my pocket are silver" does not support "If this penny were in my pocket, it would be silver." Thus it was expected that the notions of laws of nature and counterfactuals should be mutually illuminating. In a famous essay, Goodman (1947) tried to analyse counterfactuals as law-governed conditionals. He discusses this example:

> If the match had been struck, it would have lit.

He suggests that this is true if and only if there is a law of nature, and there are facts about the match and its surroundings (it was dry, there was oxygen, etc.) from which, together with the assumption that it was struck, we can deduce that it lit. A theory of the following shape emerges (I shall use throughout Stalnaker's symbol ">" for the counterfactual conditional connective):

> A counterfactual $A>C$ is true iff there is a conjunction of truths T which include a law of nature [and satisfy condition X] such that $A\&T$ entails C.

X is a placeholder for the difficult bit. In fact, the match was not struck, and did not light. Assuming that it was struck involves supposing that some things which are actually true, were not true. For instance, the match remained motionless and untouched on the table. True, but this wouldn't have been true if the match had been struck. So we need to forget about that, in considering what would have been the case if it had been struck. But we need to rely upon other things that are actually true, remaining true if the match had been struck; for instance, the fact that it was dry. What distinguishes the facts we may rely on, from those which we may not, when we make a counterfactual supposition? Using Goodman's name for the problem: which facts are *cotenable* with the assumption that the antecedent is true? (Then the square bracket in the aforementioned theory becomes "and are cotenable with *A*"). Goodman defined cotenability thus:

> *B* is cotenable with *A* iff it's not the case that if *A* had been true, *B* would not have been true.

But now circularity looms, as Goodman saw: we need cotenability in the account of counterfactuals, and we need counterfactuals in defining cotenability. Consider:

(1) If the match had been scratched, it would have lit ($S>L$).
(2) If the match had been scratched, it would not have been dry ($S>\neg D$).

Suppose (1) is true and (2) is false. How does the theory yield that result? With (1), there is a derivation from the assumption that *S*, together with a law, and facts such as *it was dry* (*D*), to the conclusion that *L*. But with (2), there is a derivation from the assumption that *S*, together with the same law, and facts such that *it did not light* ($\neg L$), to the conclusion that it was not dry. The asymmetry must lie in which facts are cotenable with the assumption that *S*. (1) is true because (inter alia) the match was dry (*D*), and this is cotenable with the assumption that *S*. (2) is false because, although the match did not light ($\neg L$), this is not cotenable with the assumption that it was struck. Applying the definition of cotenability to these claims amounts to:

> It is not the case that if the match had been struck, it would not have been dry [$\neg(S>\neg D)$].
> It is the case that if the match had been struck, it would have lit [$S>L$].

Now the circularity is obvious. Why is "$S>L$" true and "$S>\neg D$" false? Because "$S>\neg D$" is false and "$S>L$" is true.

Goodman decided he had reached a dead end. He succeeded in highlighting a pervasive problem for all theories: which facts are we entitled to hold on to, and which facts are jettisoned, when we make a counterfactual supposition?

Here are some further objections to the "law-governed" account of counterfactuals, independent of cotenability: first, it is not clear that a law of nature is always needed for a counterfactual claim. Consider:

> If I had known you were coming, I would have baked a cake.
> If you had asked me yesterday, I would have accepted.

It is unclear that there is a law of nature connecting antecedent to consequent. That is a hard question in the philosophy of mind. But our confidence in counterfactuals such as these does not seem to depend on that question.

Second, there are cases where there is definitely not a law from which the consequent can be deduced from the antecedent together with other facts, which seem to merit high confidence even if they do not merit certainty; for instance, "If you had tossed that (fair) coin ten times, you would have got at least one head." This, it seems, is highly probable, though not certain. But on the law-governed account, it should be dismissed as certainly false.

Third, it is not entirely clear that accidental generalisations never support counterfactuals. Van Fraassen (1981) gave this example: the conjuror produces a penny and claims to have gotten it from the boy's pocket. The boy responds: "That coin didn't come from my pocket. All the coins in my pocket are silver. If that coin had come from my pocket, it would have been a silver coin."

Possible-worlds semantics

Saul Kripke (1963) devised a possible-worlds semantics for modal logic, and this style of semantics came to be applied more generally. It is a framework particularly suited to counterfactuals, which are typically about unrealised possibilities. Stalnaker and Lewis, independently, developed similar theories. These have become the standard approach. The opening sentence of Lewis's *Counterfactuals* gives the gist:

> "*If kangaroos had no tails, they would topple over*" seems to me to mean something like this: in any possible state of affairs in which kangaroos have no tails, and which resembles our actual state of affairs as much as kangaroos having no tails permits it to, the kangaroos topple over.
>
> (Lewis 1973a: 1)

Stalnaker wrote: "Consider a possible world in which A is true and which otherwise differs minimally from the actual world. 'If A, B' is true (false) just in case B is true (false) in that possible world" (Stalnaker 1968: 33–4). (Note that Stalnaker gives an account of both indicative and counterfactual conditionals along these lines, while Lewis's account is specific to counterfactuals.)

Similarity to the actual world plays the role in these theories that cotenability plays in Goodman's. Stated in terms of possible worlds, Goodman's theory has the shape: $A > C$ is true iff in any possible world in which A is true and X is satisfied, C is true. For Lewis and Stalnaker, the problem of specifying X is the problem of deciding which worlds are most similar to actuality, or, as they say, "closer" to the actual world.

Lewis's truth conditions are as follows: (i) If A is true at no possible world, $A > C$ is vacuously true. (ii) $A > C$ is non-vacuously true iff some $A\&C$-world is closer to the actual world than any $A\&\neg C$-world. "In other words, a counterfactual is non-vacuously true iff it takes less of a departure from actuality to make the consequent true along with the antecedent than it does to make the antecedent true without the consequent" (Lewis 1973a: 164).

Stalnaker agrees that counterfactuals with impossible antecedents are vacuously true. For the non-vacuous case, his semantics is equipped with a "selection function" f, which yields, for any world w and proposition A, the world which is the closest A-world to w. Restricting our attention to the assessment of conditionals at the actual world, f selects the closest world to actuality in which A is true. The conditional $A > C$ is true iff C is true at the selected world.

Some of the differences between Lewis and Stalnaker are merely notational, but some are significant and we shall return to them.

Let us look at some features of the logic of counterfactuals on these accounts. First, if A is true, $A > C$ is true iff C is true. The closest world to the actual world is the actual world itself. Lewis calls this principle "centering."

The counterfactual conditional is stronger than the truth-functional conditional ⊃. Therefore, any inference pattern with conditional premises but no conditional in the conclusion which is valid for $A \supset B$, is valid for $A > B$. Thus, these principles are valid:

A; $A > B$; so B (modus ponens)
$\neg B$; $A > B$; so $\neg A$ (modus tollens)
$A \vee B$; $A > C$; $B > C$; so C.

But not all inference patterns valid for ⊃ which have conditional conclusions, are valid for >. Three principles in particular, as Lewis and Stalnaker note, fail: antecedent strengthening, transitivity, and contraposition.

The first is the inference $A > C$; so $(A \& B) > C$. Here is a counterexample: "If you had struck the match, it would have lit" may be true, while "If you had dipped it in water and struck it, it would have lit" is false. The closest strike-worlds are light-worlds; but the closest dip-in-water-and-strike worlds are not-light worlds.

Strengthening of the antecedent is a special case of transitivity, for obviously $A \& B > A$, but we can have $A > C$ yet not $A \& B > C$. Other failures of transitivity can be constructed; for instance:

> If Brown had been appointed, Jones would have resigned immediately afterwards;
> If Jones had died before the appointment was made, Brown would have been appointed;
> So: if Jones had died before the appointment was made, Jones would have resigned immediately after the appointment was made.

Departing from reality enough to get Brown appointed has Jones resigning; departing further, to get Jones dead, has Brown appointed. On this reading, the conclusion does not follow.

There is, however, a restricted form of transitivity that is valid:

$A > B$; $(A \& B) > C$; so $A > C$

That is, if the second premise $B > C$ can be acceptably strengthened in this way, the reasoning goes through.

There are also failures of contraposition. Here is Stalnaker's example (1968: 39):

> If the US had halted the bombing, North Vietnam would not have agreed to negotiate;
> But not: If North Vietnam had agreed to negotiate, the US would not have halted the bombing.

And Conditional Proof fails. "$\neg(A \& B)$; A; therefore $\neg B$" is a valid argument form. But the move from this to "$\neg(A \& B)$; therefore $A > \neg B$" is invalid. Let A be "She was hit by a bomb yesterday" and B be "She was injured yesterday." It does not follow from the falsity of $A \& B$ that if she had been hit by a bomb, she would not have been injured, i.e. that in the closest world in which she was hit by a bomb, she was not injured.

Similarity between worlds

Lewis accepts that similarity is vague. Comparing cities, or faces, or worlds, there may be no determinate answer to the question: is x more similar to z than y is? But equally, there may be

no determinate answer to the question: what would have happened if *A* had been true? Lewis's aim is to analyse one somewhat indeterminate notion in terms of another. On the other hand, similarity is not so vague as to be useless. Often, clear judgements can be made about the overall similarity of cities, faces, etc., or of how life-like as opposed to fantastical is a novel or a film.

Unlike Goodman's theory, laws of nature are not mentioned in Lewis's truth conditions. But he can explain why they have an important role in the assessment of counterfactuals. Laws of nature are important truths which say much about the character of the world. In general, the difference between two worlds with the same laws will be less than the difference between two worlds with different laws. If we keep as close to the actual world as possible, then, we tend to keep the laws fixed.

But, according to Lewis, we do not invariably do so. There is at least one kind of exception. Suppose our world is deterministic, at least in the parts relevant to a given counterfactual, such as

> If the tree had not blown over, the roof would be intact.

(Actually, the tree blew over, destroying the roof.) If we are to keep the laws exactly as in the actual world, we have to change the entire past when considering a world in which the tree didn't blow over. We stay closer to the actual world, says Lewis, if we keep the past as it actually is and consider worlds in which there is a "small miracle" (relative to the laws of the actual world) and the tree doesn't blow over.

This is the beginning of a difficulty for Lewis. His single guiding principle is overall similarity to the actual world. If we must purchase past similarity at the price of a "small miracle," why can't we purchase future similarity at the price of another "small miracle"—a thunderbolt, say—that destroys the roof despite the fact that the tree didn't blow over? In the actual world, people were injured, the family impoverished and homeless. If overall similarity is what is at issue, we stay closer to the actual world at a world where the roof is destroyed by other means.

This is one example of many which seem to suggest that these two questions can get different answers: "What would have happened if *A*?" and, "What is true in the most similar *A*-world to the actual world?" Any instance of "If *A* had happened, things would have been very different" causes problems. One example from Kit Fine (1975): If Nixon had pressed the button in 1973, there would have been a nuclear holocaust. That may well be true. But in the most similar world to the actual world in which Nixon pressed the button in 1973, there was some compensating factor and no disaster; for the disaster makes for huge differences from the actual world.

Lewis's response was to refine his account of the respects of similarity that matter on the "standard resolution of vagueness" for counterfactuals, as follows:

(1) It is of the first importance to avoid big, widespread, diverse violations of law.
(2) It is of the second importance to maximize the spatio-temporal region throughout which perfect match of particular fact prevails.
(3) It is of the third importance to avoid even small, localized, simple violations of law.
(4) It is of little or no importance to secure approximate similarity of particular fact, even in matters which concern us greatly. (Lewis 1979: 47–8)

This cunningly gets rid of many counterexamples. Return to the tree. Let w_a be the actual world, where the tree blew over at *t*, destroying the roof. To maximise the area of perfect spatio-temporal match, we consider worlds exactly like w_a until shortly before *t*, when, if necessary, a small violation of the actual laws keeps the tree upright. If we stick to the laws thereafter, the roof is not destroyed. Consider worlds in which a second "small miracle" destroys the roof. This would

not restore perfect spatio-temporal match to the actual world. It would take a large, widespread miracle to do that, and this is banned by clause (1). Certainly such worlds are more similar in many matters of particular fact, but this counts for "little or nothing." We seem to get the right answer: the "most similar" worlds, by these criteria, align with what would have happened.

The theory has become more contrived. And problems remain. Consider clause (4). In parentheses, Lewis wonders whether approximate similarity of particular fact should count for little or nothing and says he isn't sure—different cases come out differently, and he gives two references. The first is to Tichy (1976), who gave this example: when Fred goes out and it's raining, he always takes his hat. When he goes out and it's not raining, it's a random 50–50 whether he takes his hat. Today it is raining and he takes his hat. What about "If it had not been raining, he would have taken his hat"? If similarity of particular fact counts for anything at all, we should accept this, but we don't. So perhaps similarity of particular fact counts for nothing. But other examples suggest otherwise. I am offered a bet on heads, and I decline. It is tossed anyway. It lands heads. If I had bet on heads, I would have won (Morgenbesser, reported by Slote 1978). More dramatically: I miss my plane. It crashes in a totally unpredictable way. If I had caught my plane, I would be dead (Edgington 2004). In these cases, the central ground for the truth of the counterfactual is a particular fact subsequent to the departure from spatio-temporal match: the coin landed heads; the plane crashed. These particular facts are not "of little or no importance" to the counterfactual.

There may be a way of specifying the relevant worlds, on the default reading of counterfactuals, but it seems unlikely that similarity will do all the work.

Lewis versus Stalnaker

When A is false but contingently so, Stalnaker's semantics assumes that there is always a unique closest A-world, and the counterfactual is true iff C is true at that world. Lewis rejects this. There may be ties for closest; and there may be no closest, when for any A-world there is a closer A-world, without end, as in "If I were taller . . ." or "If you were to get up earlier . . ." Take a possible world in which I am an inch taller. There is another in which I am half an inch taller, and so on ad infinitum (Lewis 1973: 20).

However, examples like this do not favour Lewis. We do use counterfactuals of this form: "If you had eaten more breakfast, you wouldn't be hungry now," "If you had asked me nearer the time, I would have been able to accept," "If my bags were less heavy, I would have walked there," where we clearly don't mean infinitesimally more breakfast, lighter bags, a split second later. Rather, a reasonable range is determined by context. Again, this tells against taking similarity too seriously.

Stalnaker (1981: 98) makes another point, involving Lewis's *might*-conditionals. Lewis defines "If it had been that A, it might have been that C," which he writes as $A, A\Diamond\!\!\rightarrow C$ as true iff some $A\&C$ world is as close as any $A\&\neg C$-world. Consider: I might have weighed at least one pound less than I actually do. But "If I weighed less than I do, I might have weighed at least one pound less" comes out false for Lewis, because there are closer worlds to actuality which satisfy the antecedent and not the consequent. This remains true whatever weight, however small, you substitute for "one pound."

Turn to the more interesting question of ties. For Lewis, if there is a tie for closest, $A > C$ is true iff C is true at *all* closest A-worlds, otherwise it is false. Thus for Lewis, the counterfactual conditional connective is a kind of necessity operator. Lewis calls it a "variably strict conditional": it involves quantification over worlds, but which worlds are quantified over depends on the antecedent. (Lewis symbolises the counterfactual $A\Box\!\!\rightarrow C$.) For Stalnaker, it is not a necessity operator. For Lewis, *would* is related to *must*; for Stalnaker, *would* is related to *will*.

For Lewis, a possible world is a vast universe, spatio-temporally isolated from ours, but every bit as real. Inevitably, it has an indefinite amount of detail. On this conception, it is hard to see how there could ever be a unique non-actual closest world in which, e.g., I struck the match—think of all the tiny differences in speed and hand movements involved. For Stalnaker (and most philosophers), possible worlds are ways this universe—this spatio-temporal object—might have been. There is no compelling reason to think of them as *absolutely complete* ways the world could be. And Stalnaker does not. He writes:

> I doubt that it is plausible to believe that there is, independently of context, a well-defined domain of absolutely maximally specific possible states of the world, but I do not think [I need] a commitment to such a domain. The alternative possibilities [I need] must be exclusive possibilities which are maximally specific relative to the distinctions that might be made in the context at hand. . . . The space might be partitioned differently in different contexts, and there might be no maximally fine partition.
>
> (Stalnaker 1981: 136)

Thus, only relevant differences are taken into account, and it is more feasible to think that there is, sometimes at least, a unique closest world in which I strike the match.

Stalnaker accepts that there can be ties for closest. In those cases, it is indeterminate which selection function is appropriate. If Bizet and Verdi were compatriots, it is indeterminate whether they would both be French, or both Italian. When 90% of the red balls have black spots, it is indeterminate whether, if you had picked a red ball, it would have had a black spot. It may be indeterminate whether, if you had had the operation, you would have been cured, and so on. For Lewis, these cases are all false: it's false that Bizet would have been Italian, false that Verdi would have been French, false that if you had picked a red ball it would have had a black spot, etc.

Opinions vary, but I think Stalnaker's verdict is superior. Although "indeterminate" is not a very helpful verdict, it is correct as far as it goes, and we should try to refine it. After all, 90% of the relevant pick-red-ball-worlds are black-spot-worlds, so the counterfactual is probable, and could be said to be close to clearly true. Similarly, for other uncertain counterfactuals like "She would have been cured if she had the operation," "The dog would have bitten me if I had approached," whose consequent may be true in most but not all relevant antecedent-worlds. They are not clearly true, but, contra Lewis, not clearly false.

Stalnaker's semantics validates the Law of Conditional Excluded Middle (CEM): $(A>C)\vee(A>\neg C)$: either the closest A-world is a C-world, or the closest A-world is a $\neg C$-world—though it might be indeterminate which, as in the Bizet-Verdi case; compare, of a borderline case: "It's either red or orange, but it's indeterminate which." Lewis rejects CEM. For him, when there are ties for closest, both disjuncts are false. For Lewis, $\neg(A>\neg C)$ iff $A\Diamond\!\!\rightarrow C$. For Stalnaker, $\neg(A>\neg C)$ iff $A>C$. Lewis (1973: 80) complains that for Stalnaker, *mights* collapse into *woulds*: $A>C$ iff $A\Diamond\!\!\rightarrow C$.

Stalnaker (1981: 98–9) replies that he does not accept that there is a special *might*-conditional connective; *might* applied to conditionals is just the same as *might* applied to other claims. He does not accept Lewis's relation between *woulds* and *mights*. He shows that Lewis's view gives some odd results. Consider:

A: Would Jane have accepted if she had been offered the job?
B: No, certainly not; but she might have accepted if she had been offered the job.

This reply seems defective, but according to Lewis, it is quite in order, for if some but not all relevant offer-worlds are accept-worlds, the conditional is definitely false. Similarly, consider this reply to the same question:

B: I believe she would have accepted, but she might not have accepted.

This reply sounds perfectly sensible, but for Lewis, it is not, because "She might not have accepted" contradicts the claim that she would have accepted, i.e. that it is true in all closest *A*-worlds. Once more, Lewis's *woulds* become clearly false too easily, when there is at least one relevant *A*-world in which *C* is false.

Final thoughts

I have suggested that in cases of uncertainty, we need a probability distribution over the relevant *A*-worlds. Also, agreeing with Stalnaker's verdict of indeterminacy in cases of ties, I suggested that we need the thought that some are closer to clearly true than others. (Compare: among the borderline-tall, some are taller than others.) Adams (1975) gave a probabilistic account of indicative conditionals and does extend it to counterfactuals (1975: ch. 4; see also Edgington 2008).

I have given an account of indeterminacy or vagueness in terms of degrees of closeness to clear truth (which I call verities), and I argued that these degrees have probabilistic structure (Edgington 1996). That is not to say that probabilities and verities are the same thing: there is no temptation to say that the probability that it will rain tomorrow is a degree of closeness to truth. But in the special case of counterfactual probabilities, they may coincide. For these uncertain counterfactuals can have genuine probabilities, yet there are no further facts that would determine a yes-no answer to the question of (e.g.) which ball you would have picked, if you had picked a red ball. For counterfactuals, I suggest, uncertainty *is* indeterminacy.

References

Adams, E. (1970) "Subjunctive and Indicative Conditionals," *Foundations of Language* 6, 89–94.
———. (1975) *The Logic of Conditionals*, Dordrecht: Reidel.
Anderson, A. R. (1951) "A Note on Subjunctive and Counterfactual Conditionals," *Analysis* 12, 35–8.
Bennett, J. (2003) *A Philosophical Guide to Conditionals*, Oxford: Clarendon Press.
Fine, K. (1975) "Critical Notice of David Lewis's *Counterfactuals*," *Mind* 84, 451–8.
Edgington, D. (1996) "Vagueness by Degrees," in Keefe, R. and Smith, P. (eds.), *Vagueness: A Reader*, Cambridge, MA: The MIT Press, 294–316.
———. (2004) "Counterfactuals and the Benefit of Hindsight," in Dowe, P. and Noordhof, P. (eds.), *Cause and Chance*, London: Routledge, 12–27.
———. (2008) "Counterfactuals," *Proceedings of the Aristotelian Society* CVIII, 1–21.
Gibbard, A. (1981) "Two Theories of Conditionals," in Harper, W., Stalnaker, R. and Pearce, G. (eds.), *Ifs*, Dordrecht: Reidel, 211–47.
Gibbard, A. and Harper, W. (1981) "Counterfactuals and Two Kinds of Expected Utility," in Harper, W., Stalnaker, R. and Pearce, G. (eds.), *Ifs*, Dordrecht: Reidel, 153–90.
Goodman, N. (1947) "The Problem of Counterfactual Conditionals," *Journal of Philosophy* 44, 113–28, reprinted in Goodman, N., *Fact, Fiction and Forecast* (1955), Indianapolis: Bobbs-Merrill. Page references in the text are to Goodman (1955).
Jackson, F. (1987) *Conditionals*, Oxford: Blackwell.
Kripke, S. (1963) "Semantic Considerations on Modal Logic," *Acta Philosophica Fennica* 16, 83–94.
Lewis, D. (1973a) *Counterfactuals*, Oxford: Blackwell.
——— (1973b) "Causation," *Journal of Philosophy* 70: 556–67, reprinted in D. Lewis, *Philsophical Papers*, vol. 2. Oxford: Oxford University Press (1986), 159–72. Page references to Lewis (1986).

——— (1976) "Probabilities of Conditionals and Conditional Probabilities," *Philosophical Review* 85, 297–315, reprinted in D. Lewis, *Philosophical Papers*, vol. 2. Oxford: Oxford University Press (1986), 133–56. Page references to Lewis (1986).

——— (1979) "Counterfactuals and Time's Arrow," *Nous* 13, 455–76, reprinted in D. Lewis, *Philosophical Papers*, vol. 2, Oxford: Oxford University Press (1986), 32–66. Page references to Lewis (1986).

Lycan, W. (2001) *Real Conditionals*, Oxford: Clarendon Press.

Nozick, R. (1981) *Philosophical Explanations*, Cambridge, MA: Harvard University Press.

Stalnaker, R. (1968) "A Theory of Conditionals," in *Studies in Logical Theory, American Philosophical Quarterly*, Monograph Series 2, 98–112, reprinted in Harper, Stalnaker and Pearce, *Ifs.* Dordrecht: Reidel, 41–55. Page references to Harper, Stalnaker and Pearce (1981).

———. (1981) "A Defense of Conditional Excluded Middle," in Harper, W., Stalnaker, R. and Pearce, G. (eds.), *Ifs*, Dordrecht: Reidel, 87–104.

———. (1981) "Indexical Belief," *Synthese* 49, reprinted in Stalnaker (1999) *Context and Content*, Oxford: Oxford University Press, 130–49. Page references to Stalnaker (1999).

———. (2005) "Conditional Propositions and Conditional Assertions," in *New Work on Modality*, MIT Working Papers in Linguistics and Philosophy, 51.

Slote, M. (1978) "Time in Counterfactuals," *Philosophical Review* 87, 3–27.

Strawson, P. (1986) "'If' and '⊃'," in Grandy, R. and Warner, R. (eds.), *Philosophical Grounds of Rationality*, Oxford: Clarendon Press, 229–42.

Tichy, P. (1976) "A Counterexample to the Stalnaker-Lewis Analysis of Counterfactuals," *Philosophical Studies* 29, 271–3.

van Fraassen, B. (1981) "Essences and Laws of Nature," in R. Healey (ed.), *Reduction, Time and Reality*, Cambridge: Cambridge University Press, 189–200.

Chapter 4

IMPOSSIBILITY AND IMPOSSIBLE WORLDS

Daniel Nolan

Introduction

Possible worlds have found many applications in contemporary philosophy: from theories of possibility and necessity, to accounts of conditionals, to theories of mental and linguistic content, to understanding supervenience relationships, to theories of properties and propositions, among many other applications. Almost as soon as possible worlds started to be used in formal theories in logic, philosophy of language, philosophy of mind, metaphysics, and elsewhere, theorists started to wonder whether impossible worlds should be postulated as well. To take just one example, possible-worlds theories of mental content associate sets of worlds with beliefs (or perhaps entire belief systems): the content of a belief is (or is represented by) the set of possible worlds where that belief is true. But what should we say about beliefs that cannot possibly be true – false logical or mathematical beliefs, for example, or beliefs in metaphysical impossibilities? It would be natural to represent these beliefs with sets of *impossible* worlds. If James thinks that 87 is a prime number, the set of worlds associated with his beliefs includes worlds where 87 is prime, for example. If Jane is undecided about the principle of excluded middle, her belief worlds should include some at which the principle is correct and some where it is incorrect. And so on.

There are some direct arguments that we should take impossibilities and impossible worlds seriously: we seem to talk about, count, and compare impossible scenarios just as we talk about, count, and compare possible ones, and if that talk is taken at face value, then it commits us to impossible scenarios. But most of the arguments for accepting that there are impossible worlds involve pointing to the value of theories that postulate them. If theories postulating impossible worlds offer us satisfactory explanations and understanding, especially of otherwise puzzling phenomena, that gives us reason to endorse the existence of impossible worlds. (Not indefeasible reasons: better theories that reject the existence of impossible worlds may come along, after all.) This methodology is not uncontroversial, of course, and some might think that success of theories employing impossible worlds would only be evidence that they are useful heuristics or have some other lesser value. Debates about impossible worlds have typically centred on the quality of the theories that employ them, so this chapter will begin with a survey of some of the more important applications impossible worlds have figured in.

In this chapter, I will first introduce some of the ways theories of impossible worlds have been used to deal with problems that arise for more standard possible-worlds analyses in

philosophy. Then I will discuss some questions about the nature and extent of impossible worlds. Lastly, I will turn to some important objections to the project of employing impossible worlds in our philosophical theorizing.

Some uses for impossible worlds

The main reason philosophers have been interested in impossible worlds is the potential that they will play a useful role in good theories of other phenomena. One fruitful strategy for finding applications for impossible worlds is to look at some of the places possible worlds have been employed in philosophical theories, and to see whether limitations of possible-worlds approaches can be overcome by expanding the theories to encompass impossible worlds as well. Let us turn to some of the main uses (for more detailed presentations of a range of uses of impossible worlds, see Berto and Jago 2018; Nolan 2013).

One famous application of the theory of possible worlds has been in offering a theory of *counterfactuals*: conditional sentences such as "If the conveyor had not short-circuited, the factory would not have burned down". While there is controversy about what counts as a counterfactual conditional, paradigm cases include conditionals about what *would* have happened, had things gone otherwise than they did. Famously, just determining the truth-values of antecedents and consequents of these conditionals is not always enough to see whether they are true or false. A popular approach to counterfactuals, pioneered by Stalnaker (1968) and Lewis (1973), is to treat their truth conditions as involving other possible worlds: roughly, a counterfactual is *true* provided that in the nearest possible worlds where the antecedent is true, the consequent is true as well. (In the possibilities most relevantly like ours where the conveyor did not short-circuit, the factory did not burn down.)

Restricting attention to *possible* worlds gives intuitively the wrong result when we are asked to consider counterfactuals about what would have happened if something *impossible* had happened. Stalnaker's and Lewis's theories both predict that whenever the antecedent of a counterfactual is impossible, the counterfactual as a whole is true. Thomas Hobbes, famously, devoted time to discovering a method of squaring the circle (something we now know is impossible). "If Hobbes had squared the circle, he would have made a famous mathematical discovery" seems true (and we might say so if we are trying to explain Hobbes's motivation for his attempt). But "If Hobbes had squared the circle, he would have become a werewolf" seems false: even skilled geometers do not turn into werewolves. Counterfactuals with impossible antecedents ('counterpossibles') can still be handled in something like a Lewis-Stalnaker framework if impossible worlds are used as well, and a counterfactual is true provided that in the nearest worlds (possible or impossible) where the antecedent is true, the consequent is true as well. Hobbes-squaring-the-circle worlds are impossible, but the ones where he does so but does not turn into a werewolf are more relevantly similar to our own than ones in which he squares the circle and also becomes a werewolf. (Routley 1989 is the first I know of to extend a Lewis-Stalnaker-style semantics to impossible worlds, and this approach to counterpossibles has been defended by Nolan 1997, Vander Laan 2004, and Brogaard and Salerno 2013, among many others.)

Perhaps the earliest systematic use for impossible worlds was in providing the semantics for various logical systems: modal logics, and then logics representing the language of psychological attitude attributions (talk about belief, desire, etc.). The earliest use of worlds that are in some sense "impossible" is in possible-worlds semantics for modal logics that employ an "accessibility" relation: in models where some worlds are not accessible from the actual world, those worlds are not possible "from" the actual world (though those worlds are still standardly labelled "possible worlds"). Accessibility relations on possible worlds seem to have made their first appearance in

journal publications in Prior (1962) and Kripke (1963), though see Copeland (2002) for a fascinating pre-history of the notion.

The earliest published use of impossible worlds in a stronger sense I have been able to find is Kripke (1965), who employed "non-normal worlds" to provide the semantics for a range of modal logical systems including the Lewis systems S2 and S3. Logics of belief and desire using impossible worlds came soon after: Cresswell (1970) used "non-classical worlds" to represent differences in belief attributions, and Hintikka (1975) employed "impossible possible worlds" for a similar purpose (see also Cresswell 1973; Rantala 1982 for some other early contributions to this program). One feature of systems such as Cresswell (1970) that has become standard is that the model-theoretic definition of logical validity requires truth preservation only at all *possible* worlds in all models: that is, there is no *possible* world in any model where the premises are all true but the conclusion is not true. This feature allows us to keep a robust notion of logical consequence, while not requiring that all impossible worlds be logically well-behaved.

As far as a pure model-theory goes, so-called "worlds" need not have much of a connection to ways things cannot be, but philosophical motivations for these logics can be given in terms of genuine impossibilities, via explaining *why* the formal models are appealing ones for capturing the ideas behind the logics. An example of using the semantics of inaccessible "possible" worlds not just as a piece of formalism but as being about possible and impossible worlds is Salmon (1984), employing principles about worlds to argue against the modal logic S4.

Probably the main use of possible worlds in philosophy has been in theories of mental and linguistic representation. At the beginning of this chapter, an example of using impossible worlds to improve on a possible-worlds account of belief contents was given. Impossible worlds also play a fruitful role in improving on the "possible worlds semantics" tradition in semantics for natural languages (see Partee 1989 for a number of advantages of using possible worlds in semantics). Using only possible worlds has some limitations: it is hard to avoid running together distinct claims that are necessarily true, or distinct claims that are impossible. But we want to explain differences in meanings between these claims when explaining the meaning of mathematical language or philosophical language, for example. See Nolan (2013, pp 364–366) for a more detailed discussion of the use of impossible worlds in a theory of linguistic and mental representation.

Finally, impossible worlds seem to have a role to play in the metaphysics of the non-representational world. Possible worlds have been used in a wide variety of metaphysical theories from the 1960s: Montague (1969) is an important founding paper in this trend, and it appears throughout the metaphysical writings of David Lewis. Many of the same areas can benefit from theories employing impossible worlds. Apart from theories of counterfactuals discussed earlier in this chapter, metaphysical topics recently treated with theories of impossible worlds include explanation in general and metaphysical explanation in particular (Kment 2014), theories of essence (Brogaard and Salerno 2013, pp. 646–648), and the nature of omissions (Bernstein 2016). Nolan (2014) lists a number of other areas that seem to call for metaphysical treatments using resources more fine-grained than possible worlds: though of course there are other resources available besides analyses in terms of impossible worlds.

The nature of impossible worlds

Before deciding whether to adopt a commitment to the existence of impossible worlds, it is reasonable to ask what impossible worlds are supposed to be, and what features they are supposed to have. There is little agreement about either of these questions, though the disagreements about what kinds of things impossible worlds are do have parallels with better-known debates about what *possible* worlds would be.

Perhaps the most common type of account of possible worlds takes them to be abstract objects of some sort or other. One might take possible worlds to be sets of sentences, or sets of propositions, or each world to be a single, maximal proposition, or each world to be a maximal property, or maximal state of affairs, or possible worlds might even be sui generis abstract entities. Such views require a specification of what it is for a sentence (or claim, proposition, etc.) to be true according to one of these abstract objects: notice that a claim can be true *according to* a world without in any sense being in fact true. The family of views that treat possible worlds as abstract objects are often labelled "abstractionists". It is natural for abstractionists about possible worlds to be the same sort of abstractionists about impossible worlds. As well as sets of sentences that can all be true together, there are sets of sentences that cannot be all true together. As well as maximal possible propositions, there are propositions that cannot be true, including maximal propositions that cannot be true. As well as maximal states of affairs that can be instantiated, there seems to be no obvious bar to supposing there are maximal states of affairs that cannot be instantiated. And so on. There are sometimes metaphysical puzzles about the details, but the general shape of abstractionist theories of impossible worlds is clear (abstractionist theories of impossible worlds include those presented by Mares 1997, Vander Laan 1997, and arguably Zalta 1997, while a more recent abstractionist approach is defended in Jago 2014).

David Lewis famously claimed that possible worlds were objects of the same kind as our physical universe (Lewis 1986), and while this "concretism" remains a minority view of possible worlds, it retains some influential defenders. Concretism about impossible worlds is an even less popular view, since on the face of it they would be things that cannot exist, but do exist, and the inconsistent ones at least would be things "whereof you speak truly by contradicting yourself" (Lewis 1986, pp. 1, 7 n. 3). Despite these hurdles, several forms of concretism, which treat impossible worlds as in some sense the same kind of thing as our own cosmos, have been defended. Yagisawa (2009) is the most famous defence of this view, though Kiourti (2010) and Vacek (2013) each defend a different development of a concretist view about impossible worlds.

Some theorists have suggested that we conceive of possible worlds and impossible worlds as being different kinds of entities. The most popular "mixed option" is to treat possible worlds as concrete, but impossible worlds as abstract objects that are in some sense constructions from possible worlds. Restall (1997) was an early suggestion of this combination, but it has recently been advocated by Berto (2010), who argues that it combines benefits of concrete realism about possible worlds, such as a reductionist account of modality, with the benefits of postulating impossible worlds.

There are other options for theories of impossible worlds than treating them as existing abstract objects or existing concrete universes. One is to treat them as non-existent objects of one sort or another: see, for example, Priest (2005). There are also approaches that refuse to accept that impossible worlds are anything at all, existing or non-existing. One could treat talk about impossible worlds as a *convenient fiction* for regimenting our talk about what cannot happen, embracing either a fictionalism or instrumentalism about impossible worlds. Or one could offer a *paraphrase* of talk of impossible worlds into a theory that was not committed to such things.

Beyond questions about what kind of entity impossible worlds are, there are other questions about impossible worlds to be sorted out. One of the main choice points is whether or not to treat impossible worlds as closed under some logic: that is, whether there is some interesting logic L, and associated consequence relation $\models_L$, such that whenever a set of propositions Σ is true according to a world, and there is a proposition A such that $\Sigma \models_L A$, A will also be true according to that world. When impossible worlds are the inaccessible worlds of Kripke-style semantics for modal logics, or the non-normal worlds of semantic treatments of systems like S2

or S3, the non-modal formulas true according to the worlds are closed under classical (propositional or predicate) logic.

For some other applications, impossible worlds are all treated as if they are closed under classical logic, or whatever other logic that the actual world is supposed to be closed under: some theorists accept the existence only of metaphysically impossible worlds that are still logically possible (see e.g. Kment 2014) These theorists see the use for impossible worlds where e.g. metaphysical necessities do not obtain, or some facts that flow from the essence of objects obtain but others do not. Allowing for metaphysically impossible worlds permits e.g. non-trivial counterfactuals about what would be the case were the metaphysical facts different, and for representing ignorance of logically ideal agents who are ignorant or mistaken about some metaphysical necessities.

More radical still than these options are theories according to which all worlds are closed under an interesting and substantial logic, but that logic is considerably weaker than the (strongest) logic that the actual world is closed under. For example, the worlds might be closed only under First Degree Entailment, which ensures e.g. that conjunctions are true according to a world when their conjuncts are, but does not ensure that there are no contradictions true according to a world, or that there are no failures of excluded middle at such worlds. This approach is called the "Australasian Plan" by Priest (1997) and, following him, Berto and Jago (2018). The Australasian plan has some advantages; for example, that the truth-value, at any world, of many truth-functional compounds can be defined in terms of the truth-values at those worlds of the propositional constituents of those compounds. It also would enable us to rely on some standard logical principles, in full generality, for working out truth according to worlds that we might be considering when evaluating counterfactuals or belief contexts. Priest (1992) is one influential philosophical motivation for an Australasian-plan approach to impossible worlds.

Most radical of all is the option of treating impossible worlds as not, in general, being closed under any logical consequence relation at all (apart from identity: when B is true according to a world, B will be true according to that world). Approaches like this typically hold that logical consequence is connected to truth-preservation at possible worlds rather than truth-preservation at all worlds: recognizing a wide range of impossibilities does not automatically require a revisionary conception of which worlds are logically possible. Naturally, even if impossible worlds in general are not all closed under any particular logical consequence relation, individual worlds might still be: indeed, there can be individual impossible worlds closed under classical logic if the only impossibilities there are e.g. metaphysical ones. This approach was labelled the "American plan" for impossible worlds by Priest (1997), on the grounds that several of its defenders are American (Zalta 1997; Vander Laan 1997) or "honorary Americans" (Nolan 1997). One argument put forward for this generosity with impossible worlds in Nolan (1997, p. 547) is that otherwise a theory of impossible worlds looks like an awkward halfway house, with some impossibilities corresponding to impossible worlds and some (apparent) impossibilities not even being found among the *impossible* worlds (e.g. cases where A is true according to an impossible world, B also being true according to that impossible world, but $A\&B$ failing to obtain at that world).

The dispute between Australasian-plan approaches and American-plan approaches is likely to be settled by looking more carefully at what is required for good theories using impossible worlds. I think that once the full range of applications are tackled, sufficiently anarchic impossible worlds will need to be appealed to such that no interesting logical consequence relation will be found that they are all closed under. Champions of the Australasian plan will no doubt expect that this degree of freedom in our theorizing about impossible worlds is unnecessary.

Interestingly, Graham Priest himself seems to have shifted camps: his remarks in Priest (1997) suggest his sympathy is with the Australasians, but Priest (2005) includes "open worlds" that need not be closed under any substantial consequence relation, in order to handle the full range of phenomena involving intentional attitudes like belief.

Objections to postulating impossible worlds

Sustained philosophical attention to impossible worlds is a relatively recent phenomenon, and as might have been predicted, early discussions have tended to take place by people trying to make the case for postulating them or using them in one philosophical inquiry or another. However, some criticisms of the project of employing impossible worlds, either for specific purposes or in general, have appeared in the literature. In this section I will focus on two fairly general criticisms of the use of possible worlds, rather than criticisms of specific applications; though see Williamson (2018) for a recent criticism of the use of impossible worlds for a specific application: the theory of counterpossible conditionals.

Robert Stalnaker's influential "Impossibilities" (Stalnaker 2003) contains many concerns about impossible worlds, and since it is in the form of a dialogue, it is not clear which of them Stalnaker is advancing or even thinks can be made to work. I will focus here on just one of the important concerns that Stalnaker's dialogue suggests (though perhaps he would not say it in the way I am about to).

The concern is whether any argument for postulating "impossible worlds" would be self-defeating. Suppose we showed that for a range of tasks for which we postulated possible worlds, we need more worlds than e.g. Stalnaker would countenance: suppose we needed to postulate worlds that were not closed under classical consequence to provide a theory of belief, for example, or worlds where contradictions were true to handle some counterfactuals. Why suppose such worlds represent *impossibilities*, rather than conclude that we were too restrictive about what *possibilities* we recognize? For example, if beliefs serve to distinguish between possibilities, and we have to genuinely distinguish in belief between worlds where 123 is prime and 123 is composite, wouldn't this just show that both of those are genuine possibilities after all? (Compare Mortensen 1989, who accepts that every proposition is possible, roughly on the grounds that we can treat any proposition in the way standard theories hold we can treat only possible propositions.) Stalnaker suggests that arguments for impossible-world theories are self-defeating: at best, they are arguments for using only possible worlds, but revising how generous our theory of possible worlds should be. (At least if I have interpreted Stalnaker correctly here: I take the exchange on pp. 62–67 to suggest this, particularly the remarks about "logical" and "illogical" space on pp. 62–63.)

A defender of impossible worlds should provide a response to this challenge (whether or not it was what Stalnaker had in mind). Why are the new worlds postulated *impossible* in any interesting sense, if they play the same theoretical roles as possible worlds? Why not just count them as more *possible* worlds, and say that we previously underestimated the extent of what is possible? One immediate response to this challenge could be to point out relevant differences in how impossible worlds are deployed. If impossible worlds play some but not all the roles possible worlds are supposed to play, it is worth keeping a distinction in place (however it is labelled). Possible worlds, for example, play a distinctive role in connection with the (non-epistemic, non-deontic) modal operators: it is not enough that there be an *impossible* world according to which a proposition p obtained to ensure that p was genuinely possible. Perhaps we can introduce new operators, M' and L', perhaps, so that $M'(p)$ is true whenever p is true according to some world, possible or impossible, and $L'(p)$ is true when and only when p is true according to every

possible and impossible world. On the face of it, though, that would no more correspond to a genuine sense of possibility and necessity than introducing new operators $M''(p)$ and $L''(p)$ so that $M''(p)$ is true no matter what p is, and $L''(p)$ is false no matter what p is, would show that all propositions were really possible! Likewise, in many impossible-world theories, logical validity is a matter of truth-preservation at *possible* worlds, not all worlds.

A second response to the challenge that goes a little deeper would be to defend an understanding of possibility and necessity so that we have good reason to treat something as impossible even if it is thinkable or representable in language, if it makes sense to consider it as a supposition for counterfactual reasoning, or even if it plays some of the other roles traditionally assigned to possibilities. Many substantial theories of possibility and necessity could play this role. Perhaps a necessary condition to be possible is that a proposition not logically imply a contradiction, for example. That would require some argument, but if we made it plausible, then if we could also show that some contradictions are thinkable (and different contradictions differently thinkable), are supposable for non-trivial counterfactual reasoning, make a non-trivializing contribution to a fiction, and so on, then that combination would be a principled position which provides a response to the Stalnakerian concern about collapse. Or perhaps it is a necessary condition on a claim's being possibly true that it does not violate any analytic rules: a philosophical position which defended that claim and also showed how some statements which do violate those rules can be believed, or non-trivially supposed, or non-trivially contribute to a fiction, etc., could be used to argue that not every apparent impossibility is just a possibility in wolf's clothing. In general, there are many ways that a theory of the difference between possibility and impossibility can draw the line so that the distinction does not collapse, even if impossible worlds do some of the work traditionally thought to be the exclusive province of possible worlds. But a general theory of the divide between possibility and impossibility owes us some such answer to this Stalnakerian challenge.

Another interesting challenge to the use of impossible worlds for philosophical theorizing has been offered by Bjerring and Schwarz (2017). Their concern is based around the question of how fine-grained the distinctions need to be for impossible worlds to perform the theoretical tasks those worlds are often assigned, especially in the philosophy of mind and language. Impossible worlds, if they are to add anything, must allow the drawing of distinctions that are not marked as differences among possible worlds: e.g. two sentences true at all the same possible worlds will have to be true at different impossible worlds, if they are to be associated with different contents (e.g. if I am to be able to believe that $2 + 2 = 4$ without believing that there is a square root of -1). But then when we look at how much contents seem to come apart, very few of the sentences true at the same possible worlds will be true at the same impossible worlds. (There seem to be *very* few other mathematical propositions I must believe if I believe that $2 + 2 = 4$.) At the limit, you might think that distinct sentences must always be associated with different contents. But if contents are distinguished this finely, then impossible worlds seem to lose a lot of their explanatory power. For example, if we model worlds in full generality as associating arbitrary sets of sentences with the truth-value "true", with the complement of those sets treated as not true at the relevant world, then our model of the meaning of sentences consists just of representing the meaning of a sentence as being true whenever a set of sentences it belongs to is the set of true sentences: not a very illuminating analysis of truth conditions!

Bjerring and Schwarz also argue that it is more difficult than it appears to motivate an "intermediate" position on mental content: once you start relaxing constraints on belief worlds beyond possible ones, it is hard to keep belief worlds closed under any very interesting logic. So one line of response defenders of impossible worlds have engaged in is to look for principled points to restrict the flexibility of the impossible worlds needed e.g. for mental content (see Jago 2014,

especially ch. 8). While it makes sense to limit how fine-grained impossible worlds are for some purposes, my own suspicion is that to handle all the relevant phenomena, we will occasionally need impossible worlds that draw very fine distinctions indeed.

One place we might need very fine distinctions is in handling the beliefs of very unusual individuals, such as those in the grip of odd religious or metaphysical theories. Perhaps normally when someone believes a conjunction, they believe the conjuncts. But someone convinced that a god is beyond logic might believe e.g. that their god both exists and does not exist, in a way that it does not follow that their god does not exist: and so does not believe that their god does not exist.

So I think the spirit of Bjerring and Schwarz's challenge will be difficult to answer. There are plausible purposes for which a theory of impossible worlds will need to allow that very similar sentences, true at all the same possible worlds and many of the same impossible worlds, nevertheless differ in whether they are true according to some particularly impossible and unusual worlds. A defender of impossible worlds who accepts this owes us an account of how such a system of impossible worlds can still play a non-trivial explanatory role. I am optimistic that this can be done, but an adequate response to this general challenge will have to be left to another occasion.

Conclusion

Employing possible worlds in our theorizing runs into significant limitations when we need to distinguish between necessary equivalents. Impossible worlds offer us a straightforward way to keep the benefits of theories that employ possible worlds without running into the problems that stem from the fact that our theories seem to need distinctions that do not correspond to differences between possible worlds. Employing impossible worlds is not the only way to tackle the problem of doing justice to distinctions that go beyond modal differences: but at the current state of philosophical development, few if any rivals have been developed that offer a unified response to all the puzzles that impossible worlds can help with. Though see Duží et al. 2010 for an introduction to one systematic rival program. Work in the relevance/relevant logic tradition also offers many treatments of phenomena that trouble theories built with possible-worlds resources, but the model theory behind most relevance-logical treatments uses points of evaluation that behave just like Australasian-plan impossible worlds, so it seems better to classify relevance/relevant approaches as implementations of impossible-worlds approaches rather than as rivals to using impossible worlds.)

Theorists who employ impossible worlds disagree with each other on many questions about impossible worlds: their extent, their nature, and how best to employ them in theories of other philosophical phenomena. If past philosophical developments are any guide, these internal debates will no doubt continue. However they are resolved, the use of impossible worlds in philosophical theorizing is rapidly becoming as well entrenched as invoking possible worlds. The expansion of philosophical focus from the actual to include the merely possible will continue into appreciation for the impossible as well.

References

Bernstein, S. (2016) "Omission Impossible," *Philosophical Studies* 173.10: 2575–2589.

Berto, F. (2010) "Impossible Worlds and Propositions: Against the Parity Thesis," *Philosophical Quarterly* 60.240: 471–486.

Berto, F. and Jago, M. (2018) "Impossible Worlds," *Stanford Encyclopedia of Philosophy*. [Online]. Retrieved on 22 September 2020 from: http://plato.stanford.edu/entries/impossible-worlds/.

Bjerring, J.C. and Schwarz, W. (2017) "Granularity Problems," *Philosophical Quarterly* 67.266: 22–37.

Brogaard, B. and Salerno, J. (2013) "Remarks on Counterpossibles," *Synthese* 190: 639–660.

Copeland, B.J. (2002) "The Genesis of Possible Worlds Semantics," *Journal of Philosophical Logic* 31: 99–137.
Cresswell, M. (1970) "Classical Intensional Logics," *Theoria* 36.3: 347–372.
Cresswell, M. (1973) *Logics and Languages*, London: Methuen.
Duží, M., Jespersen, B., and Materna, P. (2010) *Procedural Semantics for Hyperintensional Logic*, Dordrecht: Springer.
Hintikka, J. (1975) "Impossible Possible Worlds Vindicated," *Journal of Philosophical Logic* 4.4: 475–484.
Jago, M. (2014) *The Impossible: An Essay on Hyperintensionality*, Oxford: Oxford University Press.
Kiourti, I. (2010) *Real Impossible Worlds: The Bounds of Possibility*, PhD thesis, University of St Andrews. [Online]. Retrieved on 21 February 2017 from: https://research-repository.st-andrews.ac.uk/bitstream/handle/10023/924/IraKiourtiPhDThesis.PDF.
Kment, B. (2014) *Modality and Explanatory Reasoning*, Oxford: Oxford University Press.
Kripke, S. (1963) "Semantical Analysis of Modal Logic I: Normal Modal Propositional Calculi," *Zeitschrift fur mathematische Logik und Grundlagen der Mathematik* 9.5–6: 67–96.
Kripke, S. (1965) "Semantical Analysis of Modal Logic II: Non-Normal Propositional Calculi," in Addison, J.W., Tarski, A., and Henkin, L. (eds.) *The Theory of Models: Proceedings of the 1963 International Symposium at Berkeley*, North Holland, pp. 206–220.
Lewis, D. (1973) *Counterfactuals*, Oxford: Blackwell.
Lewis, D. (1986) *On the Plurality of Worlds*, Oxford: Blackwell.
Mares, E. (1997) "Who's Afraid of Impossible Worlds?" *Notre Dame Journal of Formal Logic* 38.4: 516–526.
Montague, R. (1969) "On the Nature of Certain Philosophical Entities," *The Monist* 53.2: 159–194.
Mortensen, C. (1989) "Anything Is Possible," *Erkenntnis* 30: 319–337.
Nolan, D. (1997). "Impossible Worlds: A Modest Approach," *Notre Dame Journal for Formal Logic* 38.4: 535–572.
Nolan, D. (2013) "Impossible Worlds," *Philosophy Compass* 8.4: 360–372.
Nolan, D. (2014) "Hyperintensional Metaphysics," *Philosophical Studies* 171: 149–160.
Partee, B. (1989) "Possible Worlds in Model-Theoretic Semantics: A Linguistic Perspective," in Allen, S. (ed.) *Possible Worlds in Humanities, Arts and Sciences. Proceedings of Nobel Symposium 65*, New York: Walter de Gruyter, pp. 93–123.
Priest, G. (1992) "What Is a Non-Normal World?" *Logique et Analyse* 35: 291–302.
Priest, G. (1997) "Editor's Introduction," *Notre Dame Journal of Formal Logic* 38.4: 481–487.
Priest, G. (2005) *Towards Non-Being: The Logic and Metaphysics of Intentionality*, Oxford: Oxford University Press.
Prior, A. (1962) "Possible Worlds," *Philosophical Quarterly* 12: 36–43.
Rantala, V. (1982) "Impossible Worlds Semantics and Logical Omniscience," *Acta Philosophica Fennica* 35: 106–115.
Restall, G. (1997) "Ways Things Can't Be," *Notre Dame Journal of Formal Logic* 38.4: 583–596.
Routley, R. (1989) "Philosophical and Linguistic Inroads: Multiply Intensional Relevant Logics," in Norman, J. and Sylvan, R. (eds.) *Directions in Relevant Logic*, Dordrecht: Kluwer, pp. 269–304.
Salmon, N. (1984) "Impossible Worlds," *Analysis* 44.3: 114–117.
Stalnaker, R. (1968) "A Theory of Conditionals," in Rescher, N. (ed.) *Studies in Logical Theory*, Oxford: Blackwell, pp. 98–112.
Stalnaker, R. (2003) "Impossibilities," in Stalnaker, R. (ed.) *Ways a World Might Be: Metaphysical and Anti-Metaphysical Essays*, Oxford: Oxford University Press, pp. 55–67.
Vacek, M. (2013) "Concrete Impossible Worlds," *Filozofia* 68.6: 523–529.
Vander Laan, D. (1997) "The Ontology of Impossible Worlds," *Notre Dame Journal of Formal Logic* 38.4: 597–620.
Vander Laan, D. (2004) "Counterpossibles and Similarity," in Jackson, F. and Priest, G. (eds.) *Lewisian Themes*, Oxford: Oxford University Press, pp. 358–375.
Williamson, T. (2018) "Counterpossibles," *Topoi*. 37.3: 357–368.
Yagisawa, T. (2009) *Worlds and Individuals, Possible and Otherwise*, Oxford: Oxford University Press.
Zalta, E. (1997) "A Classically-Based Theory of Impossible Worlds," *Notre Dame Journal of Formal Logic* 38.4: 640–660.

Chapter 5

THE ORIGINS OF LOGICAL SPACE

Brian Leftow

The phrase "logical space" stems from Wittgenstein's *Tractatus* (e.g. Wittgenstein 1961, 1.13). The "space" of a modality is its class of possible worlds. Every truth which has that kind of necessity (physical, absolute, whatever) is true in all of this modality's space. Every contingent proposition partitions the space: in part it is true, in part it is false. Each possible world hosts a unique set of contingent truths: worlds in which the same contingent propositions are true would be, simply, the same world. Thus each possible world is or occupies a distinct part of this space. For present purposes, logical space is the space of absolutely, metaphysically or broadly logically possible worlds. To ask about its origins, then, is to ask whether there is an explanation for there being the absolute possibilities there are. I now argue the following: dissatisfaction with the "no origins" accounts of logical space on offer should lead us to at least consider "origins" accounts. Dissatisfaction with *those* should lead us to consider whether there might be a view that combined the benefits and avoided the drawbacks of both extant "origins" and extant "no origins" views. There is such a view, and I describe it here.

5.1 Does logical space have an origin at all?

Realists about absolute modality—those who think it independent of human thought, language, etc.—tend to think that it has no explanation. For most of them, the things that account for possibility are just *there*, just an ultimate framework within which actuality finds it way. The framework may consist of Platonic abstract entities (e.g. Plantinga 1974), Meinongian existence-neutral objects, Lewis's concrete universes (Lewis 1986), or Williamson's "non-concrete" objects (e.g. Williamson 2013). But whatever it is, nothing accounts for its existing or being as it is. The main realist exceptions are metaphysicians who follow A.N. Prior (e.g. Prior 1960) in holding that there are no singular possibilities for non-existent individuals. On this view, before Lincoln existed, there were no possibilities for Lincoln: it was not possible that Lincoln exist or be shot in Ford's Theatre. There were only general possibilities for individuals qualitatively just like him: it was possible that someone tall, bearded, presidential, etc., be shot in Ford's Theatre. Only once Lincoln existed was anything possible for *him*. If this is how things are, the most that is "just there" necessarily are general possible world *types*. Full possible worlds, containing possibilities for individuals, arise only given a population of actual individuals. They originate in the intersection between the (say) Platonic framework and concrete actuality's population, and so logical

space is constantly replaced: each new individual spawns an entire new space, since even if a world does not contain that individual, it must be newly related to worlds that do contain it. Just how this happens is not explained.

5.2 A case for an origin

There are reasons to be dissatisfied with "no origins" realism, at least in the main forms so far developed. Williamson requires us to believe that we existed before our parents conceived us, being then not really distinguishable from abstract objects. Lewis's version, on which possible worlds are concrete universes as full-bloodedly existent as our own, is a paradigm of ontological bloat. Lewis insists that it is actually parsimonious, since it posits only a lot more stuff of the same basic kinds as our own universe rather than adding new ontologically significant kinds like attributes (Lewis 1973, 87) and lets us reduce our number of ontological primitives (Lewis 1986, 4). But intuitively, tokens matter. If an ontology tells us that an infinity of monkeys exist, something's wrong. Lewis tells us in effect that we can have good *a priori* reason to believe that blue monkeys exist, since we know *a priori* that they are possible, and according to Lewis have good *a priori* reason to accept his views, on which whatever is possible, exists. And it is hard not to think that other existent universes are extensions of what is actual rather than alternatives to it. Meinongian theories hold that every item that ever could exist, and perhaps also every item that couldn't, "is" an object, existent or non-existent. Such theories have great difficulty making sense of their basic machinery—what it means to say that non-existent objects "are there," and the meaning of the Meinongian copula that predicates properties of these objects. They bloat the ontology as much as Lewis does: for every Lewis object existent in another universe, there is a non-existent Meinongian object which is just as much part of the ontology. Further, they bloat it beyond Lewis: Lewis "reduces" some objects to others, e.g. construing propositions as just sets of worlds (Lewis 1983, 53–5). Meinongians generally have no such reductive ambitions. For them, generally, a non-existent proposition is a proposition, not a set of anything, and so propositions are part of their ontology, though not part of the inventory of existents. And Meinongians tell a bizarre story about you: before you existed, you were there. Either you weren't three-dimensional at all, or you were three-dimensional but somehow did not occupy any space. Do you find this plausible?

This leaves Platonist theories, on which either worlds are abstract propositions and the actual world is a true proposition, or worlds are abstract properties and the actual world is an exemplified property, or worlds are states of affairs and the actual world is the totality of real facts. Here again, bloat is a worry: nominalists will be uneasy with such a massive abstract realm. Further, there are worries about whether such abstracta are really cut out for this role. Propositions seem the wrong sort of thing to be possibilities or possibility-makers. If Fido is a dog, that makes it true that Fido is a dog. Why reverse this in the modal case, and say that its being true that possibly Fido is a dog makes it the case that possibly Fido is a dog? Modal propositions should represent modal facts independent of themselves, as propositions do generally. (There are sentences like "this proposition is true" or "this proposition is false," but it is not clear that they express propositions. For if they do, the propositions they express are like non-well-founded sets—just as a non-well-founded set supposedly has itself as one of its elements, these propositions would have themselves as constituents. Both would be too much like being a proper part of oneself for comfort. So just as it seems to most that there could not be a non-well-founded set, it should seem to most that there could not be propositions these sentences express.)

There is a parallel worry about Platonic properties. If Fido is noisy, that makes it the case that Fido's essence is coexemplified with noisiness. Why reverse this in the modal case and say that

Fido's essence's ability to be coexemplified with noisiness is what makes it the case that Fido can be noisy? The facts about what properties can be co-exemplified with what should derive from the facts about what things can co-exemplify properties, not vice versa. As to states of affairs, we can do without them in all but the modal case. If Fido exemplifies noisiness, we do not need an additional entity, the state of affairs that Fido is noisy, to make it true that Fido is noisy. So if introduced in a theory of possible worlds, states of affairs are just *sui generis* entities with but one specialized purpose: ripe, low-hanging fruit for ontological elimination. I would break a lance for the further claim that they are *sui generis* mysteries: no one really understands what they are supposed to be.

Perhaps these are not all possible accounts of the nature of a "no origins" logical space, but they are certainly the most prominent so far offered. I suggest that they won't do. Their flaws seem likely to replicate if new theories arise: other abstract-entity worlds will also be things that should represent in some way modal facts established elsewhere, and other concrete sorts of world (Meinongian worlds are concrete non-existents) will produce massive bloat. If we can't come up with an adequate account of a "no origins" space, we should consider whether assigning some origin to logical space provides a better overall theory.

5.3 Current "origins" views

The main theories on offer that do assign an origin to logical space are anti-realist about absolute or metaphysical modality. Consider e.g. conventionalism: suppose we say that to be absolutely possible is just to be describable in our language without infringing any relevant convention. Before English existed, in one sense, nothing was describable in English. In another, things were: in 4000 BC, that Fido was a dog was the kind of thing one could describe in English. But if we take that second sense seriously in metaphysics, it's just a tiny step to saying that Fido would have been describable in English even if no one ever created English. Say that merely possible languages count, and actual linguistic conventions don't matter at all: we've moved from conventionalism into some theory where the key question is what languages are possible, and the ontology is of possible languages. This is not conventionalism. So for conventionalists, before English, nothing was describable in English—and before language, there was no logical space. We brought logical space to be when we began to have languages, and it has evolved over time; what is possible (now) is just what we can express (now) without infringing on convention. Suppose instead that we say that to be absolutely possible is just to be conceivable. To make this an "origin" theory, we should avoid saying that things were conceivable before there were conceivers to conceive them. If we do avoid this, logical space originated when conceivers came to be, and evolves with their conceiving powers.

Conventionalism and conceptualism have virtues. They are ontologically lean: they invoke only us and whatever must go into the ontology of our language and thought anyway. They demystify the mysterious property of necessity: it arises because our conventions do not let us say certain things coherently, or our minds do not let us coherently conceive them. Either way, necessity is the flipside of an impossibility we can explain naturalistically. Such accounts also let us understand how we have access to facts about what is possible, for our access to these is either just access to our own settled linguistic dispositions or to our own inner lives.

The problem with all this is that it's anti-realist. We might live with this when it comes to necessity. Necessity is a mystery; we're glad of a seemingly common-sense explanation. But at possibility, we jib. At the Big Bang, it was physically possible that there were dinosaurs. If physically possible, then surely absolutely possible. How could our conventions shape the past when we were not even gleams in the universe's eye? One answer might be that past, present and

future are all equally real, and so our conventions were real even at the Bang, and so "reached" backward from the first point at which they existed. But this makes commitment to a controversial theory of time the price of conventionalism—whose *lack* of large metaphysical commitments was supposed to be one of its major attractions. Further, it still requires us to say that our conventions applied before they began to exist. Even with the background theory of time just mentioned, that does not sound plausible.

5.4 The best of both worlds

Realism without origins has problems. Prior-style realism with origins inherits all of them. But our other origin stories are all anti-realist, so far. It's worth asking, then, whether we can somehow combine features: get a realist story that uses an anti-realist style origin for logical space to avoid bloat and does not get the relation between truth and reality the wrong way round. There is a clue in our anti-realist origin stories. They are both *power* theories. Conventionalism says that what is absolutely possible is just what we have the power to say coherently, given current conventions. Conceptualism says that what is absolutely possible is just what we have the power to conceive, in our actual state or perhaps in some ideal state. On either, what is possible is determined by what a particular power of ours is able to produce. But why should only our powers, and only these powers, matter? One can't help thinking that the reason is that for conventionalists and conceptualists, these powers produce the bearers of modal status, settle their truth-values, and thereby determine the modal facts. In other words, these power theories inherit the wrong-way-round problem of propositional possible worlds. Sentences and conceptions ought to reflect and report modal reality, rather than constitute it.

We might try to make progress by generalizing. Let's abandon the anthropocentrism of conventionalists and conceptualists. Let's say that what is absolutely possible is just what *any* power or group of powers is able to produce. That, however, is just a first pass. Ten billion years ago, no group of natural powers was able to produce homo sap. Nothing then naturally existent could simply act and have us as the result. But there were natural powers with the power to produce other natural things with other powers, and so on down a very long line, with the result that eventually, down that line, there would come to be things which, acting together, could produce a homo sap. So we might suggest this: possibly *P* just if there are or have been powers which, acting jointly, could produce powers which, acting jointly, etc., which could produce powers which, acting jointly, could bring it about that *P*. Let's say this as follows: possibly *P* just if a power chain reaches from some point in actuality to its being the case that *P*.

This approach has many virtues. It is realist: what powers there are does not depend on human thought or convention. It is ontologically lean: it requires only concrete things and their powers, and if one can (as I believe) give a successful nominalist account of powers, it requires only concrete things. It gets the relation between truth and reality right: it is the powers which make it true that possibly *P*. And it tells no strange tales about you. Further, grounding truths about possibility in powers is no stranger than grounding it as other realist theories do. It might be less so. For we ordinarily think that e.g. what *makes it possible* that I get to class tomorrow is that I have the power to walk there: the presence of a power *makes it possible* that things be as the power can make them. Power theories just take this ordinary locution very seriously. Again, powers theories de-mystify necessity. The conventionalist tells us (we now see) that what is impossible is what we do not have the power to say coherently, and necessity is just impossibility's flipside. If that's unmysterious, so is the generalized version that the impossible is what there is no power (or power chain) to produce, and necessity is the flipside of *that*. Finally, powers theories offer a good account of our access to possibility. On conventionalism, it was our access

to our linguistic powers. On conceptualism, it was our access to our conceiving powers. On a general power theory, it is our access to all powers, which is satisfyingly empirical, though it offers the same role to thought-experiment, conception, etc., that science does.

5.5 Natural powers aren't enough

So far, so good. But there is a fly in the ointment. Our standard ways of fixing modal belief tell us that there are many possibilities which natural powers do not give us. They tell us, for instance, that it could have been the case that $E = MC^3$, or that the universe had an infinite past, instead of the finite past it has in fact had. But no power chain reaches from anywhere in actuality to these being so. As of the Big Bang, it was too late for anything natural to give us an infinite past. (It is always too late for anything to give us any sort of counterfactual past. The past cannot be affected.) And as of the Big Bang, it was a law that $E = MC^2$, and no natural thing has ever had the power to replace that law with another. So power theorists face a choice. They must either bite the bullet and restrict absolute possibility to natural possibility, or admit non-natural powers, powers sited outside nature. Further, there are such powers only if they have bearers, and so the second alternative admits non-natural things.

5.6 So let's consider non-natural powers

The popularity and tenacity of naturalism in current philosophy is not proportioned to the strength of arguments for it. (Naturalists will, of course, say the same of the general population's non-naturalism.) If you have a naturalist knee-jerk against admitting the non-natural, ask yourself why, and then admit, honestly, that non-naturalist theories are at least worth a look, particularly when the modal cost of naturalism looks high. I now give the look, hopefully too fast for you to look away.

If there are non-natural powers, there are non-natural things. (Nominalists will particularly insist on that, since they reduce powers to things in the end.) We settle how many and of what kind by how much non-natural possibility we want to admit. Our ways of settling modal belief incline us to say that the range of the absolutely possible is something like whatever logic and obvious metaphysical necessities do not rule out: which is roughly what another age would've meant by "whatever does not imply a contradiction." Now, omnipotence is traditionally understood as power enough to bring about anything in *that* range. So what we need are powers that add up to the range of omnipotence.

We could site these powers in many small non-natural things or in one big one. The many-little-guy hypothesis brings no particular advantages. Having one big guy instead is simpler. "Ockham sez: pare away!" You might protest: we trade many little guys for the cost of one big, unfamiliar property, omnipotence. But that's not right. We could just as easily say that the one big guy has all the little powers we would've sited in the many little guys. In that case, all we've done is pare the ontology down, introducing nothing new or strange in compensation.

So what we need is a non-natural omnipotent being. Since location in spacetime or material constitution would make it natural, it cannot have either. It must be eternal to keep the ontology down: one eternal being is simpler than many short-lived ones. It also must be necessary to keep the ontology down: one being in all worlds is simpler than one actual being and some ontology that provides (somehow) for its replacement. At this point naturalist knees will surely be knocking, but why? All we're doing is following the same principle of theory-construction they favour: posit no more than we need, but enough to provide ontology for everything intuition says is there. The next question is, what *else* is true of the big guy? Well, it has powers including

the power to bring it about by means of the appearance of rationally persuasive considerations that rational agents make various choices. Without those, it would not be omnipotent. By making sentences appear, it can appear to reason. It can appear to cajole. It can appear to make offers. So it can pass the Turing test. Naturalists worth their salt will tell you that all we are is meat that can pass the test. If something without meat can pass the test, they ought to concede that it's as much a person as we are.

5.7 Wherein I pull God out of a hat

Well, heck—it's close enough. Let's just call it God, though we have as yet said nothing about its omniscience or moral attributes. If we do, we still preserve the virtues of a power theory. Our ontology adds only one concrete non-natural thing. We still have clear accounts of necessity and our access to natural possibility. Our access to non-natural possibility can't come by empirical study of the nature of God, but since God is there to undergird our ordinary epistemology of possibility—to provide possibilities for everything our ordinary intuition tells us is possible—our access to the possible is just what we think it is, and we can provide whatever natural or supernatural account of why conceiving and thinking counterfactually are reliable seems best to us. For reasons I've sketched and more beside, we ought to take power theories seriously, and the best power theory will be one that posits God.

Theist modal theories, including historical divine-power theories, generally do not think that logical space originated. On a typical divine-power view, logical space is what it is because God's powers are what they are. But on a typical view, God's powers do not originate. They are His by nature, and so they are simply there, as He and His nature are. I have no quarrel with saying this of some divine powers. But I am in a minority camp—perhaps a camp of one. I think that other divine powers do have an origin, though they are still eternally and necessarily there. Perhaps the latter implies that it would be less misleading to say that they have an eternal, necessary explanation from elsewhere in God.

5.8 God as originating logical space

I don't think God has the power to contravene logic or mathematics. Nor do you, so here no argument is needed. Where we differ, perhaps, is over why this is so. I do not think it is an external constraint on God. Rather, this is just part of His nature. As I see it, the content and necessity of logic and mathematics are engraved in God's nature—and what is engraved there settles that content rather than inheriting it from some other source. God's nature is the first *locus* of logical truth and necessity. God's powers and lacks of power are the primary truth- and necessity-maker for e.g. the principle of non-contradiction. It is necessary because God cannot for any *P* bring it about that *P* and not-*P*. Explanation here grounds out in the divine nature. I think the same about truths about the morally good, insofar as these aren't specified to creatures. But modal truths about or involving the natures of creatures seem to me to call for a different story.

The typical theist story about modality writes all creaturely natures into God's. It may say, for instance, that God's nature made Him conceive hippopotami, and hence He has by nature the power to make them. Or it may say more simply that God's omnipotence naturally includes the power to make all possible creatures, and so includes the power to make hippopotami. So one way or another, on the typical theist story, God's nature includes that of hippos. God's nature is the ultimate zoo: if we could just see far enough in, we'd see every possible animal in there. I demur. God's nature is the property which makes Him God and makes Him Himself. I don't see what hippos have to do with that. They don't help with that job. I think God could've been God

if there were no such thing as hippopotamus-nature: if hippos were not so much as impossibilia. And I think that's a *significant* counterpossible, one that isn't just trivially true but reveals something substantive about the natures involved. Sure, deity makes God omnipotent. But that just means that by nature He can *inter alia* create any creatable creature. It contains nothing about what creatures are creatable. That content, I think, comes from elsewhere in God. On the typical theist view, it is written into God's nature, part of its content, that water is H_2O. So if, counterpossibly, water had some other composition, God's nature wouldn't exist. Some other property, or no property at all, would exist instead. But if God's nature doesn't exist, then God doesn't exist. So on the typical theist view, if water goes down the drain, it carries God with it. Why should that be so? Surely God's nature is sufficiently independent of water's nature that the significant counterpossible, the one that reflects the facts, is this one: were the nature of water different, God would still be divine and exist.

If you've come this far with me and share my intuitions, you might wonder whether there is a way to accommodate them. There is. I say that the natures of creatures are free, creative inventions of God. He did not have to think up hippos. Nothing in Him constrains Him to do so. What was to be necessary—hippo nature or schmippo nature—was entirely up to Him. Retrospectively, given that He has thought up hippos, eternally and necessarily, possibly there are hippos. But it was and is "in Him" to have things come out differently. I can best explain this "in Him" by modifying some work by Kadri Vihvelin (2013). She argues that even given determinism, if you choose to do A at t, you may still be able at t to choose otherwise. She bases this on the following:

> S has the narrow ability at time t to do R (by) trying iff, for some intrinsic property B that S has at t, and for some time t' after t, if S had the opportunity at t' to do R *and* S tried to do R while retaining . . . B until . . . t', then in a suitable proportion of these cases, S's trying to do R and . . . having . . . B would be an S-complete cause of S's doing R.[1]

Abilities are narrow just if we have them regardless of whether we have the chance to use them. If I am chained to a bed but unparalysed, I have the narrow but not the all-things-considered ability to get out of bed. An S-complete cause is one complete with respect to S' havings of intrinsic properties. Vihvelin's idea is basically this: narrow abilities are clusters of intrinsic dispositions. Intrinsic dispositions rest on purely intrinsic bases. So what has a narrow ability's basis has the narrow ability, even if the past and the laws of nature manacle it, denying the chance to use it. Lacking a chance doesn't entail lacking an ability one would have used had one had the chance. Because their bases are purely intrinsic, something has an intrinsic disposition just if it has its intrinsic base, whatever its opportunities. Given chances, if one tried, one would manifest one's intrinsic ability in a suitable proportion of cases, by successfully using it.

I find this account plausible. A disposition is at a first pass the *in re* correlate of a function from "triggering" situations to final effects. For the dispositions which are narrow abilities, the triggering situation is having its base plus a chance to use it and trying to bring about what it brings about. Things do not lose dispositions because they are not in a triggering situation—diamonds are hard even when not pressed—and lack of a chance suffices to not be in a trigger-situation.

On determinism, given the past and natural law, you metaphysically cannot do otherwise. Though there are possible worlds in which you do otherwise, in none of these are the past and laws as they actually are; the past and the laws jointly entail that you do what you have done. But on Vihvelin's story, even so, you are narrow-able at t to do otherwise if your intrinsic endowment at t and your choice would S-fully account for your doing A at another time if you were still so

endowed. You are thus able just because of your intrinsic endowment, whether or not you metaphysically can use the ability in your actual circumstances.

Suppose that Vihvelin is on the right track. In God's case we can delete the "suitable proportion" clause. Narrow abilities are intrinsic. If God is omnipotent, then if He tries to do what use of an intrinsic ability suffices to do, He succeeds unless He runs up against logic, mathematics or another omnipotent will. A God necessarily omniscient and necessarily perfectly rational cannot will contra-logically or contra-mathematically. Christian theology has it that the Persons of the Trinity cannot will to oppose one another. And plausibly, if God is by necessity omniscient and perfectly rational, then if He created another omnipotent will, He would also assure that it could not contradict Him.[2] So if God tries to do what use of an intrinsic ability suffices to do, He succeeds: period.

We can also delete "as a result of trying"—God cannot act unintentionally, and so He can only act by trying. Here B is the intrinsic divine natural endowment, and so reference to retaining B later drops out: God has B eternally. We now have

> V★. God has the narrow ability at time t to do R iff He has some intrinsic property B such that for some time t' after t, if God had the opportunity at t' to do R and tried to do R, God's trying and being as He is intrinsically would be a God-complete cause of God's doing R.

This gives a legitimate sense of divine ability if Vihvelin is on target. But like Vihvelin's story, it does not imply that God possibly tries at t or at t'. It provides a sense in which God is able to do otherwise even if it is not possible that He do so. So when I say that it is in God to have thought up schmippos instead of hippos, I mean just to apply something like (V★). Given that He has not done this, it is now not possible that He do it or have the opportunity to do it. But His intrinsic endowment actually is such that if, counterpossibly, He had the opportunity, He would be able to not think up hippos and instead think up schmippos. In this limited sense—that His intrinsic endowment did not and does not constrain Him to think as He does, that it does not in any way direct Him to think up hippos and not schmippos, that He could do otherwise if He had an impossible chance—His thinking up the natures of creatures is free.

Once He has thought up creatures, He considers them. If He approves of them sufficiently, He permits Himself to make them. As naturally omnipotent, He has the power to make whatever He permits Himself to make. So once He permits, He has a specific power to make them—and once that power exists, that makes it possible that there be, say, hippos. The same goes for whole creature-involving possible worlds, which He thinks up just as He thinks up individual creatures.

This account does all the work the standard theist account does. It is just modified so as to avoid the incongruities I raised for the standard account. So I say that logical space eternally, necessarily has originated in God. The standard theist could actually say the same, deleting only the freedom of the account, if he or she held that divine powers are what make possible and are dependent for their existence on God's (eternally, necessarily) generating an idea of what they are powers to bring about.

Notes

1 Vihvelin (2013, 187). In the original, the second occurrence of "t'" is not primed, but this has to be a typo.
2 Swinburne (1994, 174–5) has a story about relations between the triune Persons that could be adapted to this.

References

Lewis, D. (1973) *Counterfactuals* (Cambridge, MA: Harvard University Press).
Lewis, D. (1983) "New Work for a Theory of Universals," *Australasian Journal of Philosophy* 61, 343–77.
Lewis, D. (1986) *On the Plurality of Worlds* (Oxford: Basil Blackwell).
Plantinga, A. (1974) *The Nature of Necessity* (Oxford: Oxford University Press).
Prior, A.N. (1960) "Identifiable Individuals," *Review of Metaphysics* 13, 684–96.
Swinburne, R. (1994) *The Christian God* (Oxford: Oxford University Press).
Vihvelin, K. (2013) *Causes, Laws and Free Will* (Oxford: Oxford University Press).
Williamson, T. (2013) *Modal Logic as Metaphysics* (Oxford: Oxford University Press).
Wittgenstein, L. (1961) *Tractatus Logico-Philosophicus* (London: Routledge and Kegan Paul).


Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

PART 2

Essentialism, ontological dependence, and modality


Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Chapter 6

ESSENTIALISM AND MODALITY

Penelope Mackie

The notions of essence and essential property have played a prominent role in the study of modality throughout much of the history of the subject. Yet there is considerable debate about how to characterize these notions, about their legitimacy, and about their significance. This chapter concentrates on some of the salient debates about essentialism and its relation to modality in analytic philosophy in the last fifty years. Over that period, the focus of attention has shifted, to some extent, to questions about how the notion of essence should be characterized, and the relation between essence and modality. The contents of this chapter reflect this shift.

In addition, principally because of limitations of space, I focus primarily on essentialism as it applies to individual things, as opposed to kinds of thing, although I briefly discuss the extension of essentialism to kinds of thing towards the end of this chapter. In particular, I avoid discussion of essentialism about natural kinds, and the related topic of scientific essentialism, hugely important though these forms of essentialism have been in analytic philosophy during the last fifty years.[1, 2]

Essentialism, essential properties, and essence

Essentialism may initially be characterized as the combination of two theses: first, that it makes sense to distinguish between the essential properties of a thing and its non-essential (or "accidental") properties, and second, that (at least some) things have (at least some) essential properties. According to this characterization, an *anti-essentialist* is someone who either denies that the distinction between essential and accidental properties is coherent, as did Quine (1953a, 1953b, 1960: 199–200), or accepts the coherence of the distinction but denies that anything has any essential properties.[3]

An essentialist might say that *being human* is an essential property of Queen Victoria, whereas *being queen of England* and *being a parent* are among her merely accidental properties. An essentialist might hold that it is an essential property of the Eiffel Tower to be made of metal, but that its exact location is one of its accidental properties. Note that these examples are purely illustrative, though. Commitment to essentialism, in the sense just characterized, does not entail any particular commitments about which of a given thing's properties are essential to it – or even to which things have essential properties.[4] An essentialist might hold that Queen Victoria is *not*

essentially human, although she is essentially a person; or that she is not essentially either of these, although she is essentially animate; or that the only things that have essential properties are abstract objects; or even (an extreme version of essentialism – "hyper-essentialism") that all Victoria's and the Eiffel Tower's properties are essential to them.

My characterization of essentialism in terms of essential properties is, of course, scarcely informative without an account of what an essential property is. Here it is impossible to say anything completely uncontroversial. But recent discussion has focused on the following two conceptions:

> (EM) A thing's essential properties are its necessary properties – those properties that it has necessarily (in the sense that it could not have existed without having them, or has in all possible worlds in which it exists).
>
> (ED) A thing's essential properties are those of its properties that define *what it is to be that thing*. This is sometimes amplified as follows: *being F* is an essential property of a thing *x* if and only if it is true in virtue of the identity (or nature) of *x* that *x* is F.[5]

On both conceptions, a thing's essence may be equated with the sum of its essential properties.[6] Following Fine (1994), these conceptions have been labelled the "modal" and "definitional" conceptions of essence and essential properties respectively. It is important not to be misled by these labels, though. Fine and other proponents of the definitional conception hold that all essential properties are necessary properties, although they deny that the converse holds (Fine 1994: 4, 8; Lowe 2012: 938). Hence these proponents of (ED) take a thing's essential properties to have modal implications – implications about its necessary properties and "modal profile". The modal conception is so called because (unlike the definitional conception) it *analyses* essence in modal terms – specifically, in terms of the (modal) notion of a necessary property.

On either conception, an essential property is a modal property. Hence essentialism as characterized earlier involves *de re* modality, in a standard sense of that notion. For it involves the attribution of modal properties to things (*res*), and not (or not merely) the attribution of modal status to statements (*dicta*) or propositions, as is the case with *de dicto* modality.

As for the type of modality associated with essentialism, again it is impossible to say anything completely uncontroversial. But it is standardly assumed that the relevant notion is that of *metaphysical* modality – in contrast to both merely nomological or physical modality and epistemic modality.[7] In what follows, I shall take this for granted.

The second conception of essential properties, (ED), obviously requires an elucidation of the relevant notion of definition – which must be applicable to *things* (including concrete particulars like Queen Victoria and the Eiffel Tower), and not (or not merely) to words or concepts. Here its proponents appeal to a broadly Aristotelian conception of *real definition*, in contrast to a verbal definition (Fine 1994: 2–3; Lowe 2012: 935–8; Hale 2013: 152–3; Oderberg 2007: 19).

Fine's critique of the modal conception of essence

Following the revival of interest in essentialism in analytic philosophy inspired principally by the work of Kripke and Putnam in the 1970s, the modal conception was dominant – indeed, usually just taken for granted. More recently, though, largely through the influence of the work of Kit Fine, the second, definitional, conception has been gaining prominence.

In "Essence and Modality" (1994), Fine presented a critique of the modal conception of essence that had two principal strands. The first was an argument that (EM)'s equation of

essential properties with necessary properties is untenable, since some of a thing's necessary properties are not among its essential properties in a metaphysically interesting sense of "essential". Fine's proposed counterexamples to (EM) included Socrates's property of *being distinct from the Eiffel Tower*, his *being such that there are infinitely many prime numbers*, and his *belonging to the set singleton-Socrates*: the set that has Socrates as its sole member (Fine 1994: 4–5). Each of these is a necessary property of Socrates, in the sense of a property that he could not have existed without having (that he has in all possible worlds in which he exists), and yet, according to Fine, is not properly classified as an essential property of Socrates, because it does not pertain to his essence or nature or identity. The second strand in Fine's critique was his diagnosis of why (EM) has these consequences. Fine argued that they are symptomatic of the fact that it is fundamentally misguided to try to explicate the notions of essence and essential property in modal terms. His counterproposal was that we should see the notion of *real definition* (rather than modality) as central to the understanding of the concept of essence (1994: 3); hence his advocacy of (ED) in place of (EM).

An important element in Fine's case was that (EM) is incapable of capturing certain asymmetries – for example, that although Socrates is not essentially a member of the set singleton-Socrates, singleton-Socrates essentially contains Socrates (Fine 1994: 5). By contrast, (ED) can accommodate the idea that the source of the necessary truth "If Socrates exists, Socrates is a member of singleton-Socrates" lies in the essence of the singleton, and not in the essence of Socrates (cf. Fine 1994: 9).

Even prior to Fine's critique, it was acknowledged that (EM) classifies, as essential properties, some that are, in an obvious sense, only trivially necessary. The clearest examples are properties that necessarily belong to everything, such as self-identity, or *being such that* 2 + 2 = 4 (see Della Rocca 1996: 3; Forbes 1986: 4; cf. McMichael 1986: 33). The standard response to this, however, was not to modify (EM) but instead (a) to distinguish trivial from non-trivial essential properties and (b) to acknowledge that, in many philosophical contexts, the versions of essentialism that are of interest are those that ascribe some *non-trivial* essential properties to things. If Fine's critique is accepted, however, this modest adjustment is insufficient to rescue the modal conception of essence.[8]

Fine's proposed counterexamples to (EM) have been very widely accepted as such. There is far less agreement, though, concerning his diagnosis.

There have been several attempts to rescue a version of the modal conception by amending the simple modal account (EM) to avoid the counterexamples – most obviously by restricting the class of essential properties to those necessary properties that satisfy some further condition. In addition, these defences of modalism are typically critical of Fine's alternative conception of essence. I cannot undertake a detailed discussion of these modalist responses here. But significant examples include Correia 2007, Cowling 2013, Zalta 2006, and Wildman 2013; for critical discussion, see Skiles 2015 and Wildman 2016. Supporters of Fine's definitional approach, on the other hand, include Hale (2013), Lowe (2008, 2012), Shalkowski (2008), and, with some qualifications, Koslicki (2012) and Oderberg (2007).[9]

What essential properties do things have?

Whichever conception of essential properties is adopted (modal or definitional), there can be disputes about what (non-trivial) essential properties things have.[10] For example, one modalist may hold that Queen Victoria is necessarily (and hence essentially) the child of her actual biological parents, while another claims that this feature of her ancestry represents a contingent fact about her, and hence a non-essential property. One modalist may hold that Queen Victoria is

necessarily (and hence essentially) a member of the human species, while another holds that she could have been a robot or a Vulcan or an intelligent chimpanzee.

Disputes among proponents of the definitional view, on the other hand, may take either of two forms. On the one hand, the disputants might agree that something is a (non-trivially) necessary property of Victoria, while disagreeing about whether it is part of her essence. Indeed, Fine maintains that it is a virtue of the definitional conception that it allows for such disputes about the "essence-source" of certain necessities (Fine 1994: 8–9). On the other hand, the disputants might disagree about whether the property is even a necessary property of Victoria, as well as whether it is part of her essence. Such disputes will be at least superficially similar to the modalists' disputes mentioned earlier – as to whether Queen Victoria is essentially human, or essentially the child of her actual biological parents, for example.

Space does not permit detailed discussion here of the debates about the attribution of non-trivial essential properties that raged, from the 1970s onwards, among adherents to the modal conception of essence. Broadly speaking, though, they concentrated on three issues. One is whether everything belongs essentially to some sort or kind, and if so, what criteria determine these "essential" sorts or kinds – an issue made prominent by the work of Wiggins (1980, 2001). The question whether Victoria is essentially human falls under this heading. A second issue concerns Kripke's notorious "necessity of origin" thesis (Kripke 1972, 1980), exemplified by the questions whether Victoria could have existed with different biological parents, and whether the Eiffel Tower could have come into existence composed of entirely different metal. A third issue is whether ordinary concrete particulars (including humans, animals, and artefacts) have non-trivial (substantial) individual essences, where an individual essence of an object x is a property, or set of properties, that is not only essential to identity with x, but also necessarily sufficient for identity with x. For discussion of some of these debates, see Mackie 2006, 2017; Mackie and Jago 2017; Robertson and Atkins 2018; Roca-Royes 2011.

In spite of extensive discussion over many decades, little consensus has emerged. Moreover, it is remarkable that such questions do not appear to become any easier to resolve if recast in terms of the definitional (rather than the modal) conception of essence. Nor, I should say, does Fine suggest otherwise. This prompts the thought that perhaps the principal significance of finding a satisfactory conception of essence lies not so much in the fact that metaphysicians are interested in "the identity of things, with what they are" (Fine 1994: 1), but rather in the connection between the concept of essence and other metaphysical concepts such as modality, substance, and ontological dependence – as suggested initially by Fine (1994) and in his later writings (e.g., 1995a, 1995b).

It might be supposed that, on one of these issues, the definitional conception must favour one side rather than the other. This is the question whether ordinary concrete things have non-trivial (substantial) individual essences.[11] It might be thought that a (real) definition of *what Socrates is* must specify properties guaranteed to be unique to Socrates and thus be distinct from the real definition of any other individual. This conception of real definition has been adopted by some of Fine's followers, notably Lowe (2008: 35) and Hale (2013: 132–3, 151, 222–3). However, this interpretation of (ED) is not mandatory, and there is evidence that Fine himself rejects it.[12] It seems compatible with (ED) that the real definition (and hence the essence) of Socrates is *not* unique to Socrates – and even that his essence might consist simply in *being a man* (*being human*), for example. Moreover, since there are notorious problems about finding plausible candidates for the role of a substantial (non-trivial) individual essence of a concrete entity such as Socrates, there is reason to think that a definitional approach to essence should try to avoid this commitment (for discussion of the problems, see Mackie 2006: chs. 2–3; Mackie and Jago 2017).[13]

Metaphysical modality based on essence

Even more striking than his claim that essence should not be understood in terms of metaphysical modality is Fine's further claim that metaphysical modality should be understood in terms of essence. According to Fine, all metaphysical modality has its source in the essences or natures of things:

> [F]ar from viewing essence as a special case of metaphysical necessity, we should view metaphysical necessity as a special case of essence. For each class of objects . . . will give rise to its own domain of necessary truths, the truths which flow from the nature of the objects in question. The metaphysically necessary truths can then be identified with the propositions which are true in virtue of the nature of all objects whatever.
>
> (Fine 1994: 9)

In taking the notion of essence to be explanatorily prior to that of modality, Fine is followed by Lowe (2012: 934). Hale (2013) makes a substantial further contribution to the debate with his development of an "essentialist" (essence-based) theory of modality, although Hale explicitly distances himself from the ambition, which he attributes to Fine, of thereby providing a reductive account of modality (2013: 134 n. 27, 150 n. 12).

Fine's revolutionary proposal is obviously of huge potential significance. As yet, though, his "essence-first" approach to modality has received rather less attention than has his attack on the modal conception of essence. One surprisingly neglected issue, of fundamental importance to the viability of any essentialist account of modality, is whether it can be guaranteed that what is true in virtue of the identity or nature or essence of an object (as these notions are understood by (ED)) is a necessary truth.[14] In addition, as well as the fact that the essence-based approach faces a backlash from proponents of the modal conception of essence, a new rival to an essence-based account of modality has recently emerged, in the form of a potentiality-based theory (Vetter 2015; cf. Vetter 2011).

Essentialism and necessary *a posteriori* truth

A discussion of essentialism in analytic philosophy over the past fifty years would be incomplete without a comment on the relation between essentialism and *a priori* truth. On certain assumptions, essentialism as characterized in the opening section of this chapter (whether understood in modal or definitional terms) generates necessary truths that can be known to be true only empirically (*a posteriori*), and not by purely *a priori* means. For example, if Mary Shelley is essentially female, and essentially a child of Mary Wollstonecraft, then, on certain plausible assumptions, the statements "If Mary Shelley exists, she is female" and "If Mary Shelley exists, she is a child of Mary Wollstonecraft" are necessary truths.[15] Yet these facts about Mary Shelley's ancestry and gender cannot be known *a priori*, but only by empirical investigation. Indeed, the fact that Kripke's (1972, 1980) arguments for essentialism were also arguments for necessary *a posteriori* truths was one of the most provocative and influential features of his work.

Whether we adopt a version of (EM) or its rival (ED), we should not expect there to be any pure *a priori* insight into the essences of particular things. As will be seen in the next section, however, the same need not always be true of the essences of kinds of thing.

De dicto modality and essentialism about kinds

Arguably, the characterization of essentialism used throughout this chapter so far is unduly restrictive, in taking essentialism to require the attribution of essential *properties* to things – thus, restricting it to a thesis about modality *de re*. A more comprehensive conception of essentialism would include claims about the essences of *kinds* of thing that need not (and perhaps should not) be cashed out in terms of the attribution of essential properties to those kinds.

It is a familiar idea in philosophy, going back at least to Socrates and Plato, that it makes sense to ask about the essence of *justice*, or *knowledge*, for example, or about the properties that are essential to being a bachelor, or a triangle, or a set, or an artwork, or a person, or a platypus, or a table, or a sample of water, or a piece of gold, or an electron. It is possible to construe these as questions about the essential properties of certain *things* (entities) – those denoted by the terms "justice", "knowledge", etc., or the kinds *bachelor*, *triangle*, *person*, *water*, *gold*, and so on. Yet there seems no good reason to insist on such a construal – and reasons to resist it.[16] This raises the possibility that we might take one form of essentialist enquiry to be a search for the conditions that are essential to *being a thing of a certain kind* (e.g., a just action, or a person, or a set, or an electron), without taking this to be an enquiry into the essential properties of a kind, regarded as an entity. It should be noted that it does not follow, from the fact that there are conditions essential to *being an F*, that individual things that are Fs are essentially Fs. This is obvious from the example of the kind *bachelor*. But (although this is sometimes overlooked) the point applies quite generally. Thus, although many would say that any person is essentially a person, or that pieces of gold and electrons are essentially pieces of gold and electrons respectively, these claims go beyond any claims about the essential conditions for being a person, a piece of gold, or an electron.

In the light of these observations, we may recognize a way in which an enquiry into essences may concern modality *de dicto* rather than modality *de re*. In fact, there is a crucial *de re*/*de dicto* ambiguity in the expression "the essence of an F". On a *de dicto* reading, this means "what is essential to *being an F*". On a *de re* reading, it means "what is essential to being a thing that is in fact an F". There seems no reason to regard "*de dicto* essences" as of any less interest to the metaphysician than "*de re* essences". Indeed, when Fine argues in the opening paragraphs of "Essence and Modality" for the importance of the concept of essence, his example of a metaphysically significant question, namely, "What is a person?" (Fine 1994: 1), is more naturally construed as a question about the *de dicto* essence of persons (what it is to be a person) than as a question about the *de re* essences of particular things that are persons.

It might be objected that if we include, within the scope of essentialism, what I am calling "*de dicto* essences", this is in danger of trivializing the subject. For example, it looks as if an enquiry into the essence of *being a bachelor* could be satisfied by a form of conceptual analysis that is innocuous even in the eyes of the most intransigent empiricist, including Quine.[17] In response, however, I suggest that, once it is admitted that essentialism about kinds can legitimately focus on the conditions for belonging to a given kind, there is no principled reason to restrict such essentialism to cases where these conditions can be known only *a posteriori* rather than *a priori*.[18]

Notes

1 On natural kind essentialism, including discussion of the seminal contributions of Kripke (1972, 1980) and Putnam (1973, 1975), see Bird and Tobin 2018. On scientific essentialism see, for example, Bird 2007 and Ellis 2001.

2 Regrettably, limitations of space also preclude me from discussing either Quine's famous scepticism about "Aristotelian essentialism" or David Lewis's counterpart-theoretic interpretation of the notion of

an essential property according to which, contrary to an assumption shared by Quine and his original opponents, attributions of essential properties may be *de re* without being referentially transparent (Lewis 1971, 1986).

3 As will be noted in this chapter, on one conception – (EM) – of the notion of an essential property, it seems impossible to deny that, if the notion is coherent at all, everything has some essential properties, even if they are only "trivial" ones. For simplicity, I ignore this here.

4 See the previous note.

5 Fine (1995a: 273–5, 1995b: 54), Correia (2012: 641), and Koslicki (2012: 190).

6 See also Fine 1995b for a refinement to the definitional conception (ED) that distinguishes "constitutive essence" and "consequential essence" and allows for a distinction between core essential properties and derivative essential properties. For discussion, see Koslicki 2012 and Correia 2012.

7 According to most accounts, metaphysical modality is also distinct from logical modality, narrowly conceived, since there may be metaphysical necessities (e.g., that water is H_2O) that are not logical necessities in a narrow sense of "logical necessity". A discussion of varieties of modality is beyond the scope of this chapter.

8 Independently of Fine's critique, this modified characterization of essentialism is still too broad to satisfy some tastes. For more on characterizations of essentialism, in relation to essential properties, see Robertson and Atkins 2018, Roca-Royes 2011: 66–8, and Yablo 1998. (A further issue, concerning the extension of essentialism beyond the attribution of essential properties to things, is discussed later in this chapter.)

9 See also Gorman 2005 for an account that rejects modalism while criticizing Fine's version of the definitional conception.

10 The qualification "non-trivial" may be redundant in the case of the definitional conception. This is suggested by Paul (2006: 366 n. 2).

11 Exactly what counts as a non-trivial (or substantial) individual essence is disputable. However, it is generally agreed that the property of *being identical with Socrates*, although it satisfies the criteria for being an individual essence, is not a non-trivial (substantial) individual essence in the relevant sense (Mackie 2006: 18–22).

12 "Just as we can think of a collection of sentences as providing a nominal definition of a term, we can think of a collection of propositions as providing a real definition of an object (*the definitions need not, in either case, be individuating*)" (Fine 1995a: 275, my italics).

13 Wiggins (1980, 2001) presents a version of essentialism focused on Aristotle's "*What is x?*" question, and which seems broadly in harmony with Fine's approach. Yet Wiggins repudiates the idea that things like human beings have substantial *individual* essences, as opposed to essences that are "shared or shareable" (1980: 120, 2001: 125–6).

14 But see Mackie 2020, Leech 2018, and Noonan 2018. Also relevant is the discussion of the notion of a "nature" in Cowling 2013. Other discussions of Fine's essence-based account of modality include those by Correia (2006, 2012), Cameron (2008, 2010), Vetter (2011), Wildman (2018), and Romero (2019).

15 For more details, see Mackie 2006: 8–11. The *locus classicus* is Kripke 1972, reprinted in Kripke 1980. This does not preclude the possibility that if these are necessary truths, they depend partly on general essentialist principles that are knowable *a priori* (Kripke 1980: 159).

16 For one thing, it requires a reification of kinds which some would find objectionable.

17 Cf. Mackie 2006: 13. I now reject the suggestion, made there, that "essentialism about kinds" should be restricted to cases where the essences of kinds cannot be known *a priori*.

18 As I interpret the views of Fine, Lowe, and Hale, they are all sympathetic to the idea that some questions about essence can be settled *a priori*, by conceptual investigation.

References

Bird, A. (2007) *Nature's Metaphysics*, Oxford: Oxford University Press.

Bird, A., and Tobin, E. (2018) "Natural Kinds," in E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2018 edition). Available at: https://plato.stanford.edu/archives/spr2018/entries/natural-kinds/.

Cameron, R. (2008) "Truthmakers and Modality," *Synthese* 164: 261–80.

Cameron, R. (2010) "The Grounds of Necessity," *Philosophy Compass* 5: 348–58.

Correia, F. (2006) "Generic Essence, Objectual Essence, and Modality," *Noûs* 40: 753–67.

Correia, F. (2007) "(Finean) Essence and (Priorean) Modality," *Dialectica* 61: 63–84.
Correia, F. (2012) "On the Reduction of Necessity to Essence," *Philosophy and Phenomenological Research* 84: 639–53.
Cowling, S. (2013) "The Modal View of Essence," *Canadian Journal of Philosophy* 43: 248–66.
Della Rocca, M. (1996) "Recent Work on Essentialism: Part 1," *Philosophical Books* 37: 1–13.
Ellis, B. (2001) *Scientific Essentialism*, Cambridge: Cambridge University Press.
Fine, K. (1994) "Essence and Modality," in J. Tomberlin (ed.), *Philosophical Perspectives 8: Logic and Language*, Atascadero, CA: Ridgeview Publishing Company, pp. 1–16.
Fine, K. (1995a) "Ontological Dependence," *Proceedings of the Aristotelian Society* 95: 269–90.
Fine, K. (1995b) "Senses of Essence," in W. Sinnott-Armstrong, D. Raffman, and N. Asher, (eds.), *Modality, Morality, and Belief: Essays in Honor of Ruth Barcan Marcus*, Cambridge: Cambridge University Press, pp. 53–73.
Forbes, G. (1986) "In Defense of Absolute Essentialism," in French, P., Uehling, T., and Wettstein, H. (eds.), *Midwest Studies in Philosophy XI: Studies in Essentialism*, Minneapolis: University of Minnesota Press, pp. 3–31.
Gorman, M. (2005) "The Essential and the Accidental," *Ratio* 18: 276–89.
Hale, B. (2013) *Necessary Beings: An Essay on Ontology, Modality, and the Relations between Them*, Oxford and New York: Oxford University Press.
Koslicki, K. (2012) "Essence, Necessity, and Explanation," in T. Tahko (ed.), *Contemporary Aristotelian Metaphysics*, Cambridge: Cambridge University Press, pp. 187–206.
Kripke, S. (1972) "Naming and Necessity," in D. Davidson and G. Harman (eds.), *Semantics of Natural Language*, Dordrecht: Reidel Publishing Company, pp. 252–355. Reprinted in revised form as a monograph by Basil Blackwell, Oxford, 1980.
Kripke, S. (1980) *Naming and Necessity*, Oxford: Basil Blackwell. (Expanded monograph version of Kripke 1972.)
Leech, J. (2018) "Essence and Mere Necessity," *Royal Institute of Philosophy Supplement* (Metaphysics) 82: 309–32.
Lewis, D. (1971) "Counterparts of Persons and Their Bodies," *The Journal of Philosophy* 68: 203–11.
Lewis, D. (1986) *On the Plurality of Worlds*, Oxford: Basil Blackwell.
Lowe, E. J. (2008) "Two Notions of Being: Entity and Essence," in R. Le Poidevin (ed.), *Being: Developments in Contemporary Metaphysics* (Royal Institute of Philosophy Supplements) 62, pp. 23–48.
Lowe, E. J. (2012) "What is the Source of Our Knowledge of Modal Truths?" *Mind* 121: 919–50.
Mackie, P. (2006) *How Things Might Have Been: Individuals, Kinds, and Essential Properties*, Oxford: Clarendon Press.
Mackie, P. (2017) "Essentialism: Postscript," in B. Hale, C. Wright, and A. Miller (eds.), *A Companion to the Philosophy of Language*, 2nd ed., Vol. 2, Oxford: Wiley Blackwell, pp. 896–901.
Mackie, P. (2020) "Can Metaphysical Modality Be Based on Essence?" in M. Dumitru (ed.), *Metaphysics, Meaning, and Modality: Themes from Kit Fine*, Oxford: Oxford University Press.
Mackie, P., and Jago, M. (2017) "Transworld Identity," in E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2017 edition). Available at: https://plato.stanford.edu/archives/win2017/entries/identity-transworld/.
McMichael, A. (1986) "The Epistemology of Essentialist Claims," in French, P., Uehling, T., and Wettstein, H. (eds.), *Midwest Studies in Philosophy XI: Studies in Essentialism*, Minneapolis: University of Minnesota Press, pp. 33–52).
Noonan, H. (2018) "The New Aristotelian Essentialists," *Metaphysica* 19: 87–94.
Oderberg, D. (2007) *Real Essentialism*, London and New York: Routledge.
Paul, L. A. (2006) "In Defense of Essentialism," *Philosophical Perspectives* 20: 333–72.
Putnam, H. (1973) "Meaning and Reference," *The Journal of Philosophy* 70: 699–711.
Putnam, H. (1975) "The Meaning of 'Meaning'," in Putnam, *Mind, Language and Reality: Philosophical Papers*, Vol. 2, Cambridge: Cambridge University Press, pp. 215–71.
Quine, W. V. (1953a) "Reference and Modality," in Quine, *From a Logical Point of View*, Cambridge, MA: Harvard University Press.
Quine, W. V. (1953b) "Three Grades of Modal Involvement," *Proceedings of the XIth International Congress of Philosophy*, Vol. 14, Amsterdam: North-Holland Publishing Company. Reprinted in Quine, *The Ways of Paradox and Other Essays* (revised edition, 1976), New York: Random House.
Quine, W. V. (1960) *Word and Object*, Cambridge, MA: The MIT Press.

Robertson, T., and Atkins, P. (2018) "Essential *vs.* Accidental Properties," in E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2018 edition). Available at: https://plato.stanford.edu/archives/spr2018/entries/essential-accidental/.

Roca-Royes, S. (2011) "Essential Properties and Individual Essences," *Philosophy Compass* 6: 65–77.

Romero, C. (2019) "Modality Is Not Explainable by Essence," *The Philosophical Quarterly* 69: 121–41.

Shalkowski, S. (2008) "Essence and Being," in R. Le Poidevin (ed.), *Being: Developments in Contemporary Metaphysics* (Royal Institute of Philosophy Supplement 62), pp. 49–63.

Skiles, A. (2015) "Essence in Abundance," *Canadian Journal of Philosophy* 45: 100–12.

Vetter, B. (2011) "Recent Work: Modality without Possible Worlds," *Analysis Reviews* 71: 742–54.

Vetter, B. (2015) *Potentiality: From Dispositions to Modality*, Oxford: Oxford University Press.

Wiggins, D. (1980) *Sameness and Substance*, Oxford: Blackwell.

Wiggins, D. (2001) *Sameness and Substance Renewed*, Cambridge: Cambridge University Press.

Wildman, N. (2013) "Modality, Sparsity, and Essence," *The Philosophical Quarterly* 63: 760–82.

Wildman, N. (2016) "How (Not) to Be a Modalist about Essence," in M. Jago (ed.), *Reality Making*, Oxford: Oxford University Press, pp. 177–96.

Wildman, N. (2018) "Against the Reduction of Modality to Essence," *Synthese*. Published online 5 January 2018. Available at: https://link.springer.com/article/10.1007%2Fs11229-017-1667-6.

Yablo, S. (1998) "Essentialism," in E. Craig (ed.), *The Routledge Encyclopedia of Philosophy*, Vol. 8, London: Routledge, pp. 417–22.

Zalta, E. (2006) "Essence and Modality," *Mind* 115: 659–93.

Chapter 7
DE RE MODALITY

Boris Kment

The study of *de re* modality is concerned with facts about the modal profiles of individuals—facts about what could have been true of them and what could not have failed to be true of them—and with the roles of individuals in the theory of possible worlds. What follows is a selective overview of issues that arise in this part of the philosophy of modality.

7.1 Conceptualist skepticism about *de re* modality

Some philosophers have doubted that it makes sense to talk about the modal profiles of individuals. One of the best-known sources of such skepticism is a view about modality that we may call "conceptualism." Roughly speaking, conceptualists maintain that all modal facts in some sense ultimately derive from features of our concepts, or of the descriptions we use to single out the things we talk about, or from our rules for describing reality. This view was supposed to contrast with "essentialism," the thesis that modal facts about an object sometimes derive from the fact that the object is more firmly tied to some of its features (those that are "essential" to it) than to others (the "accidental" ones), where this difference is unrelated to our concepts or descriptive practices. Although the distinction between the two views is impressionistic, it has been influential in philosophical thinking about *de re* modality.

Conceptualists sometimes equated—some would say "confused"—necessity with analytic or conceptual truth. This view made it natural to think that, if a sentence of the form

(1) If *a* exists, then *a* is *P*

(where "*a*" picks out a specific individual) expresses a necessary truth, then that must be so because some definite description *d* is part of the meaning of "*a*," such that ⌜If *d* exists, then *d* is P⌝ is a conceptual or analytic truth. On such a view, all necessary truths of the form (1) must be like the following sentence (the conditional scopes over both occurrences of the definite description):

(2) If the bachelor next door exists, then the bachelor next door is unmarried.

Now, the descriptions "the bachelor next door" and "the carpenter next door" might both pick out the same person. Yet, even if (2) is an analytic or conceptual truth, the same is clearly not true of "If the carpenter next door exists, then the carpenter next door is unmarried." From a conceptualist perspective, it might therefore seem as if we could make no sense of the idea that it is true or false of the person himself, independently of how we describe him, that he is necessarily such that, if he exists, he is unmarried. Necessary connections can hold between concepts, contents, and expressions, but not between individuals and specific properties they have or conditions they satisfy. Quine (1953) concluded from considerations along these lines that quantification into modal contexts, as in (3), was dubious.

(3) $\exists x \, \Box$ (If x exists, then x is human)

The truth of (3) would require that the open sentence "$\Box$ x is human" be true under some assignment of an object to the variable "x." But it makes no sense to say that this open sentence is true under such an assignment, since a variable under an assignment is not associated with any specific description that picks out the object assigned to it. To make sense of it, so it seemed, we would need to be able to draw an intelligible distinction between the essential and the accidental properties of an object, so that we could ask whether there is an object that is essentially human. But that is anathema to the conceptualist.

Conceptualism stands in an old tradition that encompasses such famed moments in the history of philosophy as the Humean denial of necessary connections in the objects and Wittgenstein's dictum that "essence is expressed in grammar" (Wittgenstein 1953: sect. 371.) However, the philosophical mood shifted against this position under the influence of work on quantified modal logic and its philosophical foundations (Barcan 1947; Kripke 1963) and the development of the theory of direct reference. In *Naming and Necessity* (Kripke 1980), Kripke presented influential arguments for the view that some claims like (4) and (5) are necessary.

(4) If Socrates exists, then Socrates is a child of Phaenarete and Sophroniscus.
(5) If Socrates exists, then Socrates is human.

At the same time, his discussion seemed to show that a name like "Socrates" does not have a descriptive content that could explain the fact that (4) and (5) are necessary.

Conceptualists can respond by trying to give novel decriptivist accounts of names that are immune to Kripke's anti-descriptivist objections, and that allow us to explain the necessity of sentences like (4) and (5) by appealing to the descriptive content of "Socrates" and to certain non-modal facts. (Perhaps "Socrates" is associated with a description of the form *the organism that is a member of the same species and has the same parents and . . . as the actual φ-er actually does*. Combined with facts about the actual species membership and parentage of the actual φ-er, this might be held to explain the necessity of (4) and (5).) However, such a view by itself would not allow us to make sense of a sentence like (3), which does not contain proper names. To explain quantification (over individuals) into modal contexts, the account would need to be supplemented with some additional apparatus. Some conceptualist options (by no means all) will be discussed in Section 7.3. But first we need to turn to another important question about how to understand *de re* modal claims.

7.2 Haecceitism and anti-haecceitism

Haecceitism is the thesis that there are, or could have been, two possible worlds that are qualitatively alike—in the sense that the same qualitative claims are true at them—but that differ in

how things stand with specific individuals.[1] As an illustration, consider a symmetrical possible world w_0 in which no material objects exist except for two qualitatively indistinguishable and eternal particles, *A* and *B*. It seems plausible that the following claims are possibly true at w_0:

(6) Everything is the way it is at w_0 before *t*, then *A* disappears at *t* while *B* continues to exist forever.
(7) Everything is the way it is at w_0 before *t*, then *B* disappears at *t* while *A* continues to exist forever.

Assuming that a claim is possibly true iff it is true at some possible world, it follows that it is true at w_0 that there is a possible world where (6) holds and a possible world where (7) holds. Haecceitists and anti-haecceitists can agree on this much but will disagree about whether the (6)-worlds are the same as the (7)-worlds. The (6)-worlds are qualitatively indistinguishable from the (7)-worlds. If there is a difference between them, it can only concern the question of which particle disappears and which particle continues to exist. Since anti-haecceitists deny that worlds can differ only in this way without differing qualitatively, they will say that the (6)-worlds are the (7)-worlds. Haecceitists, by contrast, can say that the (6)-worlds are different from the (7)-worlds.

If anti-haecceitists want to make sense of the idea that a claim about a specific individual *i* can be true at a given world *w*, they have to say that what is true of *i* at *w* is determined by the qualitative features of *i* and the qualitative claims that are true at *w*. The standard anti-haecceitist account of how this works appeals to the notion of a counterpart. Let $\varphi(x)$ and $\varphi(x_1, \ldots, x_n)$ be formulas whose sole free variables are x and $x_1, \ldots, x_n$, respectively. Moreover, let $\varphi(a)$ be the result of replacing all free occurrences of x in $\varphi(x)$ with a name for *a*, and let $\varphi(a_1, \ldots, a_n)$ be the result of replacing all free occurrences of x_1 in $\varphi(x_1, \ldots, x_n)$ with a name for $a_1, \ldots$, and all free occurrences of x_n with a name for a_n. On one possible way of formulating the counterpart-theoretic account (Lewis 1986: 8–12):

(8) (i) *a* satisfies $\varphi(x)$ at *w* iff *w* contains a counterpart *c* of *a* that satisfies $\varphi(x)$.[2]
(ii) $<a_1, \ldots, a_n>$ satisfies $\varphi(x_1, \ldots, x_n)$ at *w* iff *w* contains individuals $c_1, \ldots, c_n$ such that $<c_1, \ldots, c_n>$ is a counterpart of $<a_1, \ldots, a_n>$ and $<c_1, \ldots, c_n>$ satisfies $\varphi(x_1, \ldots, x_n)$.

$\varphi(a)$ ($\varphi(a_1, \ldots, a_n)$) is true at *w* iff *a* ($<a_1, \ldots, a_n>$) satisfies $\varphi(x)$ ($\varphi(x_1, \ldots, x_n)$) at *w*. *a* satisfies $\ulcorner\Diamond\varphi(x)\urcorner$ (and $\ulcorner\Diamond\varphi(a)\urcorner$ is true) iff *a* satisfies $\varphi(x)$ at some possible world; $<a_1, \ldots, a_n>$ satisfies $\ulcorner\Diamond\varphi(x_1, \ldots, x_n)\urcorner$ (and $\ulcorner\Diamond\varphi(a_1, \ldots, a_n)\urcorner$ is true) iff $<a_1, \ldots, a_n>$ satisfies $\varphi(x_1, \ldots, x_n)$ at some world. "$\Box$" is the dual of "$\Diamond$."[3] Whether a given individual *c* at *w* is a counterpart of *a* depends on the qualitative similarities that hold between *a* and the different individuals at *w*. For example, Lewis (1973: 39) proposes that *c* at *w* is a counterpart of *a* iff *c* has a certain minimal degree of qualitative overall similarity to *a* and *w* contains no individual that is more qualitatively similar to *a* overall than *c* is. A similar account can be given of the conditions under which one *n*-tuple is a counterpart of another.[4]

A single individual or *n*-tuple can have more than one counterpart at a given possible world. Consider the two-particle example again. Let w_1 be a possible world that is qualitatively like w_0 before *t*. Suppose that w_1 contains two particles, *C* and *D*, and that *C* disappears at *t* while *D* continues to exist. Anti-haecceitist counterpart theorists can say that $<C, D>$ is a counterpart both of $<A, B>$ and of $<B, A>$ (where *A* and *B* are the two particles at w_0). That explains how each of (6) and (7) can be true at w_1. (However, the conjunction of (6) and (7) is not true at w_1. By (8)(ii), the truth of the conjunction at w_1 would require that there be objects *a* and *b* at w_1 such that $<a,b>$ is

a counterpart of <*A*,*B*> and *a* disappears while *b* continues to exist and *b* disappears while *a* continues to exist. Obviously, there are no such objects. Truth at a world is not closed under conjunction.)

7.3 Some forms of anti-haecceitist and haecceitist conceptualism

Anti-haecceitist counterpart theorists can give a conceptualist account of the necessity of claims like (4) and (5) and of how to understand claims like (3). The explanation appeals to constraints on how to correctly *describe* the possibilities that exist. For example, the rules defining the relevant concept of a counterpart might tell us that nothing at any possible world counts as the counterpart of an actual organism *o* unless it exemplifies the same species as *o* at the actual world. Combined with the fact that Socrates is human, this determines that no world contains a non-human counterpart of Socrates. (5) comes out necessary. Moreover, there is something (namely, Socrates) of which no world contains a non-human counterpart. So, (3) is true. In this manner, conceptualism concerning modal claims about individuals can be defended without denying the Kripkean insights that presented problems for the version of conceptualism described in Section 7.1.

Haecceitists, too, could defend conceptualism about *de re* modality by appealing to rules or conceptual constraints on how to describe modal space. There are different possible ways of doing that. For example, they could start with the notion of a world, i.e. of a maximally specific scenario or way for reality to be. (Perhaps a world is just a maximally detailed description of reality that uses both qualitative predicates and individual constants). Some worlds are possible while others are impossible. The notion of a *possible* world is defined by a list of conditions that are individually necessary and jointly sufficient for a world to qualify as possible, including the following condition: the world does not describe any actually existing organism as belonging to a species different from its actual species. No scenario in which Socrates is anything other than human counts as possible by this definition. This explains the necessity of (5) and the truth of (3).[5]

7.4 Arguments for and against haecceitism

Some anti-haecceitist counterpart theorists have tried to support their view by arguing that it allows us to accommodate the apparent context-dependence and vagueness of *de re* modal talk in an elegant way, and that it helps us to dissolve a number of modal puzzles (see, e.g., Lewis 1971, 1986: ch. 4). Haecceitists in turn have given examples that seem to support the existence of qualitatively indistinguishable possible worlds that differ in facts about specific individuals. Robert Adams imagines a world that will be symmetrical until tomorrow, at which point either your planet or its doppelganger at the opposite end of the universe will be destroyed (Adams 1979). It is natural to think that there is a difference between the possibility that your planet will be destroyed and the qualitatively indiscernible possibility that its doppelganger will be destroyed. Some anti-haecceitists have responded by granting the difference between the possibilities, but denying that it is a difference between possible worlds (maximally specific ways reality could be). Instead, there is a difference between two ways that your planet and its doppelganger could be, but these two possibilities correspond to the same possible world (Lewis 1986: ch. 4). (This account resembles the "world description theory" discussed at the end of Section 7.4.)

There are other data that can be used to argue against anti-haecceitism. Let us consider some examples.

Actuality. It seems plausible that ⌜Actually, φ⌝ is true iff φ is true at the actual world. But this creates problems for anti-haecceitist counterpart theory (Hazen 1979). Suppose that some individual at some possible world w has two different counterparts, a and b, at the actual world, and that a has while b lacks the qualitative property F. Then it follows that the following sentence is true: $\Diamond\exists x$(Actually, Fx and Actually, ~Fx). (Fara and Williamson 2005) But that seems absurd. Anti-haecceitist counterpart theorists also face other, related difficulties concerning the notion of actuality, and of course there are also proposals for how to fix these problems.[6]

Counterfactuals. Kment (2012: sect. 4) The standard account of counterfactuals tells us, roughly speaking, that ⌜$P \mathrel{\square\!\!\rightarrow} Q$⌝ is true at a possible world w iff Q is true at the possible worlds closest to w at which P is true (Stalnaker 1968; Lewis 1973). Moreover, most proponents of this view accept the following thesis, which is arguably needed to obtain an attractive logic of counterfactuals (Lewis 1973):

> *Weak Centering*. A possible world is at least as close to itself as any other possible world is to it.[7]

However, these assumptions are hard to square with anti-haecceitist counterpart theory.

Some preliminaries are required before this can be shown. Consider once more the possible world w_1 described at the end of Section 7.2. w_1 contains the two particles C and D. w_1 is qualitatively like the symmetrical world w_0 before t. At t, C disappears while D continues to exist forever. Anti-haecceitist counterpart theorists have good reason to think that C is a counterpart of D. To see why, let P_C and P_D be the conjunctions of all of C's and D's (relational and non-relational) qualitative properties, respectively. Note that the claim that something has P_C necessitates every qualitative claim that holds at w_1, and the same is true of the claim that something has P_D. Consider:

(9) D has P_C

(9) is surely metaphysically possible at w_1. For, all it would have taken for D to have P_C is for D to disappear at t while C continues to exist, and it seems to be true at w_1 that could have happened. There must therefore be a possible world w at which (9) holds. Moreover, w must be qualitatively indistinguishable from w_1. By anti-haecceitist lights, that means that all claims that are true at w also hold at w_1, including (9). According to counterpart theory, w_1 must therefore contain a counterpart c of D that has P_C. Since C is the only individual at w_1 that has P_C, it follows that C is a counterpart of D.[8]

Now suppose that what happens to one of the two particles at t does not causally affect what happens to the other at t. Then it should be true at w_1 that if D had disappeared at t, then C would still have disappeared at t. The following should therefore true at w_1 as well:

> If D had disappeared at t, then two particles would have disappeared at t.

It follows that, of all the worlds where it is true that D disappears at t, those where two particles disappear at t are closer to w_1 than those where only one particle disappears at t. But given counterpart theory, that contradicts Weak Centering. Since C is a counterpart of D (as we saw before) and C disappears at t, the counterpart theorist has to say that it is true at w_1 that D disappears at t. Moreover, by Weak Centering, w_1 is at least as close to w_1 as any other world is to w_1. w_1 is therefore one of the worlds closest to w_1 where it is true that D disappears at t. And yet, at w_1 only one particle disappears at t.

Chance. Anti-haecceitist counterpart theory also seems to yield problematic consequences about physical chance. Note that it seems plausible that the sample space of the chance distribution at time t at a possible world w is the set of possible worlds that are like w up to t and conform to the natural laws of w after t. Call these worlds the "possibilities that are open at t" (at w) or the "open$_t$ possibilities" (at w). Moreover:

(10) The chance of P at t ($\mathrm{ch}_t(P)$, for short) equals the chance measure of the set of open$_t$ possibilities at which P is true.

Now consider possible world w_0 once more. To recall, w_0 contains no material objects except for the two particles A and B, both of which exist forever. Let t_0 and t_1 be times in the history of w_0, with t_0 being earlier than t_1. It seems that the following could be true at w_0:

(11) ch_{t_0}(A will disappear at t_1) = .5
(12) ch_{t_0}(B will disappear at t_1) = .5
(13) Whether A disappears at t_1 is probabilistically independent of whether B disappears at t_1.
(14) Any two open$_{t_0}$ possibilities where exactly one particle disappears at t_1 are qualitatively indistinguishable from each other.

Let D_A be the claim that A disappears at t_1 while B continues to exist, let D_B be the claim that B disappears at t_1 while A continues to exist, and let 1P be the set of all open$_{t_0}$ possibilities where exactly one particle disappears. (11)–(13) appear to entail:

(15) $\mathrm{ch}_{t_0}(D_A) = .25$
(16) $\mathrm{ch}_{t_0}(D_B) = .25$
(17) ch_{t_0}(Exactly one particle disappears at t_1) = .5.

However, (15)–(17) seem to conflict with the anti-haecceitist counterpart theorist's view. By (10) and (15), D_A is true at some of the worlds in 1P. Moreover, given (14), anti-haecceitists have to say that the same claims about A and B are true at any two worlds in 1P. It follows that D_A is true at *all* worlds in 1P. By analogous reasoning, D_B is true at all worlds in 1P as well. In other words, the open$_{t_0}$ D_A-possibilities are the same worlds as the open$_{t_0}$ D_B-possibilities—they include all and only the worlds in 1P. (Where w is any world in 1P, and where E is the particle at w that disappears at t_1 and F is the particle at w that continues to exist, $<E, F>$ is a counterpart both of $<A, B>$ and of $<B, A>$.) Given this conclusion, it follows from (10) and (15), and also from (10) and (16), that 1P has a chance measure of .25 at t_0. But that contradicts (10) and (17), which together entail that the chance measure of 1P at t_0 is .5.

Needless to say, the dialectic does not end here. There are several possible anti-haecceitist replies to these objections and possible haecceitist countermoves. I will only consider one response that anti-haecceitists might give. They could introduce entities that are more fine-grained than possible world, and which we may call "world descriptions." A *world description relative to a base world* $w^\star$ (a *world description*$_{w^\star}$, for short) is an ordered pair of a world w and a partial function from the individuals in w to individuals in $w^\star$ of which they are counterparts. (Consider the possible world w_1 where exactly one particle disappears. There are two world descriptions$_{w_0}$ involving w_1: $w_1^{(a)}$, which maps the disappearing particle in w_1 to A and the non-disappearing particle in w_1 to B, and $w_1^{(b)}$, which includes the opposite mapping.) Where $\varphi(x)$ is a formula containing x as its sole free variable, a is an individual at world $w^\star$, and $<w, f>$ is a world description$_{w^\star}$, a satisfies $\varphi(x)$ at $<w, f>$ iff w contains an object that satisfies $\varphi(x)$ and to which f assigns a. $\varphi(a)$ is true at $<w, f>$ iff a satisfies

$\varphi(x)$ at $<w,f>$. Analogous accounts can be given of satisfaction of a formula by an n-tuple of individuals at a world description, and of truth of a formula containing multiple individual constants at a world description. Even when the same qualitative claims hold at two world descriptions, they can differ in the claims about specific individuals that are true at them. (For instance, D_A is true at $w_1^{(a)}$ but not at $w_1^{(b)}$, while D_B is true at $w_1^{(b)}$ but not at $w_1^{(a)}$.) Anti-haecceitists might decide to replace worlds in their accounts of modal operators, actuality, chance, and counterfactuals with world descriptions. If this is done in the right way, it will yield an account (which we might call "world description theory") that does not face the difficulties described above.[9]

World description theory might make the right predictions, but one may wonder whether it is more adequately understood as abandoning anti-haecceitism than as vindicating it (Kment 2012: 601; also see Skow 2008). What made the debate about haecceitism interesting in the first place was the central role that possible worlds play in various philosophical theories, including theories about modality, actuality, counterfactuals, and chance. Philosophers were wondering:

(18) What are the entities like that figure in (the best versions of) these theories? If the same qualitative claims are true at two of these entities, does that mean that the same claims about specific individuals are true at them as well?

If the dispute about haecceitism is supposed to address this question, anti-haecceitism has to be understood as the thesis that the answer to question (18) is "yes" and haecceitism as the thesis that the answer is "no." However, on this way of understanding the debate, world description theory is a form of haecceitism. For it entails that the best theories of the modal operators, actuality, counterfactuals, and chance are formulated in terms of world descriptions; and it is *not* true that the same claims about specific individuals are true at two world descriptions whenever the same qualitative claims are true at them.

7.5 Modal contingentism

The rest of this chapter will consider two influential lines of argument leading from reflections on *de re* modality to one or the other of the following forms of modal contingentism.

World-existence contingentism. Some possible worlds are contingent existents.
Modal-operator contingentism. Some claims of the forms $\Box P$ and $\Diamond P$ are contingently true.

Modal-operator contingentists reject at least one of the following schemata of the modal logic S5, while their opponents accept both.

4 $\Box P \rightarrow \Box\Box P$
5 $\Diamond P \rightarrow \Box\Diamond P$

While some philosophers are happy to jettison S5, others have argued that S5 is the most attractive propositional modal logic and should be retained if possible (see, e.g., Williamson 2013).

The contingent existence of worlds. The first line of argument for modal contingentism (Fine 1977, 1985; Adams 1981; Stalnaker 2011; Kment 2014: sect. 4.5) starts from the following two assumptions.

(19) Many individuals exist contingently (for example, material objects).
(20) A possible world existentially depends on every individual that exists at it, i.e., if individual i exists at possible world w, then w fails to exist at possible worlds at which i fails to exist.

(19) and (20) can be used to motivate world-existence contingentism. Consider a contingently existing individual, such as yourself, and a world at which that individual exists, such as the actual world, @. By (20), @ does not exist at any possible world where you do not exist.

Those who accept (19) and (20) often differentiate between two notions of a claim's being true with respect to a possible world (Fine 1977, 1985; Adams 1981; Turner 2005; Stalnaker 2011: sect. 2.6; Kment 2014: sect. 4.6.2), which Robert Adams calls "truth in a world" and "truth at a world." On one way of explaining the distinction, the proposition that *P* is true *in* possible world *w* iff the proposition that *P* would have had the property of truth if *w* had been actualized. The proposition that *P* is true *at w* iff it is *actually* true that the proposition that *P* gives a correct partial description of what would have been the case if *w* had been actualized, i.e. iff it would have been that *P* if *w* had been actualized. One argument for the claim that the two notions can come apart starts from two further theses.

(21) A singular proposition about an individual existentially depends on that individual.
(22) It is impossible to have a property or stand in a relation to other things without existing.[10]

Let *Q* be the proposition that you do not exist and let *w* be a possible world where you do not exist. If *w* had been actualized, then (by (21)) *Q* would not have existed and therefore (by (22)) would not have had the property of truth. So, *Q* is not true *in w*, or *in* any other possible world where you fail to exist. (And since *Q* is the proposition that you do not exist, *Q* is not true in any possible world where you exist either. There is simply no possible world *in* which *Q* is true.) However, *Q* is true *at w*: if *w* had been actualized, then you would not have existed. These reflections also show that possibility cannot be defined as truth *in* all possible worlds but only as truth *at* all possible worlds. For, *Q* is possible but not true *in* any possible world.

If truth at a world is a relation between propositions and worlds, then the view outlined in the previous paragraph yields the result that sometimes a proposition is contingently true at a world (Kment 2014: 90). Let *w* be a possible world at which you do not exist and that does not existentially depend on you. The proposition that you do not exist (i.e., *Q*) is true at *w*. Let $w^\star$ be a possible world where *w* exists but you do not. (21) entails that at $w^\star$, *Q* does not exist. Given (22), it follows that at $w^\star$, *Q* does not stand in the truth-at relation to *w*. So, while *Q* is true at *w at the actual world*, *Q* is not true at *w at* $w^\star$.

It is controversial whether modal-operator contingentism is a consequence of the form of world-existence contingentism motivated by (19) and (20). The answer is arguably "yes" if we accept the following account of possibility:

(23) What it is for it to be possible that *P* is for there to be a possible world at which it is true that *P*.

For, (23) seems to commit us to (24).

(24) Necessarily, (it is possible that *P* iff there is a possible world at which it is true that *P*).

We saw that the form of world-existence contingentism motivated by (19) and (20) entails that it is contingent whether there is a possible world at which you exist. From that claim and (24), it follows that it is contingent whether it is possible that you exist.[11] (At the actual world, it is possible for you to exist, but at a possible world where you do not exist, it is not possible.) Modal-operator contingentism therefore comes out true. World-existence contingentists can avoid this conclusion by denying (24) and saying instead that, necessarily, $\Diamond P$ iff

there *actually* exists a possible world at which *P* (Stalnaker 2011). This account requires us to abandon (23).

The argument from (19) and (20) to world-existence contingentism can be resisted by rejecting one of its premises. Alvin Plantinga (1976, 1983) denies (20): The existence of a possible world *w* does not require the existence of the individuals that exist at *w*. It requires only the existence of the haecceities of these individuals (the haecceity of an individual is the property of being that individual). Moreover, he holds that an individual's haecceity does not existentially depend on that individual—haecceities can exist uninstantiated. Other philosophers reject (19) (Linsky and Zalta 1994, 1996; Williamson 1998, 1999, 2013). They argue that you could not have failed to exist. At best, you could have existed without being a concrete object.

Contingent modal profiles. The following argument for modal-operator contingentism is due to Hugh Chandler (1976) and Nathan Salmon (1979, 1982: 238–40). A table called "Ed" is made from the three equally large parts *A*, *B*, and *C*. Plausibly, Ed could not have been made from completely different parts. To fix ideas:

(25) Necessarily, if Ed exists, he is made from at least ⅔ of *ABC*.

However, Ed could surely have been made from *slightly* different parts, e.g. from *ABC* minus a few particles. Let us again fix ideas:

(26) Possibly, Ed is made from parts that include no more than ⅔ of *ABC*.

It is natural to think that (25) and (26) are instances of some general truths like the following:

Necessity. Necessarily, where *x* is any table and the *y*s are *x*'s parts, it is necessary that if *x* exists, *x* is made from at least ⅔ of the *y*s.

Tolerance. Necessarily, where *x* is any table and the *y*s are *x*'s parts, it is possible that *x* is made from parts that include no more than ⅔ of the *y*s.[12]

Given Tolerasnce, there might be objects *D* and *E* that do not overlap *ABC*, such that:

(27) ◊Ed is made from *BCD*
(28) □(If Ed is made from *BCD*, then ◊ Ed is made from *CDE*)

Let us suppose that such objects *D* and *E* exist. Given standard assumptions in modal logic, (27) and (28) entail (29).

(29) ◊ ◊ Ed is made from *CDE*.

However, assuming that "□" and "◊" are duals, we can infer (30) from Necessity and the assumption that Ed is made from *ABC*.

(30) ~◊ Ed is made from *CDE*.

(29) and (30) constitute a counterexample to the schema $\Diamond\Diamond P \rightarrow \Diamond P$, which (if "◊" and "□" are duals) is equivalent to **4**. **5** is violated as well: by Tolerance, ◊Ed is made from *BCD*. However, there are possible worlds where Ed is made from *EAB*, and it follows from Necessity that at these worlds, ~◊Ed is made from *BCD*. Therefore, ~□◊Ed is made from *BCD*.

The form of modal-operator contingentism motivated by Necessity and Tolerance arguably faces a problem usually associated with Humean frequentism about physical chance (Kment 2018). Frequentists believe that the chances at t are determined by the frequencies of different patterns of events throughout history. Since these global frequencies depend in part on the outcomes of post-t chance processes, it might be chancy at t what the chances at t are. For example, it could be that $ch_t(P) = x$, but that there is a non-zero chance at t that $ch_t(P) \neq x$. We can describe such cases by saying that the chances at t "undermine themselves." Allowing for the possibility of undermining yields a number of implausible consequences (Lewis 1994; Thau 1994; Hall 1994).

Necessity and Tolerance might also give rise to the possibility of undermining. For, once we accept that Ed was contingently made from the parts he was in fact made from, it is a small step to the conclusion that there could at some time have been a non-zero chance that he would be made from different parts. Suppose that there was some time t just before Ed was made from *ABC* such that (31) and (32) hold.

(31) ch_t(Ed will be made from *BCD*) = .1
(32) ch_t(Ed will be made from *EAB*) = .1

Necessity entails that at possible worlds where Ed is made from *BCD*, ~◊Ed is made from *EAB*. Therefore:

(33) At possible worlds where Ed is made from *BCD*, ch_t(Ed is made from *EAB*) = 0.

(31) and (33) entail:

(34) $ch_t(ch_t$(Ed will be made from *EAB*) = 0) = .1

(32) and (34) together entail that the chances at t undermine themselves. Such undermining cases yield problematic consequences similar to some of those that beset frequentism.

The argument from the Ed example to modal operator contingentism can be rejected without abandoning (25) or (26), provided one resists the generalization from (25) and (26) to Necessity and Tolerance. If one rejects Necessity and Tolerance, one can say that (25) and (26) are necessary truths: even at possible worlds where Ed is made from *BCD* rather than from *ABC*, it is not possible for him to be made from *CDE* (contrary to Tolerance), though it *is* possible for him to be made from *EAB* (contrary to Necessity). (28) and (29) are therefore false and there is no violation of **4** or **5**. If one rejects Necessity and Tolerance, one needs another general and systematic account of the modal profiles of objects like Ed that is consistent with (25) and (26). One such account is 'maximal multi-thingism,' the view that, necessarily, for any material object x and any set of x's non-modal properties that meets certain minimal conditions, there is an object co-located with x that has all and only the properties in this set necessarily (Leslie 2011: 287–90, Kment 2012: 194–7).[13] On this view, Ed is only one of countless table-shaped objects that occupy the same location but differ in their modal profiles.

Notes

1 This definition is essentially David Lewis's (1986: sect. 4.4). The term "haecceitism" is due to David Kaplan (1975: 722–3).
2 Lewis uses the phrases "*w* represents *a* as being *F*" (Lewis 1986) or "*a* vicariously satisfies "*Fx*" at *w*" (Lewis 1973) to mean that *a* satisfies "*Fx*" at *w*.

3 For the sake of simplicity, I have sketched a version of counterpart theory that presupposes Lewisian realism about possible worlds, i.e. the view that non-actual possible worlds are, to simplify somewhat, spatio-temporally extended individuals (like the universe we inhabit) that are spatio-temporally isolated from the actual world (Lewis 1986: ch. 1). (To see that the account assumes Lewisian realism, consider "You could have had a six-feet-tall chartreuse horse." That sentence is true. On the account I sketched, it follows that (35) holds.

(35) Some possible world *w* contains a counterpart of you who has a six-feet-tall chartreuse horse.

(35) entails that there is a chartreuse horse. However, there does not actually exist such a horse, nor is there such a horse at any spatio-temporal distance from us. So, how could (35) be true unless there were some non-actual possible world containing a six-feet-tall chartreuse horse? This world would have to be a spatio-temporally extended object and be spatio-temporally disconnected from us.) A suitably modified version of the counterpart theory I described is, however, compatible with non-realist accounts of possible worlds. See Kment 2012: 584–5.

4 For Lewis's version of counterpart theory, see Lewis (1968, 1971, 1973, 1986). For other versions, see Forbes (1982, 1987, 1990), Ramachandran (1989, 1990a, 1990b). For discussions, see Hazen (1979), Dorr (2005), M. Fara and Williamson (2005), D. Fara (2009), and Russell (2013).

5 See Cameron (2009). Also cp. Sider (2003, 2011: ch. 12). For another defense of conceptualism in the face of the data unearthed by Kripke, see Sidelle (1989).

6 For further discussion of this problem, related difficulties, and possible solutions, see the works cited in note 4.

7 Weak Centering is not completely uncontroversial. See Briggs (2012).

8 According to Lewis, *x* at *w* is a counterpart of *y* only if no individual at *w* is more qualitatively similar to *y* overall than *x* is. Presumably, *D* at w_1 is more qualitatively similar to *D* overall than *C* is. So, does the claim that *C* is a counterpart of *D* contradict Lewis's account? No. The overall similarity between two things is determined by some method of weighing their similarities and dissimilarities against each other. This method might assign zero weight to the similarities that *D* has to itself but not to *C*, so that *C* comes out as being as similar to *D* overall as *D* is to itself.

9 Strategies along these lines for solving the aforementioned problems (sometimes called "cheap haecceitism") are discussed in Lewis (1986: ch. 4), Skow (2008), Kment (2012: sect. 5), and Russell (2015), among others.

10 This principle is closely related to what Williamson (2013: sect. 4.1) calls the "Being Constraint" and what Plantinga (1983) calls "serious actualism."

11 For discussions of arguments along these lines, see Adams (1981: 28–32), McMichael (1983), Armstrong (1989: 56–63), Fitch (1996: 63–5), Bennett (2005: sect. 7), and Kment (2014: 105).

12 The name "Tolerance" is borrowed from Forbes (1985).

13 It is difficult to formulate this view precisely. For discussion, see Fairchild (2019).

References

Adams, R. (1979): "Primitive Thisness and Primitive Identity," *Journal of Philosophy* 76, pp. 5–26.

————. (1981) "Actualism and Thisness," *Synthese* 49, pp. 3–41.

Armstrong, D. (1989) *A Combinatorial Theory of Possibility*, New York: Cambridge.

Barcan (Marcus), R. (1947) "A Functional Calculus of First Order Based on Strict Implication," *Journal of Symbolic Logic* 11, pp. 1–16.

Bennett, K. (2005) "Two Axes of Actualism," *Philosophical Review* 114, pp. 297–326.

Briggs, R. (2012) "Interventionist Counterfactuals," *Philosophical Studies* 160, pp. 139–66.

Cameron, R. (2009) "What's Metaphysical About Metaphysical Necessity?" *Philosophy and Phenomenological Research* 79, pp. 1–16.

Chandler, H. (1976) "Plantinga and the Contingently Possible," *Analysis* 36, pp. 106–9.

Dorr, C. (2005) "Propositions and Counterpart Theory," *Analysis* 65, pp. 210–18.

Fairchild, M. (2019) "The Barest Flutter of the Smallest Leaf: Understanding Material Plenitude," *Philosophical Review* 128, pp. 143–78.

Fara, D. (2009) "Dear Haecceitism," *Erkenntnis* 70, pp. 285–97.

Fara, M. and T. Williamson (2005) "Counterparts and Actuality," *Mind* 114, pp. 1–30.

Fine, K. (1977) "Postscript: Prior on the Construction of Possible Worlds and Instants," in Prior, A. and K. Fine (eds.), *Worlds, Times, and Selves*, Amherst: University of Massachusetts Press, pp. 116–61.

———. (1985) "Plantinga on the Reduction of Possibilist Discourse," in J. Tomberlin and P. van Inwagen (eds.), *Alvin Plantinga*. Dordrecht: Reidel, pp. 145–86.
Fitch, G. W. (1996) "In Defense of Aristotelian Actualism," *Philosophical Perspectives* 10, pp. 53–71.
Forbes, G. (1982) "Canonical Counterpart Theory," *Analysis* 42, pp. 33–37.
———. (1985) *The Metaphysics of Modality*, Oxford: Oxford University Press.
———. (1987) "Free and Classical Counterparts: Response to Lewis," *Analysis* 47, pp. 147–52.
———. (1990): "Counterparts, Logic and Metaphysics: Reply to Ramachandran," *Analysis* 50: 167–73.
Hall, N. (1994) "Correcting the Guide to Objective Chance," *Mind* 103, pp. 505–17.
Hazen, A. (1979) "Counterpart-Theoretic Semantics for Modal Logic," *Journal of Philosophy* 76, pp. 319–38.
Kaplan, D. (1975) "How to Russell a Frege-Church," *Journal of Philosophy* 72, pp. 716–29.
Kment, B. (2012) "Haecceitism, Chance, and Counterfactuals," *Philosophical Review* 121, pp. 573–609.
———. (2014) *Modality and Explanatory Reasoning*, Oxford: Oxford University Press.
——— (2018) "Chance and the Structure of Modal Space," *Mind* 127, pp. 633–65.
Kripke, S. (1963) "Semantical Considerations on Modal Logic," *Acta Philosophica Fennica* 16, pp. 83–94.
———. (1980) *Naming and Necessity*, Cambridge, MA: Harvard University Press.
Leslie, S.-J. (2011) "Essence, Plenitude, and Paradox," *Philosophical Perspectives* 25, pp. 277–96.
Lewis, D. (1968) "Counterpart Theory and Quantified Modal Logic," *Journal of Philosophy* 65, pp. 113–26.
———. (1971) "Counterparts of Persons and Their Bodies," *Journal of Philosophy* 68, pp. 203–211.
———. (1973) *Counterfactuals*, Oxford: Blackwell.
———. (1986) *On the Plurality of Worlds*, Oxford: Blackwell.
———. (1994) "Humean Supervenience Debugged," *Mind* 103, pp. 473–90.
Linsky, B. and E. Zalta (1994) "In Defense of the Simplest Quantified Modal Logic," *Philosophical Perspectives* 8, pp. 431–58.
——— (1996) "In Defense of the Contingently Concrete," *Philosophical Studies* 84, pp. 283–94.
McMichael, A. (1983) "A Problem for Actualism about Possible Worlds," *Philosophical Review* 92, pp. 49–66.
Plantinga, A. (1976) "Actualism and Possible Worlds," *Theoria* 42, pp. 139–60.
———. (1983) "On Existentialism," *Philosophical Studies* 44, pp. 1–20.
Quine, W.V. O. (1953) "Three Grades of Modal Involvement," in *Proceedings of the XIth International Congress of Philosophy*, Vol. 14, Amsterdam: North-Holland, pp. 65–81.
Ramachandran, M. (1989) "An Alternative Translation Scheme for Counterpart Theory," *Analysis* 49, pp. 131–41.
———. (1990a) "Contingent Identity in Counterpart Theory," *Analysis* 50, pp. 163–66.
———. (1990b) "Unsuccessful Revisions of CCT," *Analysis* 50, pp. 173–77.
Russell, J. (2013): "Actuality for Counterpart Theorists," *Mind* 122: 85–134.
——— (2015) "Possible Worlds and the Objective World," *Philosophy and Phenomenological Research* 90: 389–422.
Salmon, N. (1979) "How *Not* to Derive Essentialism from the Theory of Reference," *Journal of Philosophy* 76, pp. 703–25.
———. (1982) *Reference and Essence*, Princeton, NJ: Princeton University Press.
Sidelle, A. (1989) *Necessity, Essence, and Individuation: A Defense of Conventionalism*, Cornell: Cornell University Press.
Sider, T. (2003) "Reductive Theories of Modality," in M. Loux and D. Zimmerman (eds.), *The Oxford Handbook of Metaphysics*, Oxford: Oxford University Press, pp. 180–208.
———. (2011) *Writing the Book of the World*, New York: Oxford University Press.
Skow, B. (2008) "Haecceitism, Anti-Haecceitism and Possible Worlds," *Philosophical Quarterly* 58, pp. 98–107.
Stalnaker, R. (1968) "A Theory of Conditionals," *Americal Philosophical Quarterly*, pp. 98–112.
———. (2011) *Mere Possibilities: Metaphysical Foundations of Modal Semantics*, Princeton, NJ: Princeton University Press.
Thau, M. (1994) "Undermining and Admissibility," *Mind* 103, pp. 491–503.
Turner, J. (2005) "Strong and Weak Possibility," *Philosophical Studies* 125, pp. 191–217.
Williamson, T. (1998) "Bare Possibilia," *Erkenntnis* 48, pp. 257–73.
———. (1999) "Existence and Contingency," *The Aristotelian Society: Supplementary* 73, pp. 181–203.
———. (2013) *Modal Logic as Metaphysics*, Oxford: Oxford University Press.
Wittgenstein, L. (1953) Philosophical Investigations, ed. G.E.M. Anscombe and R. Rhees, trans. G.E.M. Anscombe, Oxford: Blackwell.

Chapter 8

RELATIVIZED METAPHYSICAL MODALITY

Index and context

Benj Hellie, Adam Russell Murray, and Jessica M. Wilson

8.1 Introduction

A familiar and useful representation of modality invokes a space of *possible worlds*, and *quantification* over it: necessity and possibility are represented, respectively, with *universal* and *existential* quantification—so that necessity is truth at all possible worlds; possibility, truth at some. While modality comes in many varieties, a familiar thought gives a singular position to *metaphysical* modality, as the broadest objective variety, pertaining to 'what could not have been otherwise no matter what' (Burgess 2009, 46). Combining these ideas, metaphysical modality is represented as absolutely unrestricted quantification over the space of possible worlds.

This representation is called on by an austere, 'Carnapean' semantic analysis of the contribution of a modal operator, used metaphysically, to possible-worlds truth-conditions: in such uses, *necessarily*, φ is true in a world just if φ is true in every world, while *possibly*, φ is true in a world just if φ is true in some world (compare Carnap 1947, **41-1**, 183). A 'classical' semantics for quantified modal languages combines this analysis with comparably austere analyses of reference and quantification (compare Carnap 1947, **41-2**, 184–5). The result is attractive for its inherent simplicity; moreover, its predictions arguably capture strongly held *logical* intuitions about the various underlying subject matters—the phenomenon of modality, and its interactions with individuation and with ontology.

But various literatures, each spanning many decades, discuss an array of puzzling phenomena involving all these subject matters, leading to *metaphysical* challenges to the classical theory. Preserving the classical theory requires revising intuitive metaphysics; conversely, preserving intuitive metaphysics requires revising the classical theory. Unfortunately, these literatures presume an alliance between intuitive logic and the classical theory—thereby presenting an unappealing choice of revisionisms, in logic or in metaphysics.

Fortunately, the presumed alliance is false, holding only under a implicitly assumed 'modal absoluteness' to modality, individuation, and ontology: if instead, as we propose, what is metaphysically possible or necessary is *relativized to which world is actual*, the puzzling phenomena demand revisionism in neither logic nor metaphysics.

Our approach—*relativized metaphysical modality* (RMM: Murray and Wilson 2012; Murray 2017)—is based around a *double-indexing semantics* for modal languages, on which truth is twice over relativized to a possible world. This basis is interpreted with a metaphysics and a

pragmatics. Metaphysically, with a *moderate modal naturalism*, on which metaphysical possibility (associated with a first dimension of world-relativity—the 'horizontal' dimension, in a familiar picture: compare Stalnaker 1978) is explained in part by abstract metaphysical principles and in part by actual matters of contingent categorical fact (associated with a second, 'vertical', dimension of world-relativity). Pragmatically, with a *Context–Index* representation of modal reasoning (Lewis 1980a; compare Kaplan 1977), on which the second dimension of world-relativity is fixed independently of language, by the actual contingent categorical facts (our 'context'), while anything comprehensible as a 'modality' is restricted de jure to the first dimension of world-relativity (the 'indices' generable through the semantic powers of the language): our access to the second dimension is exclusively through an imaginative exercise of falsely supposing things to be other than they are—by 'considering worlds as actual' (as contrasted with 'as counterfactual', which requires no false supposition and is associated with modal reasoning properly so-called).

RMM may be heuristically understood as a substantive metaphysical repurposing of more familiar 'epistemic two-dimensional' (E2D: compare Chalmers 2005) semantical machinery, which similarly distinguishes between 'consideration' of a possible world 'as actual' and 'as counterfactual' (Jackson 1994). But the concerns of the approaches are very different. E2D focuses on the 'apriority' of a sentence (representing this status with its truth in every possible world when considered as actual); RMM, by contrast, focuses on what is required of semantics and pragmatics to make sense of moderate modal naturalism without revising intuitive logic (and accords no distinctive role to quantification over worlds considered as actual).

Remaining sections of this entry progress as follows. Section 8.2 describes the 'classical' approach to semantic theorizing, analyzing a nested series of modal fragments: Section 8.2.1 treats a basic propositional modal fragment; Section 8.2.2 adds predicate–term syntax; Section 8.2.3 adds quantification. Section 8.3 describes the metaphysical puzzles to each of the classical proposals in series: in each case, the puzzle motivates a familiar 'postclassical' semantics to preserve intuitive metaphysics by revising intuitive logic. In Section 8.3.1, a puzzle concerning laws of nature motivates a postclassical 'accessibility' semantics, revising intuitive 'S5' modal logic; in Section 8.3.2, a puzzle concerning individual essence motivates postclassical treatments with either accessibility or 'counterparts', revising S5 logic either way; in Section 8.3.3, a puzzle concerning ontology motivates postclassical treatments with 'contingent domains', revising intuitive 'Barcanite' logic (compare Barcan 1946).

The puzzles in Section 8.3 involve a common presupposition, to be labeled the 'In-Light Principle' (ILP), which identifies possibility *in light of which* and possibility *that* such-and-such: going through the puzzles, we identify where the ILP is exploited. Section 8.4 develops RMM as a way to reject the ILP and solve the puzzles: Section 8.4.1 calls on double-indexation to restructure the postclassical semantic proposals, identifying possibility *that* and *in light of* with the 'horizontal' and 'vertical dimension' indices, respectively; Section 8.4.2 sketches the association with moderate modal naturalism, while Section 8.4.3 turns to Context–Index pragmatics to rebut an objection. Section 8.5 concludes with historical speculation as to why the RMM solution has so long remained elusive.[1]

8.2 The classical theory

We begin by giving precise expression to our label, the 'classical theory'. This is an account of the truth-conditions and logical consequence relations for a series of increasingly complex *modal fragments*: a *propositional* fragment; and two predicate fragments, a *referential* fragment and a *quantificational* fragment.

We use this notation for expressions of the object-language: φ and ψ are 'schematic', to be substituted with any formula of the object language; $\neg$, $\wedge$, and $\supset$ are used for *negation*, *conjunction*, and the *material conditional* (with $\varphi \wedge \psi$, 'φ and ψ' abbreviating $\neg(\varphi \supset \neg\psi)$, 'it is not the case that if φ, then not-ψ'); $\Box$ and $\Diamond$ for (metaphysical) *necessity* and *possibility* (with $\Diamond\varphi$, 'possibly φ' abbreviating $\neg\Box\neg\varphi$, 'it is not the case that, necessarily, not-φ'); Γ^n and τ are schematic, to be substituted with any n-place predicate and any term, respectively; and $\forall$ and $\exists$ are used for ('metaphysical'—understood also with maximal breadth) *universal* and *existential* quantifiers: a *quantifying operator* is of form $(\forall\tau)$ or $(\exists\tau)$ (with $\exists\tau.\varphi$, 'some τ is such that φ' abbreviating $\neg\forall\tau.\neg\varphi$, 'it is not the case that every τ is such that not-φ').

We use $\models$ for a metalinguistic truth-predicate of formulae of the object-language, decorated with various sub- and superscripts to represent relativization of truth to various parameters; $\vdash$ means the relation of logical consequence, holding between a set of zero or more *premiss* formulae of the object language and a *conclusion* formula of the object language, and analyzed as truth-preservation across all parameters to which truth is relativized;[2] we strike through these turnstiles to negate them; and we use ':=' to abbreviate *is defined as*. All analyses to come analyze $\neg\varphi$ as true (relative to an appropriate list of parameters) just if φ isn't, and $\varphi \supset \psi$ as true (again, appropriately relativized) just if either φ isn't or ψ is.

Space requires us to blur over without comment various important details and background assumptions.[3]

8.2.1 A propositional modal fragment

The formulae of this fragment are generated on a base set of simple formulae, and include just those complex formulae $\neg\varphi$, $\varphi \supset \psi$, and $\Box\varphi$.

We postulate in the semantic machinery a set of possible worlds, W; metalinguistic variables like 'w' range over this set.

Truth is relativized to a possible world, with $\models_w \varphi$ meaning 'φ is true in (according to; relative to) the world w'.

The classical truth-condition for $\Box\varphi$ (*necessarily*, φ) is as follows (recall that $\Diamond\varphi$ (*possibly*, φ) abbreviates $\neg\Box\neg\varphi$, and truth-conditions for *negation* and the *material conditional* are as earlier):

$$\left(\Box_{cl}\right) \qquad \models_w \Box\varphi := \text{for every possible world } w', \models_{w'} \varphi$$

('Necessarily, φ' is true in a world just if φ is true in every world.)

Familiarly, analysis $(\Box_{cl})$ validates the following principles:

T	$\Box\varphi \supset \varphi$	(if, necessarily, φ, then φ)
4	$\Diamond\Diamond\varphi \supset \Diamond\varphi$	(if it is possibly possible that φ, then it is possible that φ)
5	$\Diamond\Box\varphi \supset \Box\varphi$	(if it is possibly necessary that φ, then it is necessary that φ)

When these are added as axioms to a certain basis,[4] the result is a logic known as **S5**. **S5** deserves special distinction as the logic that is arguably (compare Williamson 2013, 44; Williamson 2016) definitive of our concept of *metaphysical* modality, as it makes the metaphysical modal facts themselves a matter of metaphysical necessity—perhaps in contrast with narrower modalities, for which (say) the *morally* possible or necessary acts may depend on which morally possible acts have been performed.

8.2.2 A referential modal fragment

The simple expressions of this fragment include a set of terms and a set of *n*-place predicates (including a two-place identity predicate '='), with simple formulae prefixing an *n*-place predicate to a string of *n* terms; the complex formulae include those with forms as from the propositional modal fragment.

We add to the semantic machinery a domain D of individuals, with metalinguistic variables like 'd' ranging over D; an *individual concept* k maps W to D, selecting for each world $w \in W$ a unique individual $k_w \in D$.

The meaning of a term τ is given by an individual concept k^τ, with k^τ_w as the 'w-designatum' of τ (the entity τ refers to *in*, or *relative to* w: τ is a 'rigid designator', therefore, only if k^τ is a constant function—only if there is a unique individual $d \in D$ such that the individual concept for τ maps each possible world to that d). And the meaning of a predicate is given by a world-relative relation of *satisfaction in* w between an *n*-place predicate and an *n*-place sequence of individuals—intuitively, the pair of Caesar and Cleo satisfy the two-place predicate *loves* in w just if, in w, Caesar loves Cleo—with exactly the pairs $\langle d,d \rangle$ satisfying '=' in w, for every world w.

Truth-conditions for complex formulae follow the analyses for the propositional fragment; for a simple formula, the truth-condition is this:

- $\models_w \Gamma^n\tau^1 \ldots \tau^n$:= the sequence $k^{\tau^1}_w, \ldots, k^{\tau^n}_w$ satisfies Γ^n in w

(For example, 'the McCosh Professor has greeted the Pierce Chair' is true in a world just if, in that world, the entity which, in that world, is the McCosh Professor has greeted the entity which, in that world, is the Pierce Chair.)

Quite clearly, **S5** continues to be validated.[5]

8.2.3 A quantificational modal fragment

All formulae of the referential fragment are formulae of this fragment; in addition, its formulae include those prefixing a quantifying operator to a formula.

We add to the earlier semantic machinery the device of the *assignment*: an assignment g maps each term τ to an individual concept, notated alternatively as g^τ or as k^τ_g. (A *variable* term τ, accordingly, is one for which sometimes $k^\tau_g \neq k^\tau_{g'}$; otherwise, τ is *constant*. If there are constant terms, there are fewer assignments than functions from terms to individual concepts.) An assignment g' is a τ-*variant* of g, notated $g' \sim^\tau g$, just if g' does not differ from g at any term other than τ.

Truth is relativized to an assignment and a world, with $\models_{g,w} \varphi$ meaning 'φ, relative to g, is true in w'. Predicate-satisfaction continues to be relativized only to a world. The 'designatum' of τ, as pertains to the truth-condition of a simple sentence, is relativized both to an assignment and a world: namely, for a given g and w, $k^\tau_{g,w}$. Accordingly, a simple sentence has this truth-condition:

- $\models_{g,w} \Gamma^n\tau^1 \ldots \tau^n$:= the sequence $k^{\tau^1}_{g,w}, \ldots, k^{\tau^n}_{g,w}$ satisfies Γ^n in w

(For example, 'x loves y' is true in a world, relative to an assignment just if: for that individual d such that the individual concept for 'x', relative to that assignment, maps that world to d, and for that individual d' such that the individual concept for 'y', relative to that assignment, maps that

world to d', are such that in that world, d loves d'.) Truth-conditions for complex Booleanized or modalized formulae are merely 'up-parametrized' from the referential fragment.

For a complex quantified formula, the truth-condition is:

$$\left(\forall_{cl}\right) \qquad \vDash_{g,w} \forall\tau.\varphi := \text{for every } g' \sim^{\tau} g, \vDash_{g',w} \varphi$$

('Every τ is such that φ' is true in a world relative to an assignment just if φ is true in that world relative to every τ-variant of that assignment.)

The analysis preserves the validity of **S5**. With quantification analyzed with $(\forall_{cl})$, the following are also validated:

$$BF \qquad \forall\tau.\Box\varphi \supset \Box\forall\tau.\varphi$$

(if everything is necessarily such that φ, then necessarily everything is such that φ)

$$CBF \qquad \Box\forall\tau.\varphi \supset \forall\tau.\Box\varphi$$

(if necessarily everything is such that φ, then everything is necessarily such that φ)

Logics adding *BF*, *CBF*, and the **S5** principles and rules to the principles and rules of non-modal predicate logic stand in an analogous position to **S5**, as arguably (compare Linsky and Zalta 1994, 1996; Williamson 1998, 2013) definitive of the interaction of *metaphysical modality* and *metaphysical quantification*—objectual quantification in the broadest possible sense.[6]

8.3 Metaphysical challenges to the classical theory

A distinctive metaphysical puzzle challenges the classical treatment of each fragment: for the propositional fragment, the puzzle concerns laws of nature; for the referential fragment, individual essence; for the quantificational fragment, ontology. Preserving metaphysical intuition requires revising the classical analyses. The 'postclassical' analyses proposed in the literature, unfortunately, revise the attractive logical predictions of the classical theory.[7]

We set up these puzzles to illustrate the reliance of each on the *In-Light Principle* (ILP): namely, that possibility *in light of which*, φ, is possibility *that* φ. For the puzzles from *laws* and *essence*, the ILP maintains that possibility in light of possibility is possible possibility; for the puzzle from *ontology*, that existence or nonexistence in light of possibility is possible existence or nonexistence. Without the ILP, logic and metaphysics do not clash. The next section explains the RMM strategy for rejecting the ILP: to foreshadow, we associate possibility *in light of* with 'consideration as actual', possibility *that* with 'consideration as counterfactual'.

8.3.1 Laws of nature and the propositional fragment

8.3.1.1 Nomological necessity and sensitivity?

Are the laws of nature metaphysically necessary? Is it true that, if the laws of nature require that φ, it is metaphysically necessary that φ? (We will assume that nothing metaphysically impossible can be permitted by the laws of nature.)

Necessitarians say yes, citing the explanatory power of appeals to law (Loewer 1996, 2012; Fine 2005, 247): if we answer the question 'Why, if it goes up, will it come down?' with 'It is a law of

nature that what goes up comes down', necessitarianism avoids the further question 'Yeah, but what distinguishes this case from one of those cases where it is a law of nature that what goes up comes down, but things can go up without coming down?'—a peculiar question, best avoided: to the credit of necessitarianism.

But a powerful challenge comes from the idea that the laws of nature should be sensitive to the 'categorical' facts on the ground. Leading theories of laws incorporate this *sensitivity* doctrine: according to *best-systems* theory (Lewis 1973, sect. 3.3), the laws are the simplest, strongest systematization of the categorical facts; according to a broadly *Aristotelian* approach (Shoemaker 1980), the laws are generated by the categorical fact that exactly *these* properties are instantiated. Each of these theories predicts a phenomenon we may call *undermining*[8]—categorical facts yielding laws permitting categorical facts yielding different laws, which disagree with the earlier laws over what is possible.

Undermining comes in two types. Type (i), with possibly possible impossibilities: categorical facts F yielding laws L permitting categorical facts F' yielding different laws L' permitting categorical facts F'', where L forbid F''; so for the necessitarian, what counts (*in light of* facts F and their laws L) as an impossibility (namely, the facts F'') sometimes counts as a possibility *in light of* a possibility (namely, a possibility *in light of* the laws L' generated by the possible facts F')—so if (by ILP) a possibility *in light of* a possibility is a possible possibility, an impossibility is sometimes a possible possibility. Type (ii), with possibly impossible possibilities: categorical facts F yielding laws L permitting categorical facts F' and F'', yielding respectively laws L' and L'', where L' do not permit F''; so for the necessitarian, what counts (*in light of* the facts F and their laws L) as a possibility (namely, the facts F'') sometimes counts as an impossibility *in light of* a possibility (namely, an impossibility *in light of* the laws L' generated by the possible facts F')—so if (by ILP) an impossibility *in light of* a possibility is a possible impossibility, a possibility is sometimes a possible impossibility.

The classical truth-conditions, recall, validate *4* (recall, $\Diamond\Diamond\varphi \supset \Diamond\varphi$), which prohibits possibly possible impossibilities and type (i) undermining, and *5* (recall, $\Diamond\Box\varphi \supset \Box\varphi$), which prohibits possibily impossible possibilities and type (ii) undermining. But sensitivity predicts both type (i) and type (ii) undermining. So the following would seem to be jointly incompatible: classical truth-conditions; necessitarianism; sensitivity. Strategies to avoid the conflict would seem to be exhausted by: *metaphysical revisionism*, abandoning either necessitarianism (losing its explanatory advantages) or sensitivity (losing contact with leading theories of laws); or adopting nonclassical truth-conditions.

8.3.1.2 Postclassical semantics: accessibility

One 'postclassical' approach (adapting an approach to the essence puzzle, to be discussed in Section 3.2[9]) adopts *accessibilist* truth-conditions. Let A be a 'metaphysical accessibility' relation between worlds (with the intended interpretation that wAw'—w' is *accessible from* w, or w' is *w-accessible*—just if w' is metaphysically possible *in light of* w); then the classical analysis ($\Box_{cl}$) is replaced with the following:

$$\left(\Box_A\right) \qquad \vDash_w \Box\varphi := \text{for every world } w' \text{ such that } w\,Aw',\ \vDash_{w'} \varphi$$

('Necessarily, φ' is true in a world just if φ is true in every world *accessible from* that world.)

If it is assumed also that wAw' (w' is accessible from w) only if the w-laws allow w', necessitarianism combines with ($\Box_A$) to weaken the condition of the classical necessitarian—namely,

the *w*-laws require φ only if φ holds in every *w-accessible* world. This is compatible with undermining: type (i) shows that *A* is not transitive (sometimes *wAw′Aw″* but not *wAw″*—*w* accesses *w′* which accesses *w″*, but *w* does not access *w″*), while type (ii) shows that *A* is not 'Euclidean' (sometimes *wAw′* and *wAw″* but not *w′Aw″*—*w* accesses both *w′* and *w″* but *w′* does not access *w″*). But for nontransitive *A*, ($\Box_A$) invalidates the *4* principle; and for non-Euclidian *A*, ($\Box_A$) invalidates the *5* principle—an undesirable *logical revisionism*.

8.3.2 Essence and the referential fragment

8.3.2.1 'Chisholm's Paradox': moderate centering origin essentialism?

Are facts about *individual essence* metaphysically necessary? Talk of essence applies to features of a thing which are, roughly speaking, 'definitive of it'. Whatever this may amount to, we assume, as is more-or-less standard, that if *a* is essentially *F*, then necessarily, *a* is *F*;[10] and we additionally ignore the (perhaps important: Fine 1994) distinction between essential properties and those held with 'mere' necessity, assuming therefore also the converse. So it is compatible with *a*'s essence that *a* is *F* just if it is metaphysically possible that *a* is *F*; so the initial question reduces to whether it is a necessary matter *whether* it is metaphysically possible that *a* is *F*. The classical analysis ($\Box_{cl}$) says yes; unfortunately, a consideration known as *Chisholm's Paradox* (CP: locus classicus, Chisholm 1967) seems to say no.

CP arises out of this *moderate centering* assumption concerning the essence of an artifact *a* originally constituted out of a certain quantity of matter *m*: (a) for every 'small enough' part *p* of *m*, it is not essential to *a* that its originally constituting matter include *p*; but (b) it is essential to *a* that for some 'big enough' part *q* of *m*, its originally constituting matter include *q*.[11]

The problem is that moderate centering yields cases analogous to the two types of undermining. Type (i), possibly possible impossibilities: sometimes a pair of small replacements in sequence make a big replacement—so sometimes what counts by (a) as a possibility *in light of* a possibility (namely, the second small replacement *in light of* the first small replacement) counts by (b) as an impossibility (namely, both small replacements together): so if (by ILP) a possibility *in light of* a possibility is a possible possibility, sometimes an impossibility is a possible possibility. Type (ii), possibly impossible possibilities: sometimes the right pair of alternative small replacements seem big to one another—so sometimes what count by (a) as possibilities (namely, each of the small replacements), count, *in light of* one another, by (b), as impossibilities (namely, because each sees the other as a big replacement): so if (by ILP) an impossibility *in light of* a possibility is a possible impossibility, sometimes a possibility is a possible impossibility. But as before, these are in conflict with the classical truth-conditions: type (i) cases conflict with the *4* principle, type (ii) with the *5* principle.

The literature canvases various ways of resolving the contradiction. *Metaphysical revisionists* revise moderate centering. There are three options. *Immoderate antiessentialists* (Chisholm himself) reject (b), maintaining, for example, that this desk could have originated in entirely different matter; no chain of small replacements makes a replacement too big for compatibility with its essence: implausibly. *Immoderate essentialists* reject even the 'dual' of (a) (namely, that for *some* small enough part *p* of *m*, it is not essential to *a* that its originally constituting matter include *p*), maintaining that there is *no* part *p* of *m* such that it is not essential to *a* that its originally constituting matter include *p*, maintaining, for example, that this desk could not have originated in matter differing even by one atom—no replacement is small enough to be compatible with its essence (compare Kripke 1980, 114n56): implausibly. And *moderate anti-centrists* (Williamson 1990, ch. 8) reject (a) while accepting its 'dual', maintaining that for *just some* part *p* of *m* is it inessential to *a*

that its originally constituting matter include *p*, maintaining that while there is a range of possible material origins for this desk, at some point in that range—just below the threshold level of bigness of change from a certain center—had the desk had that very origin, not even the slightest difference in origin along a certain dimension (that which would push over the edge to big change from the alleged center—though only a very small change from the hypothesized origin) would have been compatible with its essence: an asymmetry of unappealing arbitrariness.

8.3.2.2 Postclassical semantics: counterparts

The alternative is to adjust the classical truth-conditions. Those canvased in the literature on CP are 'postclassical', *logically revisionary* of **S5**: a famous debate between Salmon and Lewis concerns which such way is best.

Salmon (1981, 240–52: following Chandler 1976), adopts accessibilist truth-conditions for metaphysical modality, as in ($\Box_A$).

Lewis (1986b, 248: compare Lewis 1968, 28–9 on 'transitivity' and 'symmetry'; contrast Forbes 1984[12]) prefers instead to leave the domain of metaphysical modal quantification unrestricted. Instead, he makes an appeal to an approach unavailable in the propositional fragment, and proper to the referential fragment—the complicated *counterpart-theoretic* approach to the interaction between modality and designation (locus classicus, Lewis 1968).

Confusing stereotypes of counterpart theory abound, not all of which are germane to the logical revision demanded by CP;[13] our approach adapts Fara's (2008, 2012) interpretation of the central semantic hypothesis of counterpart theory. The gist is that the possibilities for an entity *d* are given by how its counterparts are, the possible possibilities by how the counterparts of its counterparts are; but being a counterpart of *d* is different from being a counterpart of a counterpart of *d*, with a counterpart sometimes failing to be a counterpart of a counterpart and a counterpart of a counterpart sometimes failing to be a counterpart—so the possibilities for *d* differ from the possible possibilities for *d*, with some possibilities failing to be possible possibilities (against *5*) and some possible possibilities failing to be possibilities (against *4*).

Now more explicitly: (i) truth is relativized to (and logical consequence quantifies over) not just a world *w* but also a set of individual concepts $\boldsymbol{k}$, where $\boldsymbol{k}$ is apt to determine, for every term τ, an individual concept k^{τ} from that set, which establishes (relative to $\boldsymbol{k}$) the designation for the term τ; (ii) a certain *counterpart function* κ maps a 'source' concept-set $\boldsymbol{k}$ and world *w* to a 'target' concept-set $\kappa(\boldsymbol{k},w)$; and (iii) a modal, in addition to shifting a 'source' world to a 'target' world, shifts a 'source' individual concept to a 'target' individual concept determined as the counterpart of the source individual concept relative to the source world. Accordingly, an outermost iterated modal shifts an initial source world *w* and concept-set $\boldsymbol{k}$ to a first target world w' and concept-set $\boldsymbol{k}' = \kappa(\boldsymbol{k},w)$; with w' and $\boldsymbol{k}'$ as the new source, the second-outermost modal shifts those to a second target world w'' and concept-set $\boldsymbol{k}'' = \kappa(\boldsymbol{k}',w') = \kappa(\kappa(\boldsymbol{k},w),w')$—thereby shifting the designative behavior of any embedded terms, and consequently which satisfaction-relations are relevant to the truth of an embedded simple sentence, from those which hold at the initial source world.

And formally:

- $\models_{\boldsymbol{k},w} \Gamma^n\tau^1\ldots\tau^n :=$ the sequence $k_w^{\tau^1},\ldots,k_w^{\tau^n}$ satisfies Γ^n in w

(For example, 'the McCosh Professor has greeted the Pierce Chair' is true in a world, relative to a concept-set just if, in that world, the entity which, in that world, by the lights of that set, is the

McCosh Professor has greeted the entity which, in that world, by the lights of that set, is the Pierce Chair.)

$$\left(\Box_{\kappa}\right) \qquad \vDash_{\boldsymbol{k},w} \Box\varphi := \text{for every } w', \vDash_{\kappa(\boldsymbol{k},w),w'} \varphi$$

('Necessarily, φ' is true in a source-world, relative to a concept-set just if, for every target-world, φ is true in the target-world, relative to the source-world counterpart of the concept-set.)

Assembling, $\vDash_{\mathbf{k},w} \Diamond\Gamma\tau$ ('possibly, τ is Γ' is true relative to a concept-set and source world) just if for some w', $\kappa(k^{\tau},w)_{w'}$ satisfies Γ in w'; and $\vDash_{\mathbf{k},w} \Diamond\Diamond\Gamma\tau$ just if for some w' and w'', $\kappa(\kappa(k^{\tau},w),w')_{w''}$ satisfies Γ in w''. The second does not require the first: perhaps there is no w' such that the w'-counterpart of τ-in-w is Γ in w', but some w'' and w' such that the w''-counterpart of (the w'-counterpart of τ-in-w)-in-w' is Γ in w''—against the *4* principle. Similar counterexamples to the *5* principle are also available.

8.3.3 Ontology and the quantificational fragment

8.3.3.1 Are existence and nonexistence necessary?

Are facts about *ontology*—which individuals exist and do not exist—metaphysically necessary?

The answer can seem obvious: *no*—for reasons of two types. (1) *Existence* does not seem metaphysically necessary (compare Williamson 1998, 258). After all, for any given existing human, *a*, they were begat by their parents; evidently, the activity by *a*'s parents through which *a* was begat could have been avoided: there is a possibility that *a* was not begat; because *a*'s begetting is required for *a*'s existence, there is a possibility lacking a requirement of *a*'s existence; because a possibility lacking a requirement of *a*'s existence is a possibility *in light of* which *a* does not exist, there is a possibility *in light of* which *a* does not exist; so if (by ILP) a possibility *in light of* which *a* does not exist is a possibility of *a*'s nonexistence, there is a possibility of *a*'s nonexistence—in which case, something which exists (namely, *a*) possibly does not exist. (2) *Nonexistence* does not seem metaphysically necessary (compare Kripke 1963, 65–6; Williamson 1998, 258). After all, for some given pair of humans, *f* and *m*, who failed to beget, they could instead have begat: so it is a possibility that *f* and *m* beget; because begetting by human parents suffices for the existence of a human they beget, there is a possibility of a sufficiency for the existence of a human begat by *f* and *m*; because a possibility of a sufficiency for the existence of an *F* is a possibility *in light of* which an *F* exists, there is a possibility *in light of* which a human begat by *f* and *m* exists; so if (by ILP) a possibility *in light of* which an *F* exists is a possibility of the existence of an *F*, there is a possibility of the existence of a human begat by *f* and *m*; plausibly, whenever *f′* and *m′* are a pair of humans not both identical to *f* and *m*, nothing *f* and *m* could have begat could have been identical to anything *f′* and *m′* could have begat: so there is a possibility of the existence of a human nonidentical to any existing human; plausibly, any possible human is essentially human—in which case, possibly something exists which is distinct from each of those things which in fact exists.

The classical truth-conditions, recall, validate *CBF* ($\forall\tau.\Box\varphi \supset \Box\forall\tau.\varphi$—if everything is necessarily such that φ, then necessarily everything is such that φ), by which the conclusion of (1), that there is something which possibly does not exist, entails that possibly, there is something which fails to exist—but it is impossible that there is something which fails to exist: so the classical truth-conditions are incompatible with the reasoning in (1). They also, recall, validate *BF* ($\Box \forall \tau. \varphi \supset \forall \tau.\Box\varphi$—if necessarily everything is such that φ, then everything is necessarily such that φ), by which the conclusion of (2), that it is possible that there exists something distinct from

everything that exists, entails that there exists something which is possibly distinct from everything that exists—in which case that thing would have to be possibly distinct from itself, which is an impossibility: so the classical truth-conditions are incompatible with the reasoning in (2).

Preserving the classical truth-conditions requires some form or other of *metaphysical revisionism*. On (1), actual existence–possible nonexistence: perhaps no actual begetting could have been avoided—and our ordinary judgments of metaphysical modality are entirely unreliable: implausibly. Or perhaps no human requires begetting for their existence—we are not essentially human, and could have existed, unbegat, as 'nonconcrete' (Linsky and Zalta 1996; Williamson 1998, 266; Williamson 2013 passim): strangely. On (2), actual nonexistence–possible existence: perhaps every possible begetting actually occurs—again, overthrowing the reliability of ordinary modal judgments: implausibly. Or perhaps a begetting by human parents does not suffice for the existence of a human they beget—a merely possible begetting yields no novel particular, but only a novel rearrangement of qualities: a route to skepticism about our own existence. Or perhaps our parents are inessential to us: but this is of no help for cases which strictly extend the set of actual humans. Or perhaps merely possible humans exist actually as 'nonconcrete': again, strangely.

8.3.3.2 Postclassical semantics: world-relative domains

Avoiding metaphysical revisionism requires revising the classical semantics: in the literature, strategies for this are *logically revisionary*, invalidating *BF* and *CBF*. In intuitive terms, their validity stems from the 'absoluteness', or absence of world-relativity, of the individual domain on the classical approach: if a possible world in which certain parents beget requires a domain including their child, then the possibility of that world requires that child to inhabit the domain for any other possible world, including a world in which the parents do not beget; conversely, if a possible world in which certain parents do not beget requires a domain excluding their child, then the possibility of that world excludes the child from the domain for any other possible world, including a world in which the parents beget. The canonical way to avoid this, then, relativizes the individual domain to a possible world—permitting the domain for the begetting-world to include the begotten, while the domain for the nonbegetting-world excludes any such thing.

Common to the variety of approaches for implementing this (for details and references see notes 6 and 14) is a function Q mapping a world w to a 'w-relative domain', a subset of the superdomain D ($Q(w) \subseteq D$). With such a Q, for a term τ and world w, let Q^{τ}_{w} be a set of assignments containing g just if $k^{\tau}_{g,w} \in Q(w)$—just if the Q-domain for w contains the g-relative designation of τ for world w: fixing a Q, such a g can be said 'w-designating for τ'. The following postclassical analysis of quantification then replaces ($\forall_{cl}$):

$$(\forall_{Q}) \qquad \vDash_{g,w} \forall \tau.\varphi \coloneqq \text{for every } g' \in Q^{\tau}_{w} \text{ such that } g' \sim^{\tau} g,\ \vDash_{g',w} \varphi$$

('Every τ is such that φ' is true in a world relative to an assignment just if, for every τ-variant of that assignment which, according to Q, maps τ to an entity which exists in that world, φ is true in that world relative to that τ-variant.)

Adjusting the classical quantificational fragment to include ($\forall_{Q}$) yields the intended logical revision.[14]

Concerning type (1), actual existence–possible nonexistence: we have $\vDash_{g,w} \forall x.\Box \exists y.x=y$ ('everything is necessarily identical to something') just if for every g', an x-variant of g which is w-designating for x, and every world w', there is some g'', a y-variant of g' which is w'-designating for y, such that the w'-designata of x and y, by g'', are identical. Suppose that Bill is in $Q(w)$ but not

in $Q(w')$: then, relative to w, Bill is a candidate rigid designatum of x by some g'; but relative to w', Bill is not a candidate designatum of y by any g''; so the sentence is not valid. But by *CBF*, the uncontroversially valid $\Box\forall x.\exists y.\, x = y$ ('necessarily, everything is identical to something') entails it; so *CBF* is invalidated. Concerning type (2), actual nonexistence–possible existence: let F be a predicate and S a subset of the domain such that, in any w, F is satisfied by an entity just if it is a member of S and a member of $Q(w)$. Then $\models_{g,w}\forall x.\Box Fx$ ('everything is such that it is necessarily F') just if for every x-variant g' of g which is w-designating for x, and every world w', the w'-designatum of x by g' satisfies F; with x rigid, and if $Q(w) \subseteq S$, this holds by the stipulation for F. But $\models_{g,w}\Box\forall x.\, Fx$ just if for every w' and every x-variant g' of g which is w'-designating for x, the w'-designatum of x by g' satisfies F: and with x rigid, whenever $Q(w') \not\subseteq S$, this fails to hold by the stipulation for F—so *BF* is invalidated.

8.4 Relativized metaphysical modality

As announced, the RMM strategy with each of the challenges is to reject the ILP: to distinguish possibility *in light of* which, φ, from possibility that φ. This section expands on this strategy. We begin with technicalities: by stating our proposed non-ILP adjustment to the classical truth-conditions, and explaining how despite advancing nonclassical truth-conditions, we nevertheless avoid the *logical revisionism* of the postclassical approaches. After this we turn to interpretive matters, and explain how we intend also to avoid *metaphysical revisionism*.

This combination of technicalities and interpretation, in the abstract, is what constitutes RMM: though we want to stress that the various solutions to the various puzzles are 'modular', with none requiring any other; more generally, we do not identify RMM with any of these proposals, but rather with their common technical move and its interpretation.

8.4.1 Technicalities: double-indexation

In a 'double-indexing' semantics, truth is relativized to a certain parameter twice over. RMM relativizes twice over to a *world* parameter, providing a definition of truth relative to (perhaps inter alia) a *pair* of possible worlds, notated $\models^{v}_{w}\varphi$: in intuitive terms, setting the value of v to a given world involves 'considering that world as actual'; when the value of w is set to a world distinct from the value of v, then from the point of view of that v-world, the w-world is 'considered as counterfactual'. This approach is apt to reject the ILP, at least in structural terms: roughly, possibility *that* φ goes with w-relativity; possibility *in light of which* φ, with v-relativity.

RMM truth-definitions for modality and quantification merely 'up-parametrize' the classical ($\Box_{cl}$) and ($\forall_{cl}$):

$$\left(\Box_2\right) \qquad \models^{v}_{w}\Box\varphi := \text{for every } w', \models^{v}_{w'}\varphi$$

('Necessarily, φ' is true relative to worlds v and w just if, for every world w', φ is true relative to v and w'—true, 'considering v as actual', just if for every w' 'considered as counterfactual', φ is true relative to v and w'.)

$$\left(\forall_2\right) \qquad \models^{v}_{w,g}\forall\tau.\varphi := \text{for every } g' \sim^{\tau} g,\ \models^{v}_{w,g'}\varphi$$

('Every τ is such that φ' is true relative to worlds v, considered as actual, and w, considered as counterfactual, and assignment g just if, for every τ-variant of g, φ is true relative to v, w, and that τ-variant.)

Possibility *that* is represented with w-relativity, possibility *in light of* with v-relativity. (A bit more precisely: relative to v and w, possibility *that blah* is the existence of some w' for which, relative to v and w', it is true that *blah*; while, relative to v and w, possibility that *blah in light of* a possibility is the existence of some v' (a possible world relative to v and w: namely, for some φ, $\not\models^{v}_{v'} \varphi$) such that for some w', relative to v' and w', *blah*—considering v' as actual and w' as counterfactual, *blah*—in evident conflict with the ILP.)

Postclassical apparatus is reassigned to a more fundamental role:

- For the propositional fragment (and the *laws* puzzle), the accessibility relation A from $(\Box_A)$ is repurposed to restrict the truth-definition: whether $\models^{v}_{w} \varphi$ is *undetermined* unless $w \in A(v)$.
- For the referential fragment (and *Chisholm's Paradox*), the $\boldsymbol{k}$ parameter is converted to an 'upper' parameter (namely, the truth-definition is for $\models^{v,\boldsymbol{k}}_{w} \varphi$ —for the significance of this, see note 18); while the counterpart function κ from $(\Box_\kappa)$ is reassigned to, in the evaluation of a term τ, mapping the individual concept (k_τ) it determines in a context, together with a context world v, to an individual concept, with the following effect on the truth-definition for elementary sentences:[15]

$$\left(\Gamma_{2,\kappa}\right) \qquad \models^{v,\boldsymbol{k}}_{w} \Gamma^{n}\tau^{1}\ldots\tau^{n} := \text{the sequence } \kappa\left(k^{\tau^{1}},v\right)_{w},\ldots,\kappa\left(k^{\tau^{n}},v\right)_{w} \text{ satisfies } \Gamma^{n} \text{ in } w$$

- For the quantificational fragment (and the *existence* and *nonexistence* puzzles), the domain function Q from $(\forall_Q)$ is repurposed to restrict the truth-definition: whether $\models^{v}_{w,g} \varphi$ is *undetermined* unless, for every τ, $g \in Q^{\tau}_{v}$

It should be evident that despite this use of the postclassical apparatus, the truth-definitions in themselves are inadequate to invalidate **S5**, *BF*, or *CBF*: avoiding logical revisionism.

8.4.2 Interpretation

It remains to assess whether RMM avoids *metaphysical* revisionism—an issue to be settled not with a piece of mathematics but with its interpretation. As previously indicated, the v and w indices go with possibility *in light of* and possibility *that*. To complete the interpretation, we discharge this jargon by sketching, first, a 'moderate naturalist' metaphysical account of which phenomena 'in reality' answer to possibility *in light of* and possibility *that*; and, second, a 'Context–Index' pragmatical explanation of why possibility *in light of* is not, eo ipso, possibility *that*.

For reasons of space, we restrict consideration to the laws puzzle involving type (i) undermining—a useful exemplar for both its technical simplicity and the particularly stark interpretive challenge it raises; the remaining puzzles have broadly analogous resolutions.[16]

8.4.2.1 Moderate naturalist metaphysics

Our intended picture is *moderate modal naturalism*: facts about natural law (and, in turn, about metaphysical possibility) have a hybrid explanation: some such facts are underdetermined either by any abstract metaphysical principles alone, or by any actual contingent 'categorical' facts alone, but instead become determined only through the concretization of certain true such principles in actual categorical facts.[17]

Using the analysis of the propositional fragment, the 'metaphysical principle' input is represented by the course of values (varying v) of the set $N(v) = \{w \mid vAw\}$ containing just those worlds accessible from v; the 'actual categorical fact' input is represented by a possible world @ (such that for any proposition p, p is true *simpliciter* just if p is true in @). It is possible *that* such-and-such just if for some $w' \in N(@)$, the set of worlds accessible from the actual world, such-and-such is true in w'. But it is possible, *in light of* a possible world v (considering v as actual), *that* such-and-such, just if for some (counterfactually-considered) $w'' \in N(v)$, the set of worlds accessible from v, such-and-such is true in w''.

Let us now revisit type (i) undermining with this RMM story: by *sensitivity*, the totality of actual categorical facts (represented with @) yield, via the abstract principles, actual laws ($N(@)$) permitting a nonactual totality of categorical facts—an *ok-world*, $v^\dagger$—($v^\dagger \in N(@)$) yielding, via the abstract principles, nonactual laws ($N(v^\dagger)$) permitting a totality of categorical facts—a *bad-world* $w^\dagger$ — ($w^\dagger \in N(v^\dagger)$) forbidden by the actual laws ($w^\dagger \notin N(@)$). By *necessitarianism*, the bad-world $w^\dagger$, which counts *in light of* actual categorical facts (@) and their principle-determined laws ($N(@)$) as an impossibility ($w^\dagger \notin N(@)$) sometimes counts as a possibility *in light of* a possibility (namely, *in light of* the ok-world $v^\dagger \in N(@)$; and *in light of* the ok-world $v^\dagger$, the bad-world $w^\dagger$ is indeed a possibility: $w^\dagger \in N(v^\dagger)$). The ILP would therefore require that the bad-world $w^\dagger$ is a possible possibility. But by the double-indexing semantics for modals ($\Box_2$), the possible possibilities are exactly the possibilities: so the bad-world $w^\dagger$ is a possible possibility only if $w^\dagger \in N(@)$—and it isn't, making for a counterexample to the ILP.

8.4.2.2 Context–Index pragmatics

Consider a (v-independent) sentence—the *bad-sentence*, $\varphi^\dagger$—false (for every v) in every $w' \in N(@)$ but true (for every v) in the bad-world $w^\dagger$: we claim that the bad-sentence $\varphi^\dagger$ is a *mere* possibility *in light of* a possibility. One might reasonably complain: why 'mere'? Why isn't there some reasonable sense in which the bad-world $\varphi^\dagger$ remains *possible*? (A very 'broad' sense it may well be—but *metaphysical* possibility, our target, is after all supposed to be the 'broadest' form of possibility.) But otherwise, how can we claim to preserve *necessitarianism*?

Our answer appeals to the *Context–Index* (CI) pragmatics (as codified in Lewis 1980a): this interpretation of double-indexation sees the truth-value of a natural language sentence as relativized to a possible world both *as index* and *as context*. The two central ideas are these: the truth-value of a sentence φ as asserted in a possible world w^* is that of φ relative to w^* both *as index* and *as context*; by the rules of language, index-relativity is amenable to control by operators such as modals, while context-relativity is protected against such control. So in the course of compositional determination of the truth-value of an assertion in w^*, worlds other than w^* may be involved *as index*, but only w^* is ever involved *as context*.

We intend w-relativity to be *indexical*, v-relativity to be *contextual*.[18] Accordingly,[19] for a sentence φ, any assertion of φ in the actual world @ is true just if $\models^{@}_{@} \varphi$ (just if true relative to the actual world both as context and as index); and for any operator O which could be used in natural language, whether $\models^{@}_{@} O\varphi$ (whether an actual-world assertion of $O\varphi$ is true) is determined by the course of values, varying w, of whether $\models^{@}_{w} \varphi$ (of whether φ is true relative to the actual world *as context*, but relative to an arbitrary world *as index*). Presumably a natural language operator never counts as a *possibility modal* if sometimes $\models^{v}_{w} O\varphi$ (if for some context-world v and index-world w, $O\varphi$ is true relative to v and w) when for no w', $\models^{v}_{w'} \varphi$ (when there is no index-world w' with φ true relative to the old v as context and the new w' as index); so, because it is determined whether $\models^{@}_{w} \varphi$ (whether, with an actual-world context and specified-world index, φ has a truth-value) only if $w \in N(@)$—only if w is an actual-world nomological

possibility—and for all v and all $w \in N(@)$, $\not\models^{v}_{w} \varphi^{\dagger}$ (the bad-sentence is false whatever the context-world, if the index-world is nomologically accessible from the actual world), then O cannot count as a possibility modal if $\models^{@}_{@} O\varphi^{\dagger}$ (if $O\varphi^{\dagger}$ is truly asserted in the actual world). Finally, we insist on the unintelligibility of any sense in which φ is 'possible' unless natural language contains or is extensible with some possibility modal operator O for which $O\varphi$ is truly asserted: so, as desired, the fact that $\models^{v^{\dagger}}_{w^{\dagger}} \varphi^{\dagger}$ —the fact that, relative to the ok-world as context and the bad-world as index, the bad-sentence is true—yields by itself no intelligible sense in which $\varphi^{\dagger}$ is possible.

On the CI pragmatics, our actual uses of sentential operators cannot give us access to nonactual values of v: we are stuck in our actual world context, from which sentential operators are powerless to dislodge us. Yet we can access nonactual values of v: just not through language alone. In 'serious' assertion there is no hope: the v-parameter is inexorably actualized by our actual world context. Fortunately, our assertions are not always serious: sometimes, we introduce a supposition, and for a while perform assertions against that background; if the supposition is that the ok-world $v^{\dagger}$ *is actual*, then (while the supposition remains in force) what matters to us in asserting φ is not the *real* truth-value $\models^{@}_{@} \varphi$ but the truth-value *within the supposition* $\models^{v^{\dagger}}_{v^{\dagger}} \varphi$. To introduce such a supposition is to commence 'considering the ok-world $v^{\dagger}$ as actual' (Jackson 1994). While the supposition is in force, we are right to assert $\Diamond\varphi^{\dagger}$ ('possibly, *the-bad-sentence*'); once the supposition is canceled, we are right to assert what is true, namely $\neg\Diamond\varphi^{\dagger}$ ('it is impossible that *the-bad-sentence*'). The assertability of $\Diamond\varphi^{\dagger}$ within the supposition does not give any sense in which $\varphi^{\dagger}$ is possible, or possibly possible: the supposition is *false*; what we assert within it is *mere pretense*: correct assertion of possibility within a mere pretense is not eo ipso any kind of genuine possibility, and eo ipso nothing more than pretend possibility. (No more so, anyway, than pretending that this easy chair is a bison makes it a genuine bison.)

8.5 Concluding historical speculation

Double indexation is an old and venerable tool, invented in 1968 (Kamp 1968), published first in 1970 (Lewis 1970a) and then by its inventor in 1971 (Kamp 1971), and put to a wide array of high-profile philosophical uses since then (Kaplan 1977; Stalnaker 1978; Chalmers 2010). Why then do the literatures on the various metaphysical puzzles overlook our RMM solution? Several distinguishable barriers may be collectively relevant (with barriers to initial uptake persisting as barriers through the general path-dependence of scholarly traditions).[20]

First, the postclassical tools are all older than double indexation: accessibility was invented in the mid-to-late 1950s (see Copeland 2002: compare Meredith and Prior 1956, Kripke 1959); counterpart theory, first announced in a letter to Føllesdal from 6 March 1966,[21] was published soon after (Lewis 1968); domain relativity was invented in the mid or late 1950s (Kripke 1963).

Second, the literatures are themselves old (with their order of genesis reversing the complexity of the classical analyses they challenge): the *ontology* puzzle arises together with its postclassical treatment (compare Kripke 1963 on Sherlock Holmes); the *essence* puzzle arises in the late 1960s (Chisholm 1967; independent rediscovery, Lewis 1968, 28, together with its postclassical treatment); and the *laws* puzzle, discovered only somewhat recently (Carroll 1994, sect. 3.1) and overlooked until quite recently (Fine 2005, 243–8), was perhaps implicit in Lewis's early antinecessitarian development of best-systems theory (which was, moreover, in service of a *single-indexing* analysis of the counterfactual conditional: Lewis 1973).

Third, a leading figure in the literatures, David Lewis, was by his own acknowledgement (Lewis 1980a, 42) 'compartmentalized' in his theorizing in early days, in some moods using

double-indexing (Lewis 1975), in others overlooking it (Lewis 1970b); Saul Kripke, casting a long shadow on the literatures, would not to our knowledge explicitly discuss double-indexing.

Fourth, double-indexing is not yet the diagonal account of consequence as developed in Kaplan 1977: as observed briefly in note 17, the deep story makes critical appeal to this; but the metatheory of diagonal consequence poses inherent challenges, with attendant delays in its uptake in metaphysical application.

Fifth, the CI pragmatics developed gradually over the late 1970s (compare Kaplan 1977; Stalnaker 1978), and only reached its final form in Lewis's 'Index, context, and content' (Lewis 1980a)—that last, moreover, was not collected in Lewis's *Philosophical Papers* and long escaped wider attention (and is, moreover, an odd, protean paper)—and in addition, was applied most prominently to explaining apriori knowledge (Kaplan and Stalnaker, among others), an issue remote from the literatures on the metaphysical puzzles. Moreover, the tradition in analytic philosophy is to do linguistic analysis at the level of the semantics of the sentence: still poorly integrated are such acts and attitudes as *supposition* and *assertion*, to which RMM makes crucial appeal.

Sixth, grasping the commonalities among the various puzzles requires understanding counterpart theory *as a semantics*: Lewis never presented it as such, and in consequence such a treatment was entirely lacking until Fara's (2008, 2012) work of the last decades; this in particular obscured its logical significance, with prominent discussion of CP in the 1980s failing to see the connection to **S5** revision (and, anecdotally, the connection remains poorly recognized even today).

Notes

1 The issues here are wide-ranging in subject matter and straddle subdisciplinary borders, and they are also intricate technically and challenging conceptually. To give more than a sketch, a great deal should be said. Because of space, our sketch here ignores much, and often without comment: hopefully enough is sketched for novices to get the gist and for experts to fill in what we leave out.

2 But see note 17.

3 For example: care with parentheses; the free swapping of 'official' and 'unofficial' object-language expressions; the distinction between constant and variable terms; the valuation-relativity of base-expression truth and satisfaction; the distinctions among frames, structures, and models; the distinction between syntactic and semantic consequence, with familiar metalogical correspondence results presupposed.

4 Namely, 'normality'—as axioms, all propositional tautologies, and the principle *K*: $\Box(\varphi \supset \psi) \supset (\Box\varphi \supset \Box\psi)$('necessities don't entail contingencies'); as rules, *modus ponens* and *necessitation* (the latter requiring inclusion of $\Box\varphi$ for every included φ).

5 One principle that is *not* validated is *NI* (the 'necessity of identity'; compare Kripke 1980, 3–5): $\tau = \tau' \supset \Box(\tau = \tau')$—perhaps *the professor is the janitor* is an intuitive counterexample. Still, *NI* is valid for a subfragment containing only rigidly designating terms.

6 Various options from Garson's (1984, 250) systematization of quantified modal semantic theories are available: fixed-domain approaches are systems **Q1** (Kripke 1963), in a rigid subfragment with constant terms; and **QC** (Garson 1984, 265–6), with nonrigid variables. As Garson observes, quantification over individual concepts is equivalent in strength to second-order quantification, so that **QC** lacks any complete axiomatization; and moreover, **QC** problematically validates 'something is necessarily the author of counterpart theory'.

7 Moreover, in combination with the Context–Index pragmatics treatment of speech act content (Lewis 1980a, 37–8), each postclassical analysis faces what we call the 'Generalized Humphrey Problem' (locus classicus: Kripke 1980, 45 n. 13) against counterpart theory; but also, on the most charitable readings, Lewis's 'by what right' objection (Lewis 1986b, 246) against accessibility semantics and Williamson's accusation (Williamson 1998, 263) that relative domains theory is 'philosophically unsatisfying'). Very roughly, the problem is that the postclassical analyses 'change the subject', with the truth-condition for $\Box\varphi$ sometimes diverging from the condition for the content of φ to be necessary. Space prohibits further treatment.

8 For best-systems undermining, see Lewis 1986b, 20; for Aristotelian undermining, see Fine 2005, 244–5, following Carroll 1994, sect. 3.1. The terminology is from Lewis 1986a, xv, discussing Lewis 1980b—compare Fine 2005, 246 n. 16.

9 We do not know of an author advocating this approach to the laws puzzle; more generally, the literature on this puzzle is poorly developed.

10 A possible point of controversy: many advance a weaker condition, that necessarily, *if a exists*, then *a* is *F*. This pertains to the further issue of the interaction of modality and quantification, so we set it to the side.

11 The vagueness of 'small/big enough' is not relevant.

12 Forbes's approach differs importantly from Lewis's in offering a more complex, contextually ambiguous translation scheme, potentially with logically non-revisionary semantical results comparable to those available through our proposal ($\Gamma_{2,\kappa}$): space prohibits adequate consideration of Forbes's approach. Thanks to Forbes for discussion.

13 Stereotypically—and, indeed, per both leading partisans to the debate (Lewis 1968, 28 and Kripke 1980, 45 n. 13)—counterpart theory is a metaphysical doctrine that nothing exists in more than one world. But any reasonably permissive metaphysics of 'derivative' objects can insist only that 'fundamental' objects be thus 'worldbound': allowing even for rigid designation and *NI* (see note 5) with this metaphysics; through its use of individual concepts, the classical, **S5**-validating semantics is compatible with the worldbound metaphysics: so the latter cannot by itself help with the essence puzzle; and the schematic way with the essence puzzle in the main text is compatible with rigid designation and transworld individuals.

14 Further issues remain to be settled for a determinate position in the Garson taxonomy, as discussed in note 6. Of just the approaches avoiding free logic and truth-value gaps: with all terms rigid and without constant terms, the result is system **QK** (Kripke 1963); with nonrigid variables (and otherwise making maximally straightforward choices), **Q2** (Thomason 1969).

15 The contextual occurrence of $\boldsymbol{k}$ is preserved for two reasons: first, pedagogically, to highlight the contrast with the counterpart-theoretic formulation ($\Box_\kappa$), in which the parameter occurs indexically, and thereby exemplify the general strategy here of moving indexical parameters to contextual parameters; and second, doctrinally, to permit the potential for primitive metaphysical context-sensitivity in term-designation, broadly along the lines considered in Lewis 1971. Thanks to Cian Dorr for encouraging greater explicitness in stating and elucidating the view here.

16 For *laws*, see Murray and Wilson 2012, sect. 3 and Murray 2017, ch. 3 (though the former deals with a somewhat different challenge, apparently requiring partitioning modal space by laws).

For *CP*, see Murray and Wilson 2012, sect. 2 and Murray 2017, ch. 4 (though where the former invokes ($\Box_2$) and the restriction of truth-conditions as part of a more broadly 'accessibilist' outlook, the latter invokes ($\Gamma_{2,\kappa}$) on behalf of a more broadly 'counterpart theoretic' approach. Our 'official' stance is that the RMM framework makes room for both options, remaining neutral between them and permitting the choice to be settled on the merits).

For *existence* and *nonexistence*, see Murray 2017, ch. 5.

17 The full story involves 'diagonal' logical consequence. The hybrid explanation maintains that, where φ is a consistent maximally specific statement of categorical facts, there is some sentence λ^φ giving a full statement of laws, such that $\vdash \varphi \equiv \lambda^\varphi$—namely, $\models^\nu_\nu \varphi \equiv \lambda^\varphi$ for every ν. Diagonalization is needed because while φ is contingent ($\nvdash \varphi \supset \Box\varphi$), λ^φ is noncontingent ($\vdash \lambda^\varphi \supset \Box\lambda^\varphi$)—a substitution-failure between logical equivalents, both structurally analogous to the familiar equivalence between the contingent φ and the noncontingent *actually*, φ, and handled with the same logical apparatus.

18 More generally, 'upper' parameters are contextual, 'lower' parameters indexical.

19 To save space, two gross simplifications: technically, we neglect relativity of truth to non-world parameters; conceptually, we treat the propositional modal fragment with our double-indexing semantics as representative of all natural language modal discourse, sidestepping many important complications.

20 To save space, our focus here is on tractable and therefore narrowly 'technical' barriers; as for candidate barriers arising from a putative philosophical consensus—'scientism'; a striving for 'objectivity'; a 'universalist' conception of logic; Kantian hangovers from the Vienna Circle; metaphysical antinaturalism; discomfort with 'perspective'; an association between possibility and meaning—whether genuine or not, the matter is too big to settle here.

21 http://www.projects.socialsciences.manchester.ac.uk/lewis/wp-content/uploads/2018/05/Lewis-David-to-Follesdal-Dagfinn-06.03.1966.pdf

References

Barcan, Ruth C. 1946. "A Functional Calculus of First Order Based on Strict Implication." *The Journal of Symbolic Logic* 11: 1–16.

Burgess, John P. 2009. *Philosophical Logic*. Princeton: Princeton University Press.

Carnap, Rudolf. 1947. *Meaning and Necessity*. 2nd ed. Chicago: University of Chicago Press.

Carroll, John W. 1994. *Laws of Nature*. Cambridge: Cambridge University Press.

Chalmers, David J. 2005. "The Foundations of Two-Dimensional Semantics." In *Two-Dimensional Semantics: Foundations and Applications*, edited by Manuel García-Carpintero and Josep Macià. Oxford: Oxford University Press.

———. 2010. "The Two-Dimensional Argument against Materialism." In *The Character of Consciousness*, edited by David Chalmers. Oxford: Oxford University Press.

Chandler, Hugh. 1976. "Plantinga and the Contingently Possible." *Analysis* 36: 106–9.

Chisholm, Roderick M. 1967. "Identity through Possible Worlds: Some Questions." *Noûs* 1: 1–8.

Copeland, B. Jack. 2002. "The Genesis of Possible Worlds Semantics." *Journal of Philosophical Logic* 31: 99–137.

Fara, Delia Graff. 2008. "Relative-Sameness Counterpart Theory." *Review of Symbolic Logic* 1: 167–89.

———. 2012. "Possibility Relative to a Sortal." In *Oxford Studies in Metaphysics*, edited by Karen Bennett and Dean W. Zimmerman, Vol. 7. Oxford: Oxford University Press.

Fine, Kit. 1994. "Essence and Modality." *Philosophical Perspectives* 8: 1–16.

———. 2005. "Varieties of Necessity." In *Modality and Tense*. Oxford: Oxford University Press.

Forbes, Graeme. 1984. "Two Solutions to Chisholm's Paradox." *Philosophical Studies* 46: 171–87.

Garson, James W. 1984. "Quantification in Modal Logic." In *Handbook of Philosophical Logic*, edited by Dov Gabbay and Franz Guenthner, Vol. II. Dordrecht: Kluwer.

Jackson, Frank. 1994. "Armchair Metaphysics." In *Philosophy in Mind*, edited by John O'Leary-Hawthorne and Michaelis Michael. Dordrecht: Kluwer.

Kamp, Hans. 1968. "Tense Logic and the Theory of Linear Order." PhD thesis, UCLA.

———. 1971. "Formal Properties of 'Now'." *Theoria* 37: 227–74.

Kaplan, David. 1977. "Demonstratives." In *Themes from Kaplan*, edited by Joseph Almog, John Perry, and Howard Wettstein. Oxford: Oxford University Press.

Kripke, Saul Aron. 1959. "A Completeness Theorem in Modal Logic." *Journal of Symbolic Logic* 24: 1–14.

———. 1963. "Semantical Considerations on Modal Logic." *Acta Philosophica Fennica* 16: 83–94.

———. 1980. *Naming and Necessity*. Cambridge, MA: Harvard University Press.

Lewis, David. 1968. "Counterpart Theory and Quantified Modal Logic." *Journal of Philosophy* 65: 113–26.

———. 1970a. "Anselm and Actuality." *Noûs* 4: 175–88.

———. 1970b. "General Semantics." *Synthese* 22: 18–67.

———. 1971. "Counterparts of Persons and Their Bodies." *Journal of Philosophy* 68: 203–11.

———. 1973. *Counterfactuals*. Cambridge, MA: Harvard University Press.

———. 1975. "Languages and Language." In *Minnesota Studies in the Philosophy of Science*, edited by Keith Gunderson. Vol. VII. Minneapolis: University of Minnesota Press.

———. 1980a. "Index, Context, and Content." In *Philosophy and Grammar*, edited by Stig Kanger and Sven Öhman. Dordrecht: Reidel.

———. 1980b. "Veridical Hallucination and Prosthetic Vision." *Australasian Journal of Philosophy* 58: 239–49.

———. 1986a. "*Introduction*." In Lewis, p. 2.

———. 1986b. *On the Plurality of Worlds*. London: Blackwell.

Linsky, Bernard, and Edward N. Zalta. 1994. "In Defense of the Simplest Quantified Modal Logic." In *Logic and Language*, edited by James Tomberlin, 8:431–58. Philosophical Perspectives. Atascadero: Ridgeview.

Linsky, Bernard. 1996. "In Defense of the Contingently Nonconcrete." *Philosophical Studies* 84: 283–94.

Loewer, Barry. 1996. "Humean Supervenience." *Philosophical Topics* 24: 101–27.

———. 2012. "Two Accounts of Laws and Time." *Philosophical Studies* 160: 115–37.

Meredith, C. A., and Arthur Prior. 1956. "Interpretations of Different Modal Logics in the 'Property Calculus'." In *Logic and Reality: Essays on the Legacy of Arthur Prior*, edited by Jack Copeland. Oxford: Clarendon Press, 1996.

Murray, Adam Russell. 2017. "Perspectives on Modal Metaphysics." PhD thesis, University of Toronto.

Murray, Adam Russell, and Jessica M. Wilson. 2012. "Relativized Metaphysical Modality." In *Oxford Studies in Metaphysics*, edited by Karen Bennett and Dean W. Zimmerman, 7:189–226.

Salmon, Nathan U. 1981. *Reference and Essence*. Princeton: Princeton University Press.
Shoemaker, Sydney. 1980. "Causality and Properties." In *Time and Cause*, edited by Peter van Inwagen. Dordrecht: D. Reidel.
Stalnaker, Robert C. 1978. "Assertion." *Syntax and Semantics* 9: 315–32.
Thomason, Richmond. 1969. "Modal Logic and Metaphysics." In *The Logical Way of Doing Things*, edited by Karel Lambert. New Haven: Yale University Press.
Williamson, Timothy. 1990. *Identity and Discrimination*. Oxford: Blackwell.
———. 1998. "Bare Possibilia." *Erkenntnis* 48: 257–73.
———. 2013. *Modal Logic as Metaphysics*. Oxford: Oxford University Press.
———. 2016. "Modal Science." *Canadian Journal of Philosophy* 46: 453–92.

Chapter 9

ONTOLOGICAL DEPENDENCE, GROUNDING AND MODALITY

Fabrice Correia

Ontological dependence and grounding are two important items in the metaphysician's toolbox: both notions can be used to formulate important philosophical claims and to define other notions that play a central role in philosophical theorising. Philosophical inquiry *about* ontological dependence and (especially) grounding has been very lively over the past few years, making it difficult to write a short review article on any of them, let alone a short review article on both. I try to reach a good compromise between a discussion of each notion taken separately and a discussion of how they relate to one another. I begin by introducing the notions and discussing a number of their connections with modality (Sections 9.1 and 9.2), starting with grounding for systematic reasons (some important concepts of ontological dependence are defined in terms of grounding). I then further the discussion of how the notions are connected to each other, by arguing against the view that (partial) grounding is equivalent to (the converse of) ontological dependence between facts (Section 9.3). Finally, I discuss their respective roles in the theory of fundamentality (Section 9.4).[1]

9.1 Grounding and modality

9.1.1 Grounding

Those who are not sceptics about grounding disagree about a number of core issues concerning the notion.[2] Disagreement already starts on issues pertaining to logical form (see Correia 2010, §1; Correia & Schnieder 2012, §3.1). Grounding claims have been assumed to be expressible by means of statement of the following types:

(1) The fact that p grounds the fact that q
(2) Its being the case that p makes it the case that q
(3) q in virtue of the fact that p
(4) The fact that q is explained by the fact that p
(5) q because p

On the face of it, (1) and (4) are relational predications while the other items are not: in both (2) and (5) two sentential expressions flank a sentential operator, and in (3) a sentential expression and a nominal expression flank what is sometimes called a "prenective". Which of these logical

forms should we privilege in order to regiment grounding statements? The question is of central importance, since arguably the operator regimentation does not require an ontology of facts, while the other regimentations do. The operator regimentation is favoured e.g. in Correia (2010) and Fine (2012), the predicate regimentation in Rosen (2010) and Audi (2012),[3] and the prenective regimentation in Rodríguez-Pereyra (2015).[4]

Another issue pertaining to logical form regards the "arity" of grounding. It is widely accepted that grounding can be many-to-one, i.e. that several facts may jointly, but not individually, ground a given fact (the view can of course also be formulated both in the prenective language and in the operator language). Thus, for instance, many would hold that the fact that 2 = 1 + 1 and the fact that Socrates is a philosopher jointly ground the fact that 2 = 1 + 1 and Socrates is a philosopher, and deny that any of the former facts grounds the latter fact. Orthodoxy has it that cases of grounding are either one-to-one or many-to-one. A minority view holds that grounding can also be many-to-many (see Dasgupta 2014 and Litland 2016).

For reasons of space, I will leave aside this unorthodox view, and I will largely assume a predicate regimentation of grounding, leaving completely aside the prenective regimentation, and *almost* completely aside the operator regimentation.

A number of important distinctions pertaining to grounding are discussed in the literature. Let me here mention three of them. (Here and below, I shall use '□' for 'It is metaphysically necessary that', '*f*', '*g*' and '*h*' as variables for facts, and, following Fine 2012, '*f*, *g*, . . . < *h*' for '*f*, *g*, . . . ground *h*'. Throughout the chapter, each general principle with free variables should be understood as being universally closed.)

Factive vs non-factive. To say that grounding is factive is to say that a fact cannot be grounded in other facts unless it and its grounds obtain, i.e.

□(if *f*, *g*, . . . < *h*, then *f*, *g*, . . . and *h* all obtain) Factivity

The majority view seems to be that grounding is factive, but some authors hold that there is room for a non-factive notion, in terms of which a factive notion can be defined in the obvious way (see Fine 2012, pp. 48–50; Correia & Skiles 2019, pp. 655–656).

Factual vs generic. It is plausible to hold that there is a ground-theoretic connection between being white and being coloured: something's being white grounds its being coloured. There is an obvious way of formulating this connection given the linguistic resources introduced earlier: necessarily, if something is white, then the fact that it is white grounds the fact that it is coloured. Yet intuitively there is a more direct connection, which precedes, as it were, the grounding connections between whiteness facts and colouredness facts across possible worlds. This more direct connection is *generic grounding*. Factual grounding is by far the most familiar notion. Generic grounding has only very recently been recognised as a notion in its own right (Fine 2015; Correia & Skiles 2019, pp. 660–661).[5]

Worldly vs representational. It is common to distinguish between facts as obtaining states of affairs (worldly items) and facts as true propositions (representations). What kind of facts does grounding relate? Some authors (see e.g. Audi 2012) favour the first option, some go with the second option (see e.g. Rosen 2010). This distinction between two ways of conceiving the relata of grounding is closely related to a deeper distinction between *worldly* and *representational* conceptions of grounding. Say that two facts *f* and *g* are *ground-theoretically equivalent* iff they are indistinguishable from a ground-theoretic perspective, i.e. iff (i) whatever grounds *f* grounds *g*, and vice versa, and (ii) given any fact *h* and (possibly empty) plurality of facts Δ, *h* is grounded in *f*, Δ iff *h* is grounded in *g*, Δ. Grounding is worldly just in case whenever two

sentences *S* and *T* describe the world as being the same way and pick out facts that are relata of grounding, these facts are ground-theoretically equivalent, and grounding is representational just in case it is not worldly. It is natural to think that the view that grounding is worldly should demand that grounding relate entities like states of affairs, and that the view that grounding is representational should demand that grounding relate entities like propositions. Yet this can be disputed (see Correia 2020). Be that as it may, the difference between the two conceptions of grounding makes for a big difference on the acceptability of grounding statements. Consider the sentences 'Fred is a bachelor' and 'Fred is an unmarried male'. They arguably describe the world as being the same way. If grounding is worldly, then the fact *f* picked out by the first sentence and the fact *g* picked out by the second sentence should be ground-theoretically equivalent. But then, on the assumption that *f* does not ground itself, and hence a fortiori on the assumption that grounding is irreflexive,[6] *f* cannot ground *g*. By contrast, assuming that grounding is representational allows one to hold that *f* does ground *g*—as Rosen (2010) indeed does.[7]

The worldly vs representational distinction and the factual vs generic distinction are orthogonal. The factive vs non-factive distinction makes sense for factual grounding, but not for generic grounding. For reasons of space, I will completely leave aside generic grounding.

Grounding can be used to formulate general claims of central philosophical interest. Here are some stock examples:

Mental facts (e.g. the fact that John is experiencing pain in his left hand)		Neurophysiological facts (the fact that John's brain is in state S)
Dispositional facts (e.g. the fact that the glass is fragile)		Categorical facts (the fact that the glass has molecular structure S)
Modal facts (e.g. the fact that Socrates is necessarily human)		Essentialist facts (the fact that Socrates is essentially human)
Alethic facts (e.g. the fact that 'Snow is white' is true)		Non-alethic facts (the fact that snow is white)
Determinable facts (e.g. the fact that the ball is coloured)	are grounded in	Determinate facts (the fact that the ball is red)
Existential facts about sets (e.g. the fact that {1,2} exists)		Existential facts about their members (the fact that both 1 and 2 exist)
Existential facts about wholes (e.g. the fact that the sandwich exists)		Facts about their parts (the fact that the ham and the bread are suitably related)
Existential facts about universals (e.g. the fact that wisdom exists)		Facts about their exemplifiers (the fact that Socrates is wise)

Grounding can also be used to characterise important notions, e.g. truth-making (Correia 2005, Schnieder 2006a, Cameron 2018), intrinsicality (Witmer, Butchard & Trogdon 2005, Rosen 2010), real definition (Fine 2015, Rosen 2015, Correia 2017), existential dependence (see Section 9.2), and being fundamental (see Section 9.4).

Can grounding itself be characterised in other terms? Most people consider it as a primitive notion, while few (Correia 2013, Correia & Skiles 2019) propose to analyse it—in terms of essence in the first case, in terms of a notion of "generalised identity" in the second case.

9.1.2 Connections with modality

The notion of grounding of interest to us here is sometimes labelled 'metaphysical', in order to distinguish it from other notions of grounding, e.g. causal, normative and logical grounding (see Correia 2005, pp. 55–56; Fine 2012, pp. 37–40). One way of singling out metaphysical grounding invokes modality. According to orthodoxy, metaphysical grounding induces metaphysically necessary connections, as stated in the following principle:

If $f, g, \ldots < h$, then $\Box$(h obtains if $f, g, \ldots$ all obtain) — Necessitarianism

Presumably, neither causal grounding nor normative grounding induces such connections. This is not to say that they do not induce any necessary connections at all. On a plausible view, each notion of grounding induces necessary connections of a distinctive type. Thus, for instance, causal grounding induces causally (or naturally) necessary connections, normative grounding normatively necessary connections, and logical grounding logically necessary connections (see Fine 2012, p. 38). On that account, it is plausible that given any pair of distinct notions of grounding, Φ-grounding and Ψ-grounding, either Φ-grounding will fail to induce Ψ-ly necessary connections, or Ψ-grounding will fail to induce Φ-ly necessary connections, or both. If this is correct, then we have a way of distinguishing metaphysical grounding from *all* the other notions of grounding.

Two modal principles closely related to Necessitarianism also belong to orthodoxy:

If $f, g, \ldots < h$, then $\Box$($f, g, \ldots < h$ if $f, g, \ldots$ and h all obtain) — Internality
If $f, g, \ldots < h$, then $\Box$($f, g, \ldots < h$ if $f, g, \ldots$ all obtain) — Rigidity

As Stephan Leuenberger (2014, p. 155) emphasises, Necessitarianism and Internality are logically independent. Rigidity entails Internality, and given Factivity, it entails Necessitarianism. Both Necessitarianism and Internality have been directly attacked (see in particular Skiles 2015; Leuenberger 2014; Litland 2015).

Here is a simple case put forward by Alex Skiles (2015) against Necessitarianism. What grounds the fact f that all swans in Switzerland are white? Skiles holds that the following plurality Δ of facts, where $a, b, \ldots$, are all the swans in Switzerland, does the job: the fact that a is white, the fact that b is white, . . . If he is right, then we do have a counterexample to Necessitarianism, since the facts in Δ could have obtained in the presence of a black swan in Switzerland. The natural reply is that the purported grounds constitute a *mere partial* ground for f, and that f is fully grounded in the facts in Δ *together with* the "totality fact" that a, b, . . ., are all the swans in Switzerland. The reply is indeed effective, since Necessitarianism does concern full grounding, and the newly invoked plurality of facts does necessitate f. Skiles anticipates the reply but proposes some objections to this view, as well as objections to other candidate replies. The debate on this particular kind of counterexample, as well as more generally on the truth of Necessitarianism and Internality, is currently ongoing.

The converse of Necessitarianism fails for many reasons.[8] It already fails because necessitation is, while grounding is not, reflexive. It also fails because of the following types of counterexamples:

Complete irrelevance: The fact that I had breakfast this morning does not ground the fact that $2 + 2 = 4$

Partial irrelevance: The fact that (this is red and John is sad) does not ground the fact that this is coloured

Wrong ordering: The fact that {Plato} exists does not ground the fact that Plato exists

The cases of complete irrelevance suggest that the converse of Necessitarianism could perhaps be rescued by replacing strict implication with some standard notion of "relevant" entailment with modal force.[9] Yet such a modification would not help to get rid of the other problems. A more promising suggestion is to invoke instead a notion of "exact" entailment, which can be characterised in Fine's truthmaker semantical framework as follows (see Fine 2017a on the general framework and on exact entailment, which is there simply dubbed 'entailment'):

'$p \Rightarrow q$' is true iff every state that exactly verifies p also exactly verifies q

This suggestion is indeed in line with a characterisation of (a worldly notion of) grounding proposed by Fine himself. Let us define *weak grounding* thus:

$f, g, \ldots$ weakly ground h iff$_{df}$ (f obtains & g obtains & $\ldots$) $\Rightarrow$ h obtains.

On Fine's account,[10] strict grounding can be characterised in terms of weak grounding as follows (see Fine 2012, pp. 71–74; 2017b, pp. 685–687, 700–701):

$f, g, \ldots < h$ iff (i) $f, g, \ldots$ weakly ground h, and (ii) h, alone or together with some other facts, does not weakly ground any of $f, g, \ldots$

Another interesting issue connecting grounding and modality is whether the modal status of a fact dictates the modal status of its grounds. Most examples of grounding ties that are commonly given are, on the face of it, examples where either a *contingent* fact is grounded in facts whose conjoined obtaining is itself *contingent*, or a *necessary* fact is grounded in facts whose conjoined obtaining is itself *necessary*. The specific questions I have in mind here are the following:

Can there be cases where both $f, g, \ldots < h$ and h contingently obtains, while it is necessary that $f, g, \ldots$ all obtain?
Can there be cases where both $f, g, \ldots < h$ and h necessarily obtains, while it is contingent that $f, g, \ldots$ all obtain?

According to grounding orthodoxy, disjunctions are grounded in their true disjuncts (see e.g. Fine 2012, p. 58). If this is correct, then the answer to question #2 is positive: the fact that 7 is my favourite prime number (a contingent fact) grounds the fact that 7 is or is not my favourite prime number (a necessary fact). If Necessitarianism holds, the answer to question #1 must be negative. But if the principle fails, then there may be cases of the relevant sort. Assume that in actuality, there are simple (i.e. partless) substances, and that each of them is necessarily material and hence non-mental. Assume in addition that there are possible worlds in which some simple substances are mental. Then the fact f that all the simple substances are non-mental is contingent. But arguably, if Skiles is right about the Swiss swans example described earlier, then f is grounded in the fact that a is non-mental, the fact that b is non-mental, ..., where $a, b, \ldots$, are all the simple substances that there actually are, and these facts by hypothesis all necessarily obtain.

Question #2 is closely related to Simon Blackburn's (1987) dilemma about the source of necessity. Here is a reconstruction of the argument formulated in terms of grounding. The argument purports to show that the following general claim is false:

For any p such that $\Box p$, the fact that $\Box p$ is grounded — Target

It goes as follows:

Either (a) for every p such that $\Box p$ and for all facts $f, g, \ldots$ such that $f, g, \ldots <$ the fact that $\Box p$, it is necessary that $f, g, \ldots$ all obtain, or (b) not-(a).

(b) is false. For assume that $\Box p$. Then by modal axiom (4), the fact that $\Box p$ necessarily obtains. But since question #2 must receive a negative answer, it follows that the fact that $\Box p$ cannot be grounded in facts whose conjoined obtaining is not necessary.

(a) is also false if Target is true. For assume for reductio both (a) and Target. Take any particular p_0 such that $\Box p_0$. Then we have an infinite series of pluralities of facts $\Delta_1, \Delta_2, \Delta_3, \ldots$ which enter into the following infinite regress:

> The fact that $\Box p_0$ is grounded in Δ_1
>
> $\Box(\Delta_1$ all obtain)
>
> The fact that $\Box(\Delta_1$ all obtain) is grounded in Δ_2
>
> $\Box(\Delta_2$ all obtain)
>
> The fact that $\Box(\Delta_2$ all obtain) is grounded in Δ_3
>
> ...
>
> Modality appears at each stage, which makes the regress vicious.
> Therefore, Target is false.

As we saw, grounding orthodoxy gives a positive answer to question #2. Doing so is one way of blocking the argument for the falsity of (b), and hence the whole argument. For more on how to escape the dilemma, see e.g. Hale (2010) and Cameron (2010).

9.2 Ontological dependence and modality

9.2.1 Ontological dependence

Ontological dependence comes in many different varieties.[11] An ontologically dependent entity is an entity whose existence or identity (or essence; I do not make a distinction here) demands that something be the case. Each of the following open sentences expresses a relational notion of ontological dependence:

(a) x cannot exist without γ
(b) x can, by its very nature, not exist without γ
(c) For some R, it is part of the nature of x that x is R-related to γ

Using '$\Box_x$' for 'It is part of the nature of x that', (a), (b) and (c) are naturally regimented by (RD1), (RD2) and (RD3), respectively:

(RD1) $\Box(x \text{ exists} \supset \gamma \text{ exists})$
(RD2) $\Box_x(x \text{ exists} \supset \gamma \text{ exists})$
(RD3) $\exists R \Box_x R x \gamma$

On the standard modal account of essence, according to which it is part of the nature of a thing that p iff it is impossible that the thing exists and $\neg p$, (RD1) and (RD2) are necessarily equivalent. Yet as Fine (1994) has convincingly argued, the standard account of essence is flawed, and some of his arguments can be used to show that (RD1) does not entail (RD2). Thus, for instance,

Socrates arguably depends on {Socrates} in the sense of (RD1), but not in the sense of (RD2) (see Fine 1995, pp. 271, 273–274).

Fine's objections are directed only against the right-to-left direction of the biconditional supported by the standard account of essence. Foes of the standard account usually have nothing against the other direction of the biconditional. Let us take it for granted. Then (RD2) entails (RD1). (RD3) can be derived from (RD2) by second-order existential generalisation. It can be argued that (RD3) does not in turn entail (RD2): the property of being identical to Socrates depends on Socrates in the sense of (RD3), but not in the sense of (RD2) given that the property is a necessary existent while Socrates is not (see Fine 1995, p. 274). The very same example shows that (RD3) does not entail (RD1). Conversely, (RD1) does not entail (RD3): while Socrates could not have existed without the empty set (because the latter is a necessary existent), Socrates is not essentially related to sets. Thus, the notions expressed by (RD1)–(RD3) are arguably all distinct.

From here on, let us say that fact *f partially grounds* fact *g* when either *f* on its own grounds *g*, or *f* and other facts together ground *g*. (Thus, 'partial' is understood in a permissive sense.) Benjamin Schnieder (2006b) and I (2005) independently introduced similar relations of ontological dependence defined in terms of partial grounding.[12] As an illustration, among the relations of that sort that I introduced is the important binary relation expressed by the open sentence

(RD4) $\quad \Box(x \text{ exists} \supset \exists G$ the fact that Gy partially ground the fact that x exists)

(in quasi-English: x cannot exist unless its existence is partially grounded in some fact about y) where the quantifier '$\exists G$' ranges over existence-entailing conditions (Correia 2005, p. 66). Given this requirement on the second-order quantifier, and assuming that grounding is factive, (RD4) entails (RD1). The converse entailment obviously fails. Correia (2005, pp. 78–79), argues in effect that the only entailments between (RD1), (RD2), (RD3) and (RD4) are those that have been stated in this and the previous paragraph. The situation is described in the following diagram, where an item entails a distinct item iff there is an arrow from the former to the latter:



Here is a list of commonplace ontological dependency claims:

Tropes (e.g. the particular whiteness of a sheet of paper)	depend on	Their bearers (the sheet of paper)
Events (e.g. the 1996 Summer Olympics 100 m final's second false start)		Their participants (Linford Christie)
Sets (e.g. {1,2})		Their members (1 and 2)
Wholes (e.g. Tibbles the cat)		Their parts (e.g. Tibbles' tail)
Biological organisms (e.g. Queen Elizabeth II)		Their biological origins (the particular sperm and egg she originated from)
Material artefacts (e.g. a particular wooden table)		The pieces of matter they were made from (the hunk of wood it was made from)
Non-divine things (e.g. you and me)		God
Things numerically distinct from the cosmos (e.g. all the atoms that there are)		The cosmos

In each example, one can argue that the dependent objects depend on their dependees in *all* the senses of 'depend' encountered so far, including the very strong senses defined by (RD2) and (RD4).

The notions of dependence defined so far are sometimes labelled 'rigid' and are contrasted with notions of so-called "generic" dependence. The generic counterparts of the rigid notions defined by (RD1)–(RD4) can be expressed by means of the following forms:

(GD1) $\Box(Ex \supset \exists \gamma F\gamma)$
(x cannot exist without an F)

(GD2) $\Box_x(Ex \supset \exists \gamma F\gamma)$
(x can, by its very nature, not exist without an F)

(GD3) $\exists R\Box_x\exists \gamma(F\gamma \,\&\, Rx\gamma)$
(For some R, it is part of the nature of x that x is R-related to some F)

(GD4) $\Box(x$ exists $\supset \exists \gamma(F\gamma \,\&\, \exists G$ the fact that $G\gamma$ partially grounds the fact that x exists))
(x cannot exist unless its existence is partially grounded in some fact about some F)

Simple claims of rigid dependence have the logical form 'xDy', where both 'x' and 'y' are nominal variables, and are thus naturally understood as stating that two objects stand in a certain binary relation. By contrast, simple claims of generic dependence have the form 'xDF' where 'x' is a nominal variable and 'F' a *predicate* variable, and hence, according to orthodoxy at least, cannot be understood in the same way. Thus, according to orthodoxy, 'Socrates generically depend on water molecules' is *not* to be understood as stating that Socrates is related in some way to a given entity, e.g. the property of being a water molecule or the kind WATER MOLECULE.

Generic dependence is, as it were, less demanding than rigid dependence. Let me illustrate by focussing on rigid dependence in the sense of (RD2) and generic dependence in the sense of (GD2). What do I *rigidly* depend on? Perhaps nothing at all, or nothing at all except myself. Or perhaps my biological origins and some of my parts, say my brain, and the things they themselves rigidly depend on. What do I *generically* depend on? Plausibly, each case of rigid dependence gives rise to cases of generic dependence. Thus, for instance, if I rigidly depend on my biological origins, then since each of the items I originate from is essentially F for some F (a sperm, an egg, a biological cell, a concrete entity, . . .), then I generically depend on Fs for every corresponding F. But many facts of generic dependence are not accompanied in this way by corresponding facts of rigid dependence. Thus, presumably, I generically but not rigidly depend on carbon atoms, water molecules, quantities of blood, heartbeats, human bodies, spatial locations, temporal locations, conscious experiences, beliefs and so on (remove or add items as you see fit).

We briefly went through a number of applications of the notions of rigid and generic dependence. An important application of the concept of *generic* dependence that has not been mentioned so far is the characterisation of the distinction between the Aristotelian and the Platonic conceptions of universals: on the Aristotelian conception, universals generically depend on their exemplifiers (e.g. wisdom generically depends on wise beings), while on the Platonic conception, they do not. Two, related, important applications of the concept of *rigid* dependence have not been mentioned either: the characterisation of substances as those entities that are in some sense ontologically independent from certain entities, and the definition of what it is to be fundamental. On the use of notions of dependence to characterise the Aristotelian/Platonic divide about universals and to characterise substances, as well as on other uses illustrated previously, see Correia 2008, §2, and the references mentioned therein. On dependence and fundamentality, see Section 9.4.

9.2.2 Connections with modality

The connections between the eight particular notions of ontological dependence introduced earlier in the chapter—(RD1)–(RD4) and their generic counterparts (GD1)–(GD4)—and modality are to a certain extent obvious since, as it were, these notions wear their modality on their sleeves. (RD1), (RD4), (GD1) and (GD4) directly begin with a necessity operator. The connection between the other notions and modality is slightly less direct, as it goes through the general connection between essence and modality. Earlier I have assumed that this general connection is captured by the following general principle:

If $\Box_x p$, then $\Box(x \text{ exists} \supset p)$ Ess-Ex

An alternative option is to opt for the following principle, which is stronger than the previous one:

If $\Box_x p$, then $\Box p$ Ess-Ex*

Note that given either Ess-Ex or Ess-Ex*, (RD2) entails (RD1) and (GD2) entails (GD1).

9.3 Partial grounding vs the converse of ontological dependence between facts

The idiom of dependence is sometimes used in such a way that the biconditional

g is partially grounded in *f* iff *g* depends on *f* Ground-Dep

is taken to be analytic, at least on some specific, metaphysical or ontological construal of 'depends' (see e.g. Rosen 2010 and Schnieder 2020). This is unfortunate given the way 'ontological dependence' is ordinarily understood in current metaphysical debates, for it seems indeed that both directions of Ground-Dep can be consistently rejected on any standard construal of 'depends on'. Let me illustrate this with the core construals defined by (RD1)–(RD4).[13]

Let us first focus on the left-to-right direction of the biconditional. Consider f = the fact that Socrates is a philosopher and g = the fact that someone is a philosopher. (For a different example, take f = the fact that the moon is white and g = the fact that the moon is coloured.) According to grounding orthodoxy, f grounds g. Yet since there could have been philosophers without Socrates being one of them, g could have existed without f, and hence g depends on f neither in the sense of (RD1) nor in the sense of (RD2). The view that g also does not depend on f in the sense of (RD3) has a strong intuitive pull: as Fine (2012, p. 75) puts it, '[the fact that someone is a philosopher], so to speak, knows nothing of Socrates'.[14] Finally, it is hard to see how it could be maintained that g depends on f in the sense of (RD4), for in a world where there are philosophers but Socrates is not among them, which fact about f could possibly help ground the existence of g?

Let us now turn to the right-to-left conditional. Here we consider three pairs of facts:

f_1 = the fact that {Socrates} exists; g_1 = the fact that Socrates exists
f_2 = the fact that the actual world exists; g_2 = the fact that God exists
f_3 = the fact that 1 is 0's successor; g_3 = the fact that f_3 exists in virtue of the nature of 1

The view that f_1 plays a role in grounding g_1 is utterly implausible. Yet necessarily, {Socrates} exists if Socrates does, and therefore g_1 depends on f_1 in the sense of (RD1). Consider then the following two substantial yet seemingly jointly tenable assumptions concerning a Leibnizian

god: that his existence is ungrounded, and that he is by his very nature causally responsible for the existence of the actual world.[15] Suppose then that the divinity that g_2 is about satisfies these two conditions. By the first assumption, f_2 does not partially ground g_2, and the second assumption may be used to support (even though it does not entail) the view that g_2 depends on f_2 in the sense of (RD2), and consequently also in the sense of (RD1) and in the sense of (RD3).[16] Finally, it can hardly be maintained that f_3 partially grounds g_3, and yet it is plausible to hold that g_3 depends on f_3 in the sense of (RD2), and hence also in the sense of (RD1) and in the sense of (RD3). A case can actually be made that g_3 also depends on f_3 in the sense of (RD4), based on the following, plausible principle:

> Necessarily, if the fact that *p* exists, then the fact that the fact that *p* exists is grounded in the fact that *p*.

(The principle can be formulated in a significantly less cumbersome way if we use the 'because' idiom: Necessarily, if the fact that *p* exists, then this is because *p*.) By the proposed principle, necessarily, if g_3 exists, then the fact that g_3 exists is grounded in g_3; since necessarily, g_3 is a fact about f_3, and since 'exists in virtue of the nature of 1' is existence-entailing, it indeed follows that g_3 also depends on f_3 in the sense of (RD4).

I have assumed throughout this section that existence and obtainment are equivalent features for facts, more precisely that

> For a fact to exist is for it to obtain — Ex-Ob

If Ex-Ob fails, it is natural to introduce the following three notions of "obtainment-dependence" corresponding to (RD1), (RD2) and (RD4), respectively:

(RD1*) $\Box$(*x* obtains $\supset$ *y* obtains)
(RD2*) $\Box_x$(*x* obtains $\supset$ *y* obtains)
(RD4*) $\Box$(*x* obtains $\supset \exists G$ the fact that *Gy* helps ground the fact that *x* obtains)

Given Ex-Ob, each (RD$i^\star$) with $i \in \{1,2,4\}$ boils down to (RDi), but if Ex-Ob fails, then the three relations of obtainment-dependence are certainly new. Let us, then, suppose that Ex-Ob fails. My previous objections to the view that Ground-Dep holds with dependence understood along the lines of (RD1), (RD2) or (RD4) obviously carry over, *mutatis mutandis*, the view that Ground-Dep holds with dependence understood along the lines of (RD1*), (RD2*) or (RD4*). But what shall we say about the original view? A detailed reply would have to take into account various metaphysical conceptions of facts, but given the previous material, it should already be clear that the prospects for the view in question are dim.[17]

9.4 Grounding, ontological dependence and fundamentality

Both grounding and ontological dependence are commonly thought to be connected to the notion of fundamentality, in the way suggested by the following two general principles:

> If *g* is partially grounded in *f*, then *f* is more fundamental than *g* — Ground-Fund
> If *y* ontologically depends on *x*, then *x* is more fundamental than *y* — Dep-Fund

Any relation of *being more fundamental than* must be asymmetric, and hence irreflexive. Given Ground-Fund, asymmetry should transmit from *being more fundamental than* to partial grounding, and the same is true of ontological dependence given Dep-Fund. Although orthodoxy has it

that partial grounding is asymmetric, some considerations have been put forward against that view.[18] Ontological dependence in the sense of (RD1) is clearly not asymmetric (everything depends on itself in this sense, and, still in this sense, {Socrates} and Socrates plausibly depend on each other), and it can be argued that there are cases of self-dependence and cases of mutual dependence between distinct entities, where dependence is understood in the sense of (RD2), (RD3) or (RD4).[19] If such threats to the asymmetry of partial grounding or ontological dependence are taken seriously, it is always possible to claim that it is not the relation itself but its one-sided counterpart which is connected to relative fundamentality in the way orthodoxy suggests, where an entity y is one-sidedly R-related to entity x iff_{df} y is R-related to x and x is not R-related to y. For any relation R, one-sided R is indeed by definition asymmetric (and hence irreflexive).

Assuming that partial grounding (or its one-sided variant) and various relations of ontological dependence (or their one-sided variants) are connected to relative fundamentality in the way suggested by Ground-Fund and Dep-Fund, it is certainly not to be expected that the *same* notion of relative fundamentality will be involved in all cases. Indeed, it can be argued that the assumption leads to the view that there are various dimensions of relative fundamentality.

For consider again the example involving the actual world and the Leibnizian god from the previous section. The second assumption we made there supports the view that God depends on the actual world in the sense of (RD2). Now we can consistently add to our set of assumptions the further assumption that necessarily, if the actual world exists, then its existence is partially grounded in some feature of God, say his existence. This further assumption entails that the actual world depends on God in the sense of (RD4). If both notions are connected to relative fundamentality in the way suggested by Dep-Fund, then we have to say *both* that the actual world is more fundamental than God *and* that God is more fundamental than the actual world; but this cannot be if the two occurrences of 'more fundamental than' mean the same. The further assumption just introduced yields a further illustration of the same point. Given this assumption, f_2 is partially grounded in g_2, and hence by Ground-Fund, g_2 is more fundamental than f_2. Yet as we saw in the previous section, g_2 depends on f_2 in the sense of (RD2), and hence by Dep-Fund, f_2 is more fundamental than g_2. Again, this cannot be the case if the two occurrences of 'more fundamental than' mean the same.

Similar considerations hold for the connection between partial grounding and various relations of ontological dependence on one hand, and the notion of *absolute fundamentality* on the other hand. Consider the following definitions, corresponding to two standard ways of defining 'fundamental' (here as with relative fundamentality, there might be reasons to invoke *one-sided* partial grounding and *one-sided* ontological dependence, but I leave this aside for the sake of simplicity):

Fact f is $\text{fundamental}_{\text{G}}$ iff_{df} f is not grounded	Fund-Ground
Object x is $\text{fundamental}_{\text{D}}$ iff_{df} x is not ontologically dependent	Fund-Dep

Fund-Dep is a definitional schema rather than a proper definition, each particular relation of ontological dependence giving rise to its own corresponding concept of $\text{fundamentality}_{\text{D}}$, and one can argue that various relations of dependence give rise to extensionally distinct concepts of $\text{fundamentality}_{\text{D}}$. Thus, for instance, given assumptions we previously made, God is $\text{fundamental}_{\text{D}}$ in the sense of (RD4) but not in the sense of (RD2).

$\text{Fundamentality}_{\text{G}}$ and $\text{fundamentality}_{\text{D}}$ also diverge in interesting ways. Consider a view according to which some facts are literally composed of objects and properties or relations (see e.g. Russell 1918). Such facts are plausibly dependent on their constituents, and hence are not $\text{fundamental}_{\text{D}}$, in some—if not all—of the senses captured by (RD1)–(RD4). Now of course,

this is compatible with the view that some such facts—say the fact composed of a given electron and the property of having a charge of −1e, or the fact composed of God and the property of existing—are ungrounded, and hence fundamental$_G$. Conversely, there appears to be room for the view that some facts are fundamental$_D$ in some interesting sense without being fundamental$_G$. Fregean facts, i.e. true Fregean thoughts, Schnieder (2020, §4.5) argues, are existentially independent. We may take this to entail at least that they are not ontologically dependent in the sense of (RD2) or in the sense of (RD4). Yet such a conception of facts would seem to be compatible with the view that some facts are grounded, e.g. with the view that the fact that someone is a philosopher is grounded in the fact that Socrates is a philosopher.

Thus, it would seem, Fund-Dep yields various notions of absolute fundamentality, and Fund-Ground yields still another such notion.

Acknowledgements

The research that led to this publication was funded by the Swiss National Science Foundation through project BSCGI0-157792.

Notes

1 Several useful review articles on ontological dependence and grounding are available. On ontological dependence, the reader can consult Correia (2008), Koslicki (2013) and Tahko and Lowe (2015), and on grounding, Correia and Schnieder (2012), Trogdon (2013), Bliss and Trogdon (2014) and Raven (2015). The content of this chapter inevitably overlaps to a significant extent material covered by these articles. Another useful item, that was not available to me at the time of drafting this chapter, is Skiles (2020) on the connections between grounding and necessity.

2 For reasons of space, I will leave aside scepticism about grounding—and, for that matter, scepticism about ontological dependence. The review articles about grounding cited in the previous note all address the issue of scepticism about grounding. Scepticism about ontological dependence stems from scepticism about notions that are used to precisify the concept, in particular grounding, and essence understood as non-reducible to modality.

3 Jonathan Schaffer also extensively uses the predicate 'grounds' and variants thereof (see e.g. Schaffer 2009), but it is plausible that what he has in mind when using these locutions is ontological dependence rather than grounding. But see Schnieder (2020, §4) for a discussion.

4 The choice of a form among (1) to (5) to regiment grounding will arguably be guided by more than considerations of the kind just alluded to. Thus, for instance, some hold that to ground *just is* to explain, in a suitable sense of 'explain', while others take grounding to *undergird* explanations (see Raven 2015, p. 326, for references and a discussion; see also Dasgupta 2017). Only the former will take forms (4) and (5) seriously.

5 The contrast between the operator framework and the predicate framework for factual grounding also applies to generic grounding: the contrast can be captured by means of the distinction between the forms 'The property of being *F* grounds the property of being *G*' and 'Something's being *F* makes it be *G*'. The operator framework for both factual and generic grounding is actually the one favoured both by Fine (2015) and Correia and Skiles (2019). Just like the regimentation of factual grounding using a sentential operator as in (5) from earlier does not commit one to an ontology of facts, the regimentation of generic grounding using a predicational operator does not commit one to an ontology of properties.

6 This is the majority view, but see Section 9.4 in this chapter for references on dissenting voices.

7 The worldly vs representational distinction also applies in the operator framework, but there, of course, it cannot be cashed out in terms of the kinds of relata that grounding relates. See Correia (2010, pp. 257–258), and Fine (2017b, pp. 685–686). I defend a worldly conception in Correia (2010, pp. 258–259), but I now believe that there is room for both conceptions. Kit Fine (2012) defends a representational conception, although the semantics for the logic of grounding that he introduces in that paper is only suited for a worldly conception. For a thorough discussion of the worldly vs representational distinction, see Correia (2020).

8 Here and in what follows, I assume that grounding is not factive, but it is clear how the discussion would go if factivity were assumed instead.
9 I have in mind notions of entailment regimented by the relevant logics E and NR (see Mares 2014).
10 I am here ignoring the fact that Fine favours an operator regimentation of grounding.
11 In this section, I heavily draw on Correia (2008).
12 To be accurate, Schnieder works with a notion of grounding that is conceptual in character, rather than with the notion of metaphysical grounding. Similar notions are discussed in Fine (1982) and Lowe (1998, pp. 145 ff).
13 Grounding is certainly a *determination* relation: if *f* grounds *g*, then *f*'s being the case makes *g* be the case, and hence there is a clear sense in which *f* can be said to determine *g*. In some debates, most notoriously perhaps in debates about causation, 'dependence' and 'determination' are understood as expressing converse relations. These two points may jointly help explain why Ground-Dep is sometimes simply taken for granted.
14 I am saying that the pull is strong, not that it is irresistible. I indeed argue elsewhere (Correia 2013) that it can be resisted.
15 Here and in what follows, I use 'the actual world' as a rigid designator for our world.
16 Appeal to the actual world/Leibnizian god pair is made in Correia (2005, p. 67), to make a related but different point, namely that something (in the proposed example: God) may depend on some other thing (the actual world) in the sense of (RD2) while failing to depend on it in the sense of (RD4).
17 Schnieder (2020, §4.6) objects to both directions of Ground-Dep with ontological dependence understood as existential dependence (what this exactly amounts to is left to a certain extent underspecified), precisely in a context where fact-existence and obtainment are not taken to be equivalent.
18 Similar arguments found in Fine (2010), Krämer (2013) and Correia (2014) threaten the irreflexivity of partial grounding. Jenkins (2011) indirectly challenges the view that grounding is irreflexive, by challenging the view that if *f* grounds *g*, then *f* and *g* must be distinct (in order to avoid any misunderstanding: what Jenkins calls 'irreflexivity' is the property attributed to grounding by the latter view). Of course, if Leibniz's Law applies to grounding contexts, then the two views are equivalent.
19 See Correia (2005, pp. 80–81).

Bibliography

Audi, P. (2012). A Clarification and Defense of the Notion of Grounding. In F. Correia & B. Schnieder (eds.), *Metaphysical Grounding: Understanding the Structure of Reality*, Cambridge: Cambridge University Press, 101–21.

Blackburn, S. (1987). Morals and Modals. In C. Wright & C. McDonald (eds.), *Fact, Science, and Value*, Oxford: Blackwell.

Bliss, R. & Trogdon, K. (2014). Metaphysical Grounding. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, https://plato.stanford.edu/archives/win2016/entries/grounding/.

Cameron, R. (2010). On the Source of Necessity. In B. Hale & A. Hoffman (eds.), *Modality: Metaphysics, Logic and Epistemology*, Oxford: Oxford University Press, 137–51.

———. (2018). Truthmakers. In M. Glanzberg (ed.), *The Oxford Handbook of Truth*, Oxford: Oxford University Press, 333–54.

Correia, F. (2005). *Existential Dependence and Cognate Notions*, Munich: Philosophia.

———. (2008). Ontological Dependence. *Philosophy Compass*, 3(5), 1013–32.

———. (2010). Grounding and Truth-Functions. *Logique et Analyse*, 53(211), 251–79.

———. (2013). Metaphysical Grounds and Essence. In B. Schnieder, M. Hoeltje & A. Steinberg (eds.), *Varieties of Dependence: Ontological Dependence, Grounding, Supervenience, Response-Dependence*, Munich: Philosophia, 271–96.

———. (2014). Logical Grounds. *Review of Symbolic Logic*, 7(1), 31–59.

———. (2017). Real Definitions. *Philosophical Issues*, 27(1), 52–73.

———. (2020). Granularity. In M. Raven (ed.), *Routledge Handbook of Metaphysical Grounding*, New York: Routledge, 228–43.

Correia, F. & Schnieder, B. (2012). Grounding: An Opinionated Introduction. In F. Correia & B. Schnieder (eds.), *Metaphysical Grounding: Understanding the Structure of Reality*, Cambridge: Cambridge University Press, 1–36.

Correia, F. & Skiles, A. (2019). Grounding, Essence, and Identity. *Philosophy and Phenomenological Research*, 98(3), 642–70.

Dasgupta, S. (2014). On the Plurality of Grounds. *Philosopher's Imprint*, 14(20), 1–28.
———. (2017). Constitutive Explanation. *Philosophical Issues*, 27, 74–97.
Fine, K. (1982). *Dependent Objects*. Unpublished manuscript.
———. (1994). Essence and Modality. *Philosophical Perspectives*, 8: 1–16.
———. (1995). Ontological Dependence. *Proceedings of the Aristotelian Society*, 95(3), 269–90.
———. (2010). Some Puzzles of Ground. *Notre Dame Journal of Formal Logic*, 51(1), 97–118.
———. (2012). Guide to Ground. In F. Correia & B. Schnieder (eds.), *Metaphysical Grounding: Understanding the Structure of Reality*, Cambridge: Cambridge University Press, 37–80.
———. (2015). Unified Foundations for Essence and Ground. *Journal of the American Philosophical Association*, 1, 296–311.
———. (2017a). A Theory of Truthmaker Content I: Conjunction, Disjunction and Negation. *Journal of Philosophical Logic*, 46(6), 625–74.
———. (2017b). A Theory of Truthmaker Content II: Subject-Matter, Common Content, Remainder and Ground. *Journal of Philosophical Logic*, 46(6), 675–702.
Hale, B. (2010). The Source of Necessity. *Noûs*, 36(16), 299–319.
Jenkins, C. (2011). Is Metaphysical Dependence Irreflexive? *The Monist*, 94, 267–76.
Koslicki, K. (2013). Ontological Dependence: An Opinionated Survey. In B. Schnieder, M. Hoeltje & A. Steinberg (eds.), *Varieties of Dependence: Ontological Dependence, Grounding, Supervenience, Response-Dependence*, Munich: Philosophia, 31–64.
Krämer, S. (2013). A Simpler Puzzle of Ground. *Thought*, 2(2), 85–89.
Leuenberger, S. (2014). Grounding and Necessity. *Inquiry*, 57(2), 151–74.
Litland, J. (2015). Grounding, Explanation, and the Limit of Internality. *Philosophical Review*, 124(4), 481–533.
———. (2016). Pure Logic of Many-Many Ground. *Journal of Philosophical Logic*, 45(5), 531–77.
Lowe, E. J. (1998). *The Possibility of Metaphysics*. Oxford: Clarendon Press.
Mares, E. (2014). Relevance Logic. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, https://plato.stanford.edu/archives/spr2014/entries/logic-relevance/.
Raven, M. (2015). Ground. *Philosophy Compass*, 10(5), 322–33.
Rodríguez-Pereyra, G. (2015). Grounding Is Not a Strict Order. *Journal of the American Philosophical Association*, 1(3), 517–34.
Rosen, G. (2010). Metaphysical Dependence: Grounding and Reduction. In B. Hale & A. Hoffman (eds.), *Modality: Metaphysics, Logic, and Epistemology*, Oxford: Oxford University Press, 109–36.
———. (2015). Real Definition. *Analytic Philosophy*, 56, 189–209.
Russell, B. (1918). "The Philosophy of Logical Atomism", *The Monist*, 28(4), 495–527.
Schaffer, J. (2009). On What Grounds What. In D. Chalmers, et al. (eds.), *Metametaphysics*, Oxford: Oxford University Press, 347–83.
Schnieder, B. (2006a). Truth-Making without Truth-Makers. *Synthese*, 152, 21–47.
———. (2006b). A Certain Kind of Trinity: Dependence, Substance, Explanation. *Philosophical Studies*, 129, 393–419.
———. (2020). Grounding and Dependence. *Synthese*, 197, 95–124.
Skiles, A. (2015). Against Grounding Necessitarianism. *Erkenntnis*, 80, 717–51.
———. (2020). Necessity. In M. Raven (ed.), *The Routledge Handbook of Metaphysical Grounding*, New York: Routledge, 148–63.
Tahko, T. & Lowe, E. J. (2015). Ontological Dependence. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, https://plato.stanford.edu/archives/win2016/entries/dependence-ontological/.
Trogdon, K. (2013). An Introduction to Grounding. In B. Schnieder, M. Hoeltje & A. Steinberg (eds.), *Varieties of Dependence: Ontological Dependence, Grounding, Supervenience, Response-Dependence*, Munich: Philosophia, 97–122.
Witmer, D. G., Butchard, W. & Trogdon, K. (2005). Intrinsicality without Naturalness. *Philosophy and Phenomenological Research*, 70: 326–50.

Chapter 10
MODALISM

Scott A. Shalkowski

Introduction

The philosophy of modality divides into three main areas: semantics, metaphysics, and epistemology. The semantics of modality divides into two main interpretations of modal discourse: realist and anti-realist. Realist interpretations similarly divide into two: reductivist and non-reductivist. Modalism is a realist, non-reductivist thesis about modal discourse. The metaphysical reality to which truth-directed modal discourse is aimed has no "deeper" explanation. The modal character of reality is irreducible. At its most fundamental, reality is modal in nature. Having argued for primitive modality, modalism has been implemented to illuminate propositions (Fine 1980), sets (Fine 1981; Forbes 1985), facts (Fine 1982), individual essences, substances, properties, events (Forbes 1985), to clarify and to justify logical pluralism (Bueno and Shalkowski 2009), and to provide an account of the logical constants (Bueno and Shalkowski 2013).

We should resist the temptation to conflate semantic irreducibility with metaphysical irreducibility. Strictly speaking, semantic issues are not issues about the nature of modality. They are issues about the meanings of concepts or expressions, not directly about what it is to be thing or attribute referred to or characterized by those concepts or expressions. For those thinking of language as complex, abstract, platonic objects, semantic issues pertain to the intrinsic and relational character of *those* objects. Treating meaning anthropocentrically is to treat it as attaching directly to mental states or, more conventionally, to linguistic expressions. Those expressions may be ordinary, such as 'necessary' or 'possible', or formal expressions, such as '□' or '◊'. Either way, on the anthropocentric view, issues of meaning are issues about us, our mental states, inscriptions, linguistic habits, and the like. Nothing internal to theories of meaning requires that either platonic or anthropocentric linguistic facts are especially illuminating about the natures of things. The history of human intellectual endeavors is partly a history of finding or creating linguistic tools to represent reality accurately, involving significant change as it has.

Absent some good reason to think that there is pre-established or otherwise inevitable harmony between the structures of modal reality and the platonic structures of our languages or some well-founded confidence that we have established such harmony ourselves for structures we have created, semantic issues are at best indirectly relevant to understanding the structure of reality itself. Thus, the remainder of this chapter focuses directly on the thesis that reality is itself modal.

Reductive motivations

Typically, reductive theories are prompted by epistemological, ontological, or ideological considerations. Epistemological concerns might take the rather weak form of demanding merely that metaphysical claims are "joined up" with a credible account of how those claims are to be known. (See Sonia Roca-Royes's entry on this "integration challenge" in this *Handbook*.) More ambitious and less philosophically secure is some commitment to empiricist epistemological standards, along with the common assumption that only how things *are* and not how things *must* or *could* be can be observed. If so, then if reality has any modal character to which we are rationally entitled to believe, that character must be a matter of the way things are, non-modally.

A separable metaphysical motivation is the belief that reality is categorical. Those who find thinking about modal matters illuminated by a framework of possible worlds can maintain that our (actual) world is just the totality of how things are. How things could be but are not is a matter of how things are in some other world(s), and how things must be is a matter of how things are in all worlds, niceties about which worlds are "accessible" from any given world(s) aside. Swans are white or black in ours, but blue in some other, perhaps. Rome's having been sacked in 1527 and Washington's likeness appearing on Mount Rushmore are facts of our world, with no apparent implications for other worlds. No modal expression appears in the sentences we use to express those facts. Since our world is not unique on this count, reality is fundamentally categorical, and modal matters must themselves be matters of the wider categorical realities of possible worlds. The clearest example is David Lewis's development of "genuine modal realism" (Lewis 1986), with a rather comprehensive treatment of the philosophy of possible worlds found in Divers (2002).

The tools we use to formulate claims about reality are sometimes the subject of philosophical orderings. Those of "Okhamist" tendencies may take "Don't multiply entities beyond necessity" as a genuine imperative. If they treat their theories as at least aiming at the truth, then the imperative's rationale must be that ontologically more parsimonious theories are more probably correct. The parsimony that tends toward correctness, however, could be ideological rather than ontological. If reality's tale can be told adequately in non-modal terms, then the modal may be accounted for, if at all, by way of reduction to reality's non-modal character. The development of philosophy along the lines of ideological parsimony is clear in Sider (2011).

These motivations are compatible. More significantly, each is controversial. It might well be incumbent on proponents of any philosophical program to address the applicable version of the integration challenge so that rational entitlement to the program's primary claims is not mysterious by, at least, the lights of the program's proponents. Empiricist answers are not, so far as the philosophy of modality is concerned, demanded. Only that some accounting of how one could come to believe rationally what one believes about the possible is required.

Though the appearance of non-modal reality is easy to achieve by failing to use modal expressions, the accuracy of that appearance deserves scrutiny. It is easy enough to affirm that some object, *a*, is *F*. '*Fa*' contains neither '□' nor '◊'. Similarly, for 'Kermit is green'. Does what it is to be Kermit carry modal content that is typically hidden from view? Perhaps he is not only a frog, but must be one, or must be one if he exists at all. His being green is a matter of what? How he appears now? In the dark he does *not* appear green, so perhaps it is a matter of how he *would* appear in the appropriately lit circumstances when observed by observers sufficiently like the typical likes of us. Treating greenness as a matter of reflectance falls for the same reason. It might be that Kermit's surface has one of a number of physical structures, rather than that his surface *is* or *is observed to be* doing anything. This will require further analysis of what it is to have

that structure, composed as it is of chemicals, themselves with constituents that, like electrons, seem to be what they *are* as a matter of what they *would* do in, say, magnetic fields.

Ideological efficiency might be a virtue for some purposes, but not all. Axiomatic presentations of logic are useful when metalogical results are the primary concern. Object theory proofs, however, are typically more tortuous when the only available rule of inference is *modus ponens*, than they are when more plentiful "natural deduction" rules are available. Inquiries into the nature of the modal are truth-directed. Their proponents aim to provide an accurate account of reality. It is not perfectly obvious, however, that maximizing ideological economy best facilitates reality's accurate accounting. Too few resources make that account impossible while unnecessary resources will, at best, engender redundancy. Conceding that the realist's project regarding the metaphysics of modality is one of, in Sider's terms, "carving reality at its joints", the correctness of an ideologically parsimonious theory cannot be inferred from its relative parsimoniousness. That the more parsimonious theory can deliver comprehensive correctness requires the adequacy of the parsimonious ideology, which must be demonstrated rather than assumed. Not multiplying conceptual resources beyond necessity is not a virtue that permits virtuous theories to wear their alethic merits on their parsimonious sleeves. Only after determining reality's modal character can a theory be deemed virtuously parsimonious rather than viciously mean or profligate.

Since slogans, preferences, and orientations will not settle the nature of reality's modal character, the best course is to attempt reductions of the modal in terms of the non-modal. Reductions require two components. First, there should be a tolerable correlation of modal reality with non-modal reality. Proposed reductions (or, for that matter, their rejections) should not rest on too much disputed metaphysics, if the case is to be widely persuasive. Second, there should be good grounds for thinking that the modal really is a matter of the non-modal, rather than for thinking that the non-modal is itself constrained by the modal. Modalists, by virtue of maintaining that reality is fundamentally modal, contend that modality is inescapable. They maintain that the non-modal resources available to reductionists are too poor, either metaphysically or epistemologically. The bases for reduction must themselves be modal for the account to be extensionally adequate, but thereby undermining its reductive adequacy, or else there is no good, principled way to inspire confidence that the relevant non-modal reality suffices for all possibilities while also not inadvertently including what is not possible (Shalkowski 1994, 2004). Following the very brief presentation and critiques of two reductive projects of both intuitive and historical significance, is a third that has had tremendous recent influence. The modalist's response is then detailed, with a concluding note on a common philosophical strategy on which recent reductionism has depended.

Conceivability theories

If metaphysics and its attending epistemology are to be wedded to discharge the integration challenge, ideal would be a reduction of modality in epistemological terms. Conceivability theories are recommended precisely for this reason. If the possible is the conceivable, then the impossible is what cannot be conceived, and necessary is what cannot be conceived to be otherwise.

As attractive as this is for conducting armchair modal metaphysics (see Yablo 1993), it is a non-starter as a reductive program. First, there are no grounds for thinking that the domain of the conceivable is sufficiently rich to accommodate the entirety of the modal. Typically, it is assumed that the conceivers in question—because we are here trying to address our own philosophical questions and puzzles—are us: humans with their obvious limitations. Beyond those limitations, the question "Conceivable by whom?" matters, since humanity is not monolithic. There is well-known variety regarding cognitive abilities of all sorts, including the conceptual.

Second, to the degree that it is plausible that what has been conceived, no matter whether by some chosen individual or by someone or other, does not yet exhaust what will be conceived, and to the degree that caution should be exhibited regarding whether some hard limit has been reached at any given time, caution should be exercised whether at any given time the results of conceiving really do exhaust the modal. Indeed, to the degree that objectivity of the modal is maintained and that the project of modal metaphysics is one of discovery rather than invention, this program appears to be ill-conceived precisely because conceptual achievements are achieved on the basis of what has gone before. As developments in mathematics facilitate later development, so it is for conceptual achievements more generally.

Third, being hostage to the limitations of actual acts of conceiving is solved by basing a reduction on *ideal* conceivability (Chalmers 2002), but with the consequence that proposals along this line break conceivability's apparent ability to solve *our* integration challenge, since we are far from ideal conceivers. Hope might not be lost completely, if there is yet a further story about how an ideal conceiver illuminates our acts of conceiving, but that is a story not usually told.

Fourth, if apparently modal vocabulary signals the modal character of the reality characterized with the vocabulary, then while appeals to the *conceivable* rather than to the conceived, avoids the problem just mentioned, the facts of conceivability are ill-suited to the task of accounting for the nature of the modal, modal as it is itself. (Fuller treatment is found in M. Oreste Fiocco's entry for this volume.)

Linguistic theories

Appeals to conceivability are most natural when thinking of accounts that give pride of place to possibility, with necessity being defined. Finding the source of modality in language and/or linguistic conventions is more natural when emphasizing the source of necessity, with possibility defined. Attempts in this theoretical neighborhood can be found in Ayer (1946) and Carnap (1954, 1956). Notable critiques are Pap (1958) and Quine (1936, 1951, 1963).

Some subtlety is required here. It is not merely a matter of taste to speak once of the meaning of the concept TRIANGLE and again of the meaning of the term 'triangle' If concepts are platonic entities to which we bear a relation in order to "grasp" that concept, then concepts are not mental representations "in the head", nor are they complex social practices encoded in dictionary entries for 'triangle'. Each orientation has its own challenges.

Platonism, generally, faces integration challenges, most notably formulated by Paul Benacerraf (1965, 1973). Some accounting of the relation with the relevant abstracta is called for, if this is to be a serious theory of the fundamentals of modality in terms of meaning. Additionally, some accounting is required of how objects like the abstract concept TRIANGLE could serve to constrain the characteristics of concrete items said to "fall under" those concepts. The additional ontology is not an immediate solution to the problem at hand.

Mental representations are unsuitable, being limited as were acts of conceiving. If dinosaurs could or could not have done such-and-so, the mental representations the contents of which permit or require that could not be those of us who came much later. Provisos similar to those required for ideal conceiver theories would be required for mental representors more up to the modal tasks than are we.

Under this heading are strategies for finding (necessary) truth by convention. Linguistic complexity may be reduced by stipulating explicit definitions of new terms. Entailments arising from those definitions are good candidates for necessary truth, such as "All bachelors are unmarried". It is by no means obvious, however, that such an account can be comprehensive and

appropriate for a full reduction. Quine (1936, 1963) argued that some logical truths—which, themselves must be necessary, if their definitional transformations are to be—are assumed in this approach. For a more recent defense, see Sidelle (1989). Whatever there is to recommend a thorough-going conventionalism about necessity, it does not serve the realist project of uncovering an accounting of modality in more fundamental terms.

Possible worlds

Far and away the most popular and most fully developed reductive program regarding modality has been David Lewis's development of "genuine modal realism" (Lewis 1968, 1973, 1986). The central feature of Lewis's account of modality is his expansion of the domain of concrete objects to include not merely our own concrete universe in which Socrates is snub-nosed and Nixon went to China, but for each way that a concrete universe could be, there is one that is that way. Lewis's worlds are maximal spatio-temporal wholes, no two of which have any common parts, strictly speaking. Socrates could have had an aquiline nose not because our man Socrates also inhabits another possible world, but because our Socrates has a counterpart in some other world with an aquiline nose. Exactly what and how many other-worldly objects count as counterparts of Socrates need not detain us. What matters is that he has at least one in some world with an aquiline nose, which suffices for the possibility that our Socrates has an aquiline nose. If all counterparts of Socrates, scattered across the plurality of possible worlds as they may be, have a particular characteristic, then it is essential to Socrates. Anything that is so from the standpoint of every single world is necessarily so.

Implicit here is that all of the primary features of Lewis's worlds are non-modal. Required nuances just noted aside, the story of the possible and the necessary is the story of how the totality and the parts of this plurality of worlds *is*. *Being* is primary, with *truth* dependent on being. *Is* is primary, with *possibility* and *necessity* parasitic. The prospects of possible worlds for accounting for modality are best assessed in the light of perceived deficiencies brought by modalists.

Modalism

The primary contrast with reductive theories of modality is modalism, according to which modality is a primitive, irreducible feature of reality. No amount of *is* suffices to tell the story of possibility. Advocates have been Arthur Prior and Kit Fine (1977: esp. postscript), and Graeme Forbes (1985, 1989), with applications by Otávio Bueno and Scott A. Shalkowski (2009, 2013, 2015). If the critiques of conceivability and linguistic theories stand and if Lewis's expansion of the domain of worlds is deemed egregious or unwarranted, there seems to be little choice but to be either eliminativist about modality or to treat it as a feature of reality that stands on its own at the ground level, as it were.

Modalism has been cast by Fine and Forbes as a thesis about the *language* of possibility. Specifically, can the formal languages of modal logic, extended to include an actuality operator, 'A', suffice for formal treatments of ordinary modalizing? Concerning ourselves here only with the simplest sentential forms, can our ordinary claims involving possibility and necessity be expressed using '$\Diamond P$' (Possibly, *P*), '$\Box P$' (Necessarily, *P*), and 'A*P*' (Actually, *P*) respectively, or must we quantify over worlds to formalize the same claims using '$\exists$w*P*(w)' (there is a world at which *P* holds) and '$\forall$w*P*(w)' (all worlds are worlds at which *P* holds), '*P*' sufficing in any context of use to do the duty of 'A*P*'? What here has been treated primarily as a metaphysical issue is now treated as a linguistic issue. In particular, it has been treated as an issue of comparative expressive power. Are modal languages at least as expressively powerful as those quantifying over worlds?

This casting of issues is not without merit. If we begin with intellectual puzzlement about what our claims involving modal expressions could come to, we may really be interested in what accounts for the truth of some and the falsity of others. Many find it natural to generate in others understanding of our modalized claims by asking them to consider whether there is a possible world in which the alleged possibility is so. One genuinely bereft of a grasp of 'possibly' is unlikely to find much traction in 'possible world'. Lewis's framework has the advantage that in the context of modal puzzlement, there is no typically-attending puzzlement regarding claims about how things are, *qua* claims about how things are. His metaphysics simply involves more things that are than others had seriously supposed, and the content of the claim that there are these "extra" things is not, usually, seriously questioned. If he can articulate the *is* conditions under which our *possibly* claims hold, then he has accounted for the truth of said claims. Thus, while the ascent into a metalanguage is not the nature of the philosophy of modality, it is not unnatural to lapse into thinking of matters in terms of modal truths and their truth conditions, even if that is proxy for object language claims about how things stand. Fine (Prior and Fine 1977) and Forbes (1989) claim that Lewis's "possibilist" discourse can be reduced to or exchanged sufficiently well for modalist discourse.

Joseph Melia (1992, 2003: ch. 4) complains that in order to represent some ordinary modal claims, Forbes must not only introduce the actuality operator, but he must also index operators to track accurately the scopes of multiple operators, and that to understand how the indices work to serve that function, the modalist derives their meanings from an implicit quantification over worlds. Additionally, Melia worries that the framework Forbes develops cannot express adequately claims like "My car could have been the same [color] that yours actually is" because the 'actually' does not seem to be an operator on sentences, but part of the predicate about the color of your car (Melia 2003: 98).

Much to-ing and fro-ing will be required to adjudicate these matters properly, as was begun in Forbes (1992). They might, however, be bypassed by, at least in this instance, noting that sometimes expressions like 'the same color yours actually is' is primarily a device that permits us to set forth some claim about sameness of color when in ignorance of the color of your car. Knowing that your car is teal, I could just say straightforwardly that my car, too, could have been teal. I just chose midnight blue from the manufacturer's palette, instead. When in ignorance of the color of your new car, but knowing that regarding automotive materials, there is no reason why my Honda could not have been painted the same color as your Audi, regardless of what color yours happens to be, I resort to linguistic devices such as the thorny one Melia cites. Superficially, 'actually' appears to be pointing us to one world amongst many (this one) but with no less plausibility; it is a device of conversational pragmatics that permit us to accurately convey that though your car may be special in so many ways, it is not so special regarding its color. All we need is a way to distinguish the claim that my car could be the same color as yours, i.e., they could be both red, or both green, etc., from the claim that my car could be the same color as yours, i.e., my car could be *that* color.

This casting of modalism as an issue of expressive power, which traces back to Lewis (1986: 13–17), might prove to be misguided, anyway. Suppose that a language quantifying over Lewisian spatio-temporal wholes really does afford expressiveness that boxes, diamonds, and actuality operators—regardless of how indexed—cannot. That is interesting, but it is either a fact of platonic complexities or it is a fact of our own creativity. Nothing follows obviously from either regarding whether there is a plurality of worlds within which every possibility is instanced. Nor does it follow that the plurality does not instance what it should not, i.e., genuine, real impossibilities. Even more irrelevant seems to be any conclusion about the dependence of our grasp of formal operator treatments of ordinary modal claims on prior grasp of quantificational

treatments. That would be a fact about us, our psychologies, and their attending linguistic competencies. However much philosophical interest there might be in these, it has no obvious bearing on the nature of modal reality. In domains such as mathematics, we reconcile ourselves to the fact that some are able to grasp quite advanced ideas straight away, while others must be cajoled with analogies and many examples before, at long last, things finally "click". Platonic reality, however, just is what it is, regardless of the means required for any to grasp it. Similarly, for the anti-platonist alternatives regarding mathematics. The linguistic turn is not a helpful turn, if it is to lead us to a correct account of reality.

Some have complained that there is a circularity in the Lewisian framework (Lycan 1991; Shalkowski 1994; Divers 1997 being somewhat more cautious). That complaint could be framed as a problem regarding the analysis of 'possibility', a project in linguistic or conceptual analysis, but it need not be. For Lewis's plurality of worlds to be the substance of modal reality and that which renders our modal claims true or false, it must avoid both the problem of under-generation and over-generation. Lewis, of course, is committed to no generation of worlds, unlike some versions of modern multiverse physical theories. Lewis's worlds just are. The concerns about under- and over-generation come to this. If the worlds are to be modal reality, there must be sufficiently many to account for all contingency. Otherwise, too many things will come out true in all worlds, some possibilities being unaccounted for by the plurality. There must also not be too many worlds. The concern is not redundancy of worlds, though the grounds for thinking that the plurality contains no such redundancy are not obvious. The concern is rather that the plurality contains too many worlds by virtue of containing *impossible* worlds. If there is a plurality of worlds, what makes it the case that there are "Goldilocks" many worlds, i.e., just the right number? Is it not, so this version of the circularity complaint goes, that even if there is a plurality of worlds, they must satisfy the modal constraints that each is possible and that the plurality lacks no possibility? By constituting the constraints on an appropriate plurality, possibility must "already" be "in place".

When using talk of constraints, this is plausible. Defenders of the Lewis program, however, should resist such talk. There are no metaphysical boundaries within which a world generator must confine its activities. Indeed, if the metaphysical version of the circularity problem concedes that the plurality lacks no possibility and contains no impossibility (Shalkowski 1994), then conceded are the key adequacy conditions for a reduction of modality in terms of the plurality (Sider 2003). If upon inspection the plurality omits no possibility and includes no impossibility, where is the failing in the possible worlds reductive project? 'Upon inspection', however, are the problematic words.

Exposed, now, is a key concern that was either overlooked or conflated with questions of acquiring or analyzing modal vocabulary. Lewis proposed a reductive project that famously produced incredulous stares when others first encountered the project (Lewis 1986: 133ff). His ontology was deemed to be "extreme" or "extravagant". Because hardly any engaged in realistic metaphysics had already embraced an ontology of multiple spatio-temporal wholes, Lewis proposed an account of modality in terms few took seriously. Lewis's early argument from paraphrase (Lewis 1973: ch. 4) gained traction only to the degree that one did not take the paraphrase of 'Possibly *P*' as 'There is a possible world at which *P*' to involve one with multiple spatio-temporal wholes. To the degree that it did, the stares ensued.

When engaged in conceptual/linguistic analysis, the analysis might bring together what is already familiar, say 'bachelor', 'unmarried', and 'male human', or 'cause' and 'constant conjunction'. To the extent that warranted belief in unfamiliar ontology was to be the fruit of Lewis's efforts, he needed one to accept (amongst other things, to be fair) the reduction. Strictly speaking, he required that one accept conditionals like: if the plurality of worlds exists, then modality (amongst other things) can be accounted for in its terms.

Once the ontologically committing paraphrase is rejected because it seems unfaithful to the way many wield their modal vocabulary, the possible worlds project regarding modality became one of convincing others that there is, indeed, ontology to do the philosophical work to which Lewis wanted to put it. How could Lewis make the case that the plurality exists and that it neither under- nor over-generates worlds? The overarching strategy was an inference to the best explanation (Lewis 1986: 3–4). Ineliminable, however, is that the argumentative strategy is both conditional (as stated earlier) and comparative. The allegation is that when comparing the corresponding conditional for other accounts of modality—particularly when that account is set within a wider philosophical context—modal realism does more things better than do competitors.

Set aside the predicament that such philosophical disputes are plagued with standards of better and worse that tend to be in the eyes of their beholders, though that is not insignificant. Why should metaphysicians, regardless of their specific projects, think that reality is "best" along any parameters, much less along their own preferred parameters? It is what it is. In our context, the best that advocates of the possible worlds reduction of modality can hope for is that if there is no plurality of concrete worlds that neither under- nor over-generates possibilities, there is no account of modality in other terms.

That is not nothing, but one wholly convinced of that, as many a modalist might be, can infer the Lewisian ontology only when the *modus tollens* is completed with the minor premise that there is an account of modality in other terms. Modalists are entitled to think that many philosophical advances were made in the spilling of much ink over the various possible worlds accounts of this and that, including of modality. Yet, they are entitled, also, to think that the key step needed for the move away from modalism has yet to be supplied. If "metaphysical circularity" is not afoot, the central question has still been begged.

To be fair, Lewis's project has a resource—one that no party rejected—to parry the charge that the plurality risked over-generating worlds, namely that anything real is possible. In metalinguistic terms, anything true is thereby possibly true. If the plurality contains a world, then it (and, consequently, anything contained therein) is possible. Lewis's argumentative strategy, however, contains the seeds of the destruction of this way of partially undercutting the circularity concern.

Prosecuting the defense of conditional claims as part of an inference to the best explanation of ontology has led some to note an uncomfortable consequence of Lewis's own account of counterfactual conditionals, i.e., all with necessarily false antecedents are true. They do not all seem to be true, though. It is false that were I to square the circle, the moon would be made of green cheese.

An increasingly common way of accounting for the differences between these counterfactual claims is to posit the existence of impossible worlds and/or impossible objects, as detailed in Daniel Nolan's contribution to this *Handbook*. This maneuver, if put forth in the same spirit as exhibited by Lewis when he enjoined us to embrace his plurality of worlds, warrants the similar embrace of impossibilia. If the argumentative strategy warranted realism about possibilia, then it warrants realism about impossibilia. Not only does this undermine faith that the ontology does not over-generate worlds, it guarantees that it does.

Drawing a line in the philosophical sand by insisting that the merit of the strategy extends to possibilia and no further should now, itself, draw incredulous stares. If one has assumed that there are possibilia alongside actualia, as it were, but no impossibilia, then either an argumentative question has been begged or else the account is circular in the way suggested earlier, since the totality of the items to do the reduction of modality must be all and only the *possible* worlds. When the strategy warranted merely an existence claim about a more plenitudinous totality, one could still maintain that existence entails possibility. No longer, at least until modal reductionists can supply reasons for the relevant line in the philosophical sand based on a principle deeper than "to possibility, but not beyond". This problem is not confined to modal metaphysics, but to

all projects extending to realist embraces of ontology that do not permit some independent verification beyond the variation of an argument from theoretical virtues (Bueno and Shalkowski 2020). In this respect, the possible worlds reductive project is in good company, as it were.

References

Ayer, A. (1946) *Language, Truth and Logic*, London: Gollancz.

Benacerraf, P. (1965) "What Numbers Could Not Be," *The Philosophical Review*, 74: 47–73.

Benacerraf, P. (1973) "Mathematical Truth," *The Journal of Philosophy*, 70: 661–679.

Bueno, O. and Shalkowski, S. (2009) "Modalism and Logical Pluralism," *Mind*, 118: 295–321.

——— (2013) "Logical Constants: A Modalist Approach," *Noûs*, 47: 1–24.

——— (2015) "Modalism and Theoretical Virtues: Toward an Epistemology of Modality," *Philosophical Studies*, 172: 671–689.

——— (2020) "Troubles with Theoretical Virtues: Resisting Theoretical Utility Arguments in Metaphysics," *Philosophy and Phenomenological Research*, 101: 456–469.

Carnap, R. (1954) *The Logical Syntax of Language*, London: Routledge & Kegan Paul Ltd.

Carnap, R. (1956) *Meaning and Necessity: A Study in Semnatics and Modal Logic*, Chicago: University of Chicago Press.

Chalmers, D. (2002) "Does Conceivability Entail Possibility?" in T. Gendler and J. Hawthorne (eds.), *Conceivability and Possibility*, Oxford: Oxford University Press.

Divers, J. (1997) "The Analysis of Possibility and the Possibility of Analysis," *Proceedings of the Aristotelian Society*, 97: 141–160.

——— (2002) *Possible Worlds*, London: Routledge.

Fine, K. (1980) "First-Order Modal Theories," *Studia Logica*, 39: 159–202.

———. (1981) "First-Order Modal Theories I—Sets," *Noûs*, 15: 177–205.

Fine, K. (1982) "First-Order Modal Theories III: Facts," *Synthese*, 53: 43–122.

Forbes, G. (1985) *The Metaphysics of Modality*, Oxford: Clarendon Press.

——— (1989) *Languages of Possibility*, London: Basil Blackwell.

Forbes, G. (1992) "Melia on Modalism," *Philosophical Studies*, 68: 57–63.

Lewis, D. (1968) "Counterpart Theory and Quantified Modal Logic," *The Journal of Philosophy*, 65: 113–126.

——— (1973) *Counterfactuals*, Cambridge, MA: Harvard University Press.

——— (1986) *On the Plurality of Worlds*, Oxford: Blackwell.

Lycan, W. (1991) "Two—No, Three—Conceptions of Possible Worlds," *Proceedings of The Aristotelian Society*, 91: 215–227.

Melia, J. (1992) "Against Modalism," *Philosophical Studies*, 68: 35–56.

Melia, J. (2003) *Modality*, Chesham: Acumen.

Pap, A. (1958) *Semantics and Necessary Truth*, New Haven: Yale University Press.

Prior, A. and Fine, K. (1977) *Worlds, Times and Selves*, London: Duckworth.

Quine, W. (1936) "Truth by Convention," in *Philosophical Essays for Alfred North Whitehead*, New York: Longmans, Green and Co., 90–124. Reprinted in his (1976) *The Ways of Paradox*, Cambridge, MA: Harvard University Press, 77–106.

——— (1951) "Two Dogmas of Empiricism," *The Philosophical Review*, 60: 20–43.

——— (1963) "Carnap and Logical Truth," in Schilpp, P. (ed.), *The Philosophy of Rudolf Carnap*, La Salle, IL: Open Court Publishing Co., 385–406. Reprinted in his (1976) *The Ways of Paradox*, Cambridge, MA: Harvard University Press, 107–132.

Shalkowski, S. (1994) "The Ontological Ground of the Alethic Modality," *The Philosophical Review*, 103: 669–688.

Shalkowski, S. (2004) "Logic and Absolute Necessity," *The Journal of Philosophy*, 101: 55–82.

Sidelle, A. (1989) *Necessity, Essence, and Individuation: A Defense of Conventionalism*, Ithaca: Cornell University Press.

Sider, T. (2003) "Reductive Theories of Modality," in Loux, M. and Zimmerman, D. (eds.), *The Oxford Handbook of Metaphysics* Oxford: Oxford University Press.

——— (2011) *Writing the Book of the World*, Oxford: Clarendon Press.

Yablo, S. (1993) "Is Conceivability a Guide to Possibility?" *Philosophy and Phenomenological Research*, 53: 1–42.

PART 3

Modal anti-realism


Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Chapter 11
MODAL ANTI-REALISM

John Divers

11.1 Motivation

For the purposes of this chapter, I will supply modal anti-realist positions with only a very general and broad motivation. Other things equal, modal anti-realist positions – like anti-realist positions in general – are (intended to be) metaphysically less committed, or more economical, than their realist competitors. Such metaphysical economy is usually held to make for epistemological economy as well. The thought is that the less substantial and less independent of us the modal "facts" are, the less we demand – ceteris paribus – by way of intellectual capacities that are able to reach them. Whatever other complications (semantic, etc.) attach to the anti-realist positions, their proponents will typically contend that they are delivering a satisfactory explanation of our modalizing that is overall of better value than that offered by the realist opponent. The realist opponent will typically disagree, and usually on the grounds that the favourable "costs" of the anti-realist position are moot, since the "benefits" they offer are, in fact, inadequate attempts at explanation. Hereafter my concern will be with the taxonomy of modal anti-realist positions, and not with the specific benefits and costs that are alleged to attach to them.

11.2 The basics

"Modal anti-realism" refers to different views in different contexts. That is because "modal realism" refers to different views in different contexts, and "modal anti-realism" is a label for any view that involves refusal to accept what is called "modal realism" in that context. The primary distinction between kinds of modal realism is between: (i) those committed to modalities being real (by which I mean, irreducible and, in some sense, objective/mind-independent) and (ii) those committed to the existence of modality-relevant entities of some sort. I call the former *ideological modal realism* (IMR) and the latter *ontological modal realism* (OMR). Aristotle (1998) is an exemplary proponent of IMR because he believed that the status of attributes as essential or accidental to a substance are fundamental features of reality. Aristotle is not obviously a proponent of any version of OMR. Hume (1978: sect. VII) is an exemplary opponent of IMR. For Hume thought that the very purpose of our modal concepts makes them unfit to copy anything real: rather, he thought, our modal concepts are for organizing and coping with a world that is, in reality, non-modal. Lewis (1986) is an exemplary proponent of OMR, for he believed in the

existence of an infinite pluriverse of non-actual possible worlds and possible individuals: entities no less real than the universe and the individuals (such as ourselves) in it that we call "actual". Where the modal realism is thus Lewisian, almost all philosophers are, in that context, modal anti-realists. Despite being an exemplary proponent of OMR, Lewis joins Hume in opposing IMR, although Lewis does not accept Hume's supporting narrative about the function of our modal concepts. Hereafter, I shall classify modal anti-realist positions in relation to various species of OMR or of IMR that we may discern.

11.3 Some methodological issues[1]

Firstly, the present coverage of modal anti-realism is intended to be inclusive and descriptive. I am not interested in the precious questions of which, if any, of the positions described is *really* deserving of the name "modal realism" (thence, "modal anti-realism") or cuts to the heart of the matter of what modal realism (thence, modal anti-realism) is *really* about.

Secondly, any discussion of options in a debate involving would-be realists and anti-realists is in need of a policy about where to place those who might naturally or plausibly be characterized as neutral or agnostic. Where we have, in a context, a version of modal realism, one way of departing from it is by rejecting it, but another is by refusing to accept or reject. Again, there is a precious kind of question afoot in which I have no interest – that is: whether those who wish to occupy the neutral position, or wish to be neutral between neutrality or opposition etc. are *really* anti-realists rather than non-realists. My policy is to opt for two categories and use the terminology accordingly: once a version of modal realism is identified, anyone who (considers and) does not accept the thesis in question is classified, in that context, as a modal anti-realist. This matter introduces another category that is of genuine significance.

Thirdly, OMR and IMR are metaphysical theses. Accordingly, a massive question that lies behind our issue is that of the status of metaphysics. I believe that it is important to outline the methodological subtleties involved, and especially as they apply to this particular topic. However, I signal here that I will exit by lumping together again as modal anti-realists a range of thinkers who have quite different attitudes towards the status of metaphysics.

As I am thinking of modal realists, of any sort, they think of themselves as making a given metaphysical commitment and present themselves accordingly. A variety of non-positive reactions to such a commitment might ensue, and at the extreme ends of that range are these. At one end, we have the non-positive reaction that is most engaged with the (modal) realist metaphysician: it is perfectly accepting of metaphysics but refuses the particular metaphysical commitment that is on offer. At the other end, the most radical of non-positive reactions is refusal to debate any such metaphysical contention because metaphysics in general is not to be entertained. The latter case makes for dialectical complexities. For such a radical (modal) anti-realist will often take herself to be offering a theory of modal thought that is non-metaphysical, and the better for it. And that may be so while those who are engaged in the metaphysical project will classify the position taken by such a philosopher as one that is metaphysically anti-realistic when considered within the frame of a metaphysical debate. This subtle methodological issue looms large over the history of anti-realism about modality in particular. For that history is interwoven with the history of empiricism, and the empiricist tradition is shot through with scepticism about the classical metaphysical project. In all, then, modal anti-realists may come from one of three meta-metaphysical positions: there are straightforward metaphysicians who refuse a particular metaphysical thesis (Lewis 1986); there are rejectors of metaphysics who refuse to make anything of a thesis that is characterized in self-consciously metaphysical terms (Ayer 1936) and – between these extremes, we must also note – there are the explicators of metaphysics who are prepared

to engage with theses that are presented as metaphysical, but only insofar as they can map these onto something considered properly intelligible and tractable, perhaps in semantic, epistemology or empirical science (Quine 1948; and according to Divers 2017b, Quine more generally). In a simplifying move, I will write hereafter as though we always have a straightforward dispute between philosophers who are engaged in the classical metaphysical project. But the limitations of that conceit have been advertised.

Fourthly, since modalities are of various kinds there is space for views that are metaphysically heterogeneous, in being realistic about some modalities and anti-realistic about others. Here, in another simplifying move, I will restrict attention to the modalities that are the strongest candidates for being "objective". Accordingly, and for example, I will not consider modalities of cognitive attitude (what must be so, given what an agent knows or believes) or modalities of obligation (what must be done, according to some legal or moral code). I will consider the family of *prima facie* objective mind-independent modalities that includes the causal, the nomological and the physical. Indeed, by default, I will be considering the "head" of that family, which is to say metaphysical modality, as per Kripke (1972). Metaphysical possibility will be considered as the most inclusive, easiest to come by, kind of such possibility: what is possible causally, etc., is possible metaphysically but not always vice versa. Metaphysical necessity will be considered as the least inclusive, hardest to come by, kind of such necessity: what is metaphysically necessary is nomologically necessary, but not always vice versa. This sweep is not intended to be comprehensive in speaking to all of the modalities. It is also intended to leave open further important questions about the basis on which a modality deserves to be called "causal", etc., and about how and when different descriptors ("metaphysical" modality and "logical" modality, for example) are equivalent.

Fifthly, there is a time and a place for the question of what the criteria are for being modal. In material mode, what makes a feature of the world modal: in formal mode, what makes a word or a concept modal? But that question is not for this time and place. Here, it is presumed, we know the modal when we see it: or rather, when it is called "necessity", "possibility", "contingency", "potentiality", "essentiality", etc.

Many issues have been raised in this section that the dedicated student of modal philosophy may wish to pursue further. But for present purposes, they will (largely) be left behind to allow quicker coverage of the ground.

11.4 Ontological modal anti-realism

The primary thesis of OMR is the thesis that possibilia exist, also known as *possibilism*. These are sometimes called "mere" possibilia, indicating that we wouldn't, pre-theoretically, think of them as things that actually exist. If sperm-cell *s* and ovum *e* had fused, they would have formed a human embryo *h*; but they did not fuse and so, while *h* might have existed it does not (actually) exist. Similarly, blue lemons do not (actually) exist, although some might have. Or so it seems. The possibilist move at this point is to defend, in some version, the thesis that *h* does exist, or that blue lemons do exist, once we consider *all* the things there are. The typical possibilist generalizes to hold that for any one thing that might have been, such a thing is (exists). The possibilist move, historically associated with Meinong (1960) comes in two main varieties. Firstly, we may retain our views about what *actually* exists, so not *h*, for example, but position what is actual or what actually exists as only a small part of what unrestrictedly exists. This move is often compared to expanding our concept of what unrestrictedly exists beyond our concept of what exists at the present time. This is the possibilism of Lewis (1986), in which blue lemons exist but do not exist in the part of reality that we call "actual". Secondly, we may retain our commitment

that there is no conceptual or logical distinction between what unrestrictedly exists and what actually exists, but expand our beliefs about what actually and unrestrictedly exists. This is the possibilism of Linsky and Zalta (1994, 1996) and of Williamson (1998, 2002, 2013) according to which the mere possibilia are esoteric, contingently non-concrete, posits that actually exist but which we don't recognize pre-theoretically. The dominant motive for accepting possibilism has been the desire to have access to theories that fall under the heading of "possible-worlds semantics". For it was theories of this kind that promised first to deliver a philosophically satisfactory interpretation of quantified modal logics (along with a meta-theory) and of *de re* modal discourses in general. The most straightforward way of appropriating all such theories is to take literally their apparent quantification over a domain of all possibilia. Among the earliest versions of possible-worlds semantics, and the version that is still the best known and frequently discussed by philosophers, is that presented in Kripke (1959, 1963). It is worth noting at this stage, however, that Lewis is doubly distanced from this standard motivational picture. For Lewis (1968, 1986) is not an advocate of interpreting modal discourse by appealing to quantified modal logic, and the case presented by Lewis (1986) for believing in possibilia extends significantly beyond their semantic utility.

The modal anti-realist, in this context, will be one who does not accept the existence of whatever possibilia are postulated. This, however, gives rise to a tricky case of classification. The orthodox position in the heyday of possible-worlds semantics was that variously described as that of the actualist, moderate or ersatz realists. At the ontological heart of such positions lay the idea that there were things we (philosophers) ought to believe in already and which can, at least, "play the role" of the possible *worlds* postulated in elaborations of the Kripkean semantic theories. So there exist(s) not just one actualized possible world but also many non-actualized possible worlds. The candidates to play the role were various maximal combinations of sentences, propositions, properties or states of affairs where these had to be taken to outstrip the true propositions, the instantiated properties or the actualized states of affairs (respectively). Thus, see Plantinga (1974) and Stalnaker (1976). This project was extended to having candidates to play the role of those possibilia that are not supposed to be worlds – things like *h* or the blue lemons. Despite normally being called "realists" in this respect, because they are appealing to the existence of entities of some kind or other, the proponents of these views might be found claiming (variously): that they believe in and are realists about possible worlds, that they have surrogates for possible worlds or that they have dispensed with or eliminated possible worlds, having shown that something else can play the role. Again, this is a point at which there is no profit in forcing the question of who really deserves to be called an X – in this case "a possibilist". These ersatzists, as we might generically call them, are, however, to be distinguished from another breed of philosophers who are entirely conscious and thorough in their determination to avoid taking on ontological commitments of the kind that the ersatzists accept. Philosophers of this kind tend to say that they do not believe that there are things – neither possibilia nor surrogates for them – that correspond to the variables of the semantic theory. Thus, for example, we have the agnosticism described in Divers (2004, 2006). More typically, such views aspire to a sophistication that allows the following combination of claims: (i) we can truly believe that there are such things as possible worlds but (ii) also maintain our ontological anti-realism (because we have in stock a demonstration that quantification over them is ultimately dispensable). Thus, for example, we have the *modalism* of Fine (1977, 2003) and Forbes (1985, 1989) in which the fundamental theory contains modal operators, and the *fictionalism* of Rosen (1990) in which the fundamental theory contains a fictive prefix (of the sort, "According to the realist theory . . .").

In increasingly popular extensions of possibilist ontology, many more expansive ontological realists are happy to add impossibilia to their ontological commitments – impossible worlds

and/or impossible individuals. This extension also tends to be semantically motivated, and in particular by the desire to offer more discriminating and finer-grained truth-conditions for various modal expressions than would be allowed if they are considered equivalent for being true at the same possible worlds. One massive issue arising from consideration of the admission of impossibilia is the extent to which this commits one to countenancing contradiction. For these issues, see Berto (2013).

Finally, in classifying versions of OMR, we should note that this can take the form of ontological commitment to modally relevant entities that are not of the category of objects – for example: the modal properties endorsed by Plantinga (1974) or the anti-Humean necessitation relations discussed by Armstrong (1983). Local versions of ontological anti-realism are available accordingly.

11.5 Ideological modal realism: Caveat emptor!

The proponent of IMR is committed to modalities being real (by which I mean, irreducible and, in some sense, objective/mind-independent). And to make for clear contrast with OMR, we shall now say, this metaphysical status does not derive from the existence of any modally relevant entities. The most familiar straightforward version of such a thesis would have it that there are real modal ways of things being or modal aspects of reality. Further, that this modal reality is intimated by the use of one of the modal *operators* of standard modal logics or their natural language analogues. In the mouths of those who do not believe that operators refer to entities (and the semantic category of operators is usually introduced for that very reason), this is an ideological rather than an ontological claim. To exemplify the *de dicto* case, one modal way for things to be is that *it is contingent that* there is an absolute maximum velocity. To exemplify the *de re* case, another modal way for things to be is that some things are *necessarily* human. By the conventions in place up until the very late twentieth century (and shared in the literature of Marcus, Quine, Kripke and Lewis), we could take these modal terms to be interchangeable with terms from Aristotelian essentialist metaphysics – thus: "It is *accidental* that there is an absolute maximum velocity " and "Some things are *essentially* human." That consensus no longer holds (following Fine 1994). But I intend ideological *modal* realism (IMR) inclusively, to extend to realism about essentiality (etc.). Before we can look profitably at the varieties of IMR, some serious potential misunderstandings must be forestalled.

The first potential misunderstanding is that taking modal terms to express irreducibly modal concepts is the same as (or entails) taking them to intimate metaphysical (ideological) primitives. This confusion is promoted by inattention to decidedly different uses of certain philosophical terms of art. In the heyday of possible-worlds semantics, when philosophers were concerned with "truth-conditions", it was, almost invariably, because they sought a theory of meaning. Applying the general thought to the modal case, the idea was that if truth-conditions for modal sentences could be generated and presented in an appropriate way, and in the context of an appropriate theory, this would illuminate the meanings of modal terms or contribute to the analysis of modal concepts. Philosophers of that era considered whether truth-conditional semantic theories for modal discourses had commitments of either of two kinds: ontological commitments to the existence of possibilia (as discussed earlier) and commitments to any irreducible (undefined) modal *concept*. Crucially, however, and in the absence of a great deal of further argument, the question of whether a theory is committed to irreducible modal concepts is removed from the question of whether the significance of any expression is to intimate metaphysically irreducible (fundamental) modality. Indeed, it merits emphasis that the opponents of IMR – excepting Lewis and his followers – are happy to take modal *concepts* as irreducible

(primitive, undefinable). Thus see, for example, Blackburn (1984, 1986), more of whom anon. But the distinction between irreducible concepts and irreducible aspects of reality was, I fear, blurred amidst an important change in the culture. Once the possible-worlds semantic theories, and the philosophical articulations of these theories, were out there, analytic philosophy underwent a metaphysical turn. So then a whole generation of philosophers, I surmise, presumed that their predecessors' concerns with "analysis" and "truth-conditions" and "primitives" all had to be metaphysical. This tendency was encouraged because the old debates about the semantics of modal discourse tended to be described by the parties involved as debates in the metaphysics of modality; but really, they were debates about the ontology and primitive concepts of semantic theories (as these applied to metaphysically neutral modal and essentialist discourse, more of which to come). The continuing confusion between questions about modal concepts and those about modal ideology (metaphysics) is the unfortunate legacy of that transitional period between the dominance of the theory of meaning and that of metaphysics.[2] In a similar vein, the non-inevitability of the hegemony of metaphysics has to be taken into account when we attempt to understand philosophical work that was done (just) before the heyday of possible-worlds semantics. It is clear that the likes of Marcus (1993) and Prior (1957) were staunch defenders of quantified modal logics and of our right to take the modal operators as logical primitives. But it is – to say the least – unclear whether these philosophers always intended the view they defended to be accompanied by a corresponding metaphysical commitment to primitively modal aspects of reality.

The second potential misunderstanding is that by acknowledging a metaphysical modality and, or, modality *de re*, one is committing to ideological modal *realism*.

It is certainly not *only* ideological modal realists who claim the privilege of using modal claims of these sorts and of acknowledging their truth. Granted, there are forms of ideological anti-realism, in particular older forms of expressivism, or non-cognitivism (see Section 11.6) that do concede truth (or *real* truth) to a local realist opponent. But I trust that the case is now well made that not all forms of ideological anti-realism need do so – or, at least, that we do not improve our understanding of the issues if we immediately foreclose the right of ideological anti-realists to make the case for their entitlement to acknowledge truth (cf Wright 1992). If we were discussing anti-realism about colour or mathematics or morals, this point would, perhaps, not merit emphasis. But there are features of our case – or at least the surrounding terminology – that are especially pernicious in this regard and which need to be identified as such. Firstly, I have stipulated that we are (by default) focusing on *metaphysical* necessity. From there it is easy to drift towards speaking of a conception of necessity as metaphysical, thenceforth to necessity that is in reality and so to a pseudo-equivalence between a metaphysically neutral categorization of a certain conception of necessity as metaphysical and a metaphysically substantive (realist) conception of necessity. What Kripke (1972) compelled us all to think about is a kind of necessity that is not strictly correlated with the analytic or the a priori, which might be conceived as the strongest form of physical necessity, which may nonetheless be absolute and which is implicated in plausible and naturally made essentialist claims. This description, I contend, is neutral with respect to any interesting question of realism: it does not rule out an ideologically anti-realistic theory of such a necessity. If Kripke had chosen an adjective other than "metaphysical", much presumption of realism might have been avoided, but we persevere with the terminology. Secondly, further to there being a necessity that is metaphysical and in good standing, Kripke (1972) also offers support for there being necessity that is *de re* and in good standing. But it is tendentious to understand one who takes that point as being committed to the ancient, medieval and metaphysical (realism-relevant) connotation of some modality being "of things" (language-independent) things and some modality being "of what is said" (language-dependent).

For, as with the case of metaphysical necessity, there is a perfectly decent way of taking Kripke's point here without committing to any form of modal realism. All we need do is appeal to the standard syntactic characterization(s) of modality *de re* (and of modality *de dicto*) as follows. In quantified modal logics, the *de re* modal construction is quantification across a modal operator into a variable ($\exists x \Box Hx$). And, if we wish, we can extend the term "*de re*" to apply to the occurrence of a name in a sentence (such as $\Box Ha$) when we can infer by existential introduction on that name any sentence such as the former. Once enough syntactic conditions on the *de re* have been added, let any other occurrences of the modal operator be *de dicto*. In English, the *de re* modal construction is in evidence when, for example, we have a modal modifier attaching directly to a non-modal adjective to form a complex modal adjective – thus: "Socrates is necessarily human" or "Hesperus is necessarily identical to Phosphorus". Now it is not, of course, to be ruled out that certain versions of ideological modal anti-realism will have more trouble with these *de re* constructions than they do with the *de dicto*. But in order to start at a decent level of debate, all parties do well to recognize that there is a good *prima facie* case for holding true various claims that are taken to intimate necessity de re and, or metaphysical necessity.

A third potential misunderstanding arises from characterizations of Kripke (1972) as having championed or rehabilitated modal metaphysics or the metaphysics of modality. In fact, what Kripke does is twofold. In the first place, he makes the case for an intuitively compelling and workable notion of a necessity that is absolute, without being equivalent to the narrowly logical, epistemological or semantic (so call it "metaphysical"). In the second place, he puts that modal notion to work in doing some metaphysics *by* modality, as when arguing against the mind-brain identity thesis (ibid, Lecture III). It is also to be acknowledged that Kripke practices a breed of "modal metaphysics" that involves puzzling over which controversial modal and/or essentialist claims we should hold true: claims of essentiality or necessity of origin, of kind-membership and of identity. What Kripke does not (obviously) do is to advocate a metaphysics *of* modality in the sense that is presently salient for us. As far as I can discern, Kripke (1972) engages with no question of IMR at all. If an analogy is wanted, the "modal metaphysics" of Kripke (1972), especially when it involves argument about the essentiality of origin, etc., is analogous to the practice of normative ethics, while advocacy of IMR would be analogous to taking a (realist) position in meta-ethics.

11.6 Ideological modal anti-realism proper

Following the metaphysical turn, we can identify genuinely metaphysically modalist philosophers who explicitly intend their use of modal expressions to express a commitment to metaphysically irreducible modality. The clearest examples I can find of such IMR proper are in Plantinga (1987), Shalkowski (1994), Williamson (2013) and Hale (2013).[3, 4] The family of ideological modal anti-realists is unified by their, metaphysically Humean, refusal to accept the presence in reality of irreducible modal aspects. The differences within that family are to be traced in terms of the supplementary story that is told about how we are to explain modal truth, if not on this simplest of models that connects modal words to a fundamental modal reality. Firstly, however, we must consider Quine.

An often reproduced version of Quine has him saying something very radically anti-realistic about modality that goes far beyond the straightforward rejection of IMR and begins with rejection of modal truth. Further, the infamous "Quinean modal scepticism" has it that modal statements, and especially those involving modality *de re*, are not just systematically false, but irredeemably defective, confused, unintelligible, etc. Burgess (2008) exposes this version of

Quine as mythical, and the theme is developed in Divers (2017a). Tentative remarks about what a positive Quinean story about modal truth might look like are to be found at the end of Divers (2017b).

Despite Lewis's explicit warning about his own non-metaphysical usage of the term "analysis" (Lewis 1992, 215), many seem to have failed to distinguish Lewis's claims about the analysis of modal concepts from his metaphysical claims (see Divers & Fletcher 2018). Nonetheless, the beginning of the properly metaphysical turn in modal philosophy might be associated with the work of Lewis, for whom there is no semantically irreducible modality, nor any metaphysically irreducible modality. The focus of the early concerns of Lewis (1968) semantics, and the semantic analysability, of (*de re*) modal discourse. These concerns gradually expanded to include an increasingly explicit concern with the metaphysical reducibility of modality, and the endorsement of IMR is clear in Lewis (1986). It is against the spirit of Lewisian philosophy to join with the orthodox opponents of IMR: for they, but not he, will foreground a story about why modality matters to us and seek a non-objectivist explanation of modal truth. For Lewis, all that need concern us philosophers is which body of sentences we hold true before doing philosophy: these are our data. And the theory that best systematizes all that we take to be true is one in which modal judgements are not treated as anything other than descriptions of objective fact – albeit, ultimately non-modal fact about an infinite and plenitudinous pluriverse.

Having considered the two prominent ideological modal anti-realists, Quine and Lewis, in their own right, the remaining ground is best covered by the consideration of two traditions. What these traditions have in common is the idea that modal truth is to be explained in terms of limits that are set by the meanings of words, or by minds. One of these traditions is reductive: modal truth is to be explained on the correspondence model, it obtains in virtue of (non-modal) facts about (the limits of) meanings or minds. The other tradition is non-reductive: modal truth is not to be explained on the correspondence model; it is to be explained (somehow) in terms of the limits of meanings or minds, but not in *that* way. With such reductionism, once again, when we require clear metaphysical intent of the part of proponents, it is not obvious that there have been many ideological modal anti-realists of the kind intimated: those who held the metaphysical reducibility of modal facts to facts about (the limits of) meanings or minds. Carnap (1947), if read as though he was with the metaphysical project, might be one: but that (I hardly need add) is a big "if". A more promising candidate may be Sidelle (1989) who might be read as claiming: (a) that metaphysical necessity is metaphysically reducible to analyticity and (b) when we explain analyticity as truth in virtue of meaning, a full-blooded metaphysical reading of "in virtue of" is appropriate. This last consideration is crucial to establishing a reductive version of opposition to IMR.

Many logical empiricist philosophers, including the champion of logical positivism, Ayer (1936), can be found saying that necessary truths are those that hold in virtue of the meanings of words and the conventions that govern their use. One problem that arises when usage of that kind is careless is that it does not distinguish clearly which truth is at issue. Is it a claim of the type *it is necessary that P*, or is it a non-modal claim *P*, which is (held to be) necessary? Our present concern is resolutely with the former. But having got past that issue, we must take into account that a claim to the effect that such a truth "obtains in virtue of . . ." need not have meant the same in the mouths of those who did not work within a metaphysical culture than it does in the mouths of those who do. There is abundant evidence that Ayer, for example, had in mind a great deal of the time a kind of explanation of necessary truth that is clearly aligned with what I have recently classified as the non-reductive tradition of ideological modal realism (see Thomasson 2007). My view, with Thomasson (2007), is that the metaphysically reductive

understanding of "true in virtue of . . .", has delineated and made salient a disastrously weak version of the otherwise attractive thought that modal truth is to be explained in terms of (the limits of) meanings or minds. And so, finally, I turn to the non-reductionist versions of ideological modal realism.

This non-reductive tradition of ideological modal realism encompasses positions that are variously known as *non-cognitivist*, expressivist, *non-descriptivist*, or *quasi-realist*. The first and most important steps towards framing a position of that kind which is apt to avoid the well-taken criticisms of its precursors is, in my view, taken in the modal quasi-realist manifesto of Blackburn (1986) (see also Blackburn 1984, ch. 6; Blackburn 1993). The modal quasi-realist narrative, as I would appropriate it and endorse it, has the following profile. Firstly, it refuses IMR and so aspires to explain our modalizing practices without appeal to modal aspects of reality. Secondly, it is non-cognitivist in spirit because it embraces the Humean idea that our modal locutions are born of the need to cope with, or organize, a reality that is non-modal: not the need to copy a reality that is modal. Thirdly, it is minimalist and ecumenical about truth, allowing truth to spread all across our language to the parts of it, the modal being a case in point, in which we are exercising our attempts to cope alongside our attempts to copy. Thus the conceit of a universal and substantial correspondence theory of truth is abjured. Fourthly, it is guided by naturalism: the dual trigger for our judgements of impossibility is our finding ourselves (in some way) unable to make anything of a certain proposition and subsequently finding no naturalistic explanation of our own limitation. Fifthly, it is not radically conventionalist, skeptical or error-theoretic. It proceeds from the presumption that the locutions of (metaphysical) modality do not betray an arbitrary decision, or a failure or a mistake on our part. Were we to look to proceed from this last presumption to a positive account of the function of our (metaphysical) modalizing, we might, I suggest, seek clues in Wittgenstein (1964) as interpreted by Wright (1980, ch. 20) in thinking of its operating, characteristically, in certain cases where we manifest our need to insist upon and to "regulate the distinction between appearance and reality". Further developments of a modal quasi-realist position may benefit from considering in a certain light the relationship between necessity and counterfactuality. That is by: (a) delineating the actual-world role and value of our counterfactual suppositions, (b) motivating our interest in necessity through its correlations with generalizations about counterfactuals and (c) delineating the metaphysically least demanding conception of necessity that is so formed.

I believe that it is with this program that the most promising future for modal anti-realism lies.

Notes

1 For the reader who is happy to make a quick survey of the modal anti-realist landscape, without forming a sense of what lies beneath and beyond, this section may be better skipped at first pass and read last (if at all).

2 Further to this theme, see the remarks on Lewis later in this chapter.

3 Perhaps there are earlier clear examples of IMR proper. The title of Plantinga (1974) certainly intimates a metaphysics of modality – thus: "The Nature of Necessity". I am not sure that we find there a clear distinction between the intention to take necessity as a metaphysical primitive as opposed to a conceptual primitive. But I am prepared for scholars to show otherwise. Plantinga (1987), however, is clearly with the project of defending ideological modal realism, reacting as he is to a Lewis whose concerns had also become explicitly metaphysical.

4 Hale (2013) discerns a metaphysical hierarchy within the broad class of modalities that are under consideration here. Thus, the facts of metaphysical necessity are not equivalent to the facts of essentiality, the latter being the more fundamental explainers of the former.

References

Aristotle (1998) *The Metaphysics*, Translated with an Introduction by Hugh Lawson-Tancred, London: Penguin Classics.

Armstrong, D. M. (1983) *What Is a Law of Nature?* Cambridge: Cambridge University Press.

Ayer, A. J. (1936) *Language, Truth and Logic*, London: Gollancz.

Berto, F. (2013) "Impossible Worlds," Substantive Revision in *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta. Available at: https://plato.stanford.edu/archives/win2013/entries/impossible-worlds/.

Blackburn, S. (1984) *Spreading the Word*, Oxford: Clarendon Press.

——— (1986) "Morals and Modals," in G. MacDonald and C. Wright (eds.), *Fact, Science and Morality: Essays on A. J. Ayer's Language, Truth and Logic*, Oxford: Blackwell, 119–42.

——— (1993) "Addendum to 'Morals and Modals'," in S. Blackburn (ed.), *Essays in Quasi-Realism*, Oxford: Oxford University Press, 73–4.

Burgess, J. (2008) "Quinusab omni naevo vindicatus," in *Mathematics, Models and Modality*, Cambridge: Cambridge University Press, 203–35.

Carnap, R. (1947) *Meaning and Necessity*, Chicago: University of Chicago Press.

Divers, J. (2004) "Agnosticism about Other Worlds: A New Antirealist Programme in Modality," *Philosophy and Phenomenological Research* 49: 659–84.

——— (2006) "Possible-Worlds Semantics without Possible Worlds: The Agnostic Approach," *Mind* 115: 187–225.

——— (2017a) "De Re Modality in the Late 20th Century: The Prescient Quine," in M. Sinclair (ed.), *The Actual and the Possible: Modality and Metaphysics in Modern Philosophy*, Oxford: Oxford University Press, 217–35.

——— (2017b) "How Skeptical Is Quine's 'Modal Skepticism'?" *Monist* 100: 194–210.

Divers, J. & Fletcher, J. (2018) "(Once Again) Lewis on the Analysis of Modality," *Synthese*, https://doi.org/10.1007/s11229-018-1698-7.

Fine, K. (1977) "Postscript: Prior on *The Construction of Possible Worlds and Instants*," in A. N. Prior & K. Fine (eds.), *Worlds, Times and Selves*, London: Duckworth, 116–61.

——— (1994) "Essence and Modality," *Philosophical Perspectives* 8: 1–16.

——— (2003) "The Problem of Possibilia," in M. Loux & D. Zimmerman (eds.), *The Oxford Handbook of Metaphysics*, Oxford: Oxford University Press, 161–80.

Forbes, G. (1985) *The Metaphysics of Modality*, Oxford: Clarendon.

——— (1989) *The Languages of Possibility*, Oxford: Blackwell.

Hale, B. (2013) *Necessary Beings*, Oxford: Oxford University Press.

Hume, D. (1978) "An Enquiry Concerning Human Understanding," in L. A. Selby-Bigge & P. H. Nidditch (eds.), *Enquiries Concerning Human Understanding and Concerning the Principles of Morals*, Oxford: Clarendon Press.

Kripke, S. A. (1959) "A Completeness Theorem in Modal Logic," *Journal of Symbolic Logic* 24: 1–15.

——— (1963) "Semantical Considerations on Modal Logic," *Acta Philosophica Fennica*, 16: 83–94. Reprinted in L. Linsky (ed.), *Reference and Modality*, Oxford: Oxford University Press (1971): 63–87.

——— (1972) "Naming and Necessity," in D. Davidson and G. Harman (eds.), *Semantics for Natural Languages*, Dordrecht: Reidel, 253–355; 763–9. Reprinted as monograph with new "Preface," Oxford: Blackwell (1980).

Lewis, D. (1968) "Counterpart Theory and Quantified Modal Logic," *Journal of Philosophy* 65: 113–26.

Lewis, D. (1986) *On The Plurality of Worlds*, Oxford: Blackwell.

——— (1992) "Review of D. M. Armstrong, *A Combinatorial Theory of Possibility*," *Australasian Journal of Philosophy* 70: 211–24.

Linsky, B. & Zalta, E. (1994) "In Defense of the Simplest Quantified Modal Logic," *Philosophical Perspectives* 8: 431–58.

——— (1996) "In Defense of the Contingently Nonconcrete," *Philosophical Studies* 84: 283–94.

Marcus, R. (1993) *Modalities: Philosophical Essays*, Oxford: Oxford University Press.

Meinong, A. (1960) "The Theory of Objects," in Roderick M. Chisholm (ed.), *Realism and the Background of Phenomenology*, Glencoe, IL: Free Press.

Plantinga, A. (1974) *The Nature of Necessity*, Oxford: Clarendon.

Plantinga, A. (1987) "Two Concepts of Modality: Modal Realism and Modal Reductionism," *Philosophical Perspectives* 1: 189–231.

Prior, A. N. (1957) *Time and Modality*, Oxford: Oxford University Press.
Prior, A. N. & Fine, K. (1977) *Worlds, Times and Selves*, London: Duckworth.
Quine, W.V. O. (1948) "On What There Is," *Review of Metaphysics* 2: 21–38.
Rosen, G. (1990) "Modal Fictionalism," *Mind* 99: 327–54.
Shalkowski, S. (1994) "The Ontological Ground of the Alethic Modality," *Philosophical Review* 103: 669–88.
Sidelle, A. (1989) *Necessity, Essence and Individuation*, Ithaca, NY: Cornell University Press.
Stalnaker, R. (1976) "Possible Worlds," *Nous* 10, 65–75.
Thomasson, A. (2007) "Non-Descriptivism About Modality: A Brief History and Revival," in *The Baltic International Yearbook of Cognition, Logic and Communication, Vol. 4: 200 Years of Analytical Philosophy*, 1–26.
Williamson, T. (1998) "Bare Possibilia," *Erkenntnis* 48: 257–273.
——— (2002) "Necessary Existents," in A. O'Hear (ed.), *Logic, Thought and Language*, Cambridge, Cambridge University Press, 233–51.
——— (2013) *Modal Logic as Metaphysics*, Oxford: Oxford University Press.
Wittgenstein. L. (1964) *Remarks on the Foundations of Mathematics*, E. Anscombe (Trans.), Oxford: Blackwell.
Wright, C. (1980) *Wittgenstein on the Foundations of Mathematics*, London: Duckworth.
——— (1992) *Truth and Objectivity*, Cambridge, MA: Harvard University Press.

Further readings

Borghini, A. (2016) *A Critical Introduction to the Metaphysics of Modality*. London: Bloomsbury. This treats in greater detail most of the versions of (anti-)realism discussed here.
Divers, J. (2009) "Possible Worlds and Possibilia," in LePoidevin, R. et al. (eds.), *The Routledge Companion to Metaphysics*, Oxford: Taylor & Francis, 335–45. This summarizes the classical debate between genuine and ersatz realists and provides references to the primary literature.
Thomasson, A. (2007) "Non-Descriptivism About Modality: A Brief History and Revival," in *The Baltic International Yearbook of Cognition, Logic and Communication, Vol. 4: 200 Years of Analytical Philosophy*, 1–26. This is a comprehensive survey of and introduction to the literature on non-reductive ideological modal anti-realism.
Williamson, T. (2013) *Modal Logic as Metaphysics*, Oxford: Oxford University Press. This extensive and demanding work is the most influential and comprehensive contemporary defence of a combined ontological and ideological modal realism.

Chapter 12
MODAL CONVENTIONALISM

Ross P. Cameron

Necessity, analyticity and convention

Necessity is both metaphysically and epistemologically troubling. Metaphysically, it is hard to see how it is that necessity gets bestowed on a proposition. The world being the way it is makes things *true*, but how can things being the way they are make things *necessarily true*? Epistemologically, it is hard to see how we could ever *recognize* necessity. Our observations are, seemingly, of things merely being the case; how could we detect their *necessarily* being the case?

One way of making necessity metaphysically and epistemologically tractable is to see it as less a feature of the world itself and more a feature of the way in which we describe the world. Conventionalism about modality is one way of doing this. The conventionalist sees necessity as arising not because of some modal features of reality itself, but because of how we decide to speak about reality.

For example, it is true (let us suppose, at least) that all vixens are shy. Foxes in general are shy. But this might not have been true: foxes have been made shy by circumstance, and there could have been bold foxes. It is also true that all vixens are female. But this could not have failed to be true. It is necessary that all vixens are female. There could have been a bold vixen, but not a male vixen. So we have two true generalizations about vixens, one of which is necessary, the other contingent.

The first version of conventionalism we will look at explains this difference as follows. 'Vixen' just *means* 'female fox'; and so, the linguistic conventions we adopted that govern the term 'vixen' ensure that 'All vixens are female' is true. Thus (the thought goes), the world doesn't have to play any role in making it true that all vixens are female: it is an *analytic* truth – true solely in virtue of meaning, or true solely in virtue of the linguistic conventions we have adopted.[1] This is supposed to explain the necessity of this claim, for if the world plays no role in its truth, there is no worldly variation for its truth-value to be sensitive to. What's going on in the world, says the conventionalist, is irrelevant to the truth-value of 'All vixens are female': the meaning of the words alone suffices for the truth of the claim. By contrast, it is no part of the meaning of 'vixen' that vixens be shy. No conventions in place surrounding our use of the words determine that vixens are shy. For this to be true, the world has to play a role: vixens have to be disposed to certain behaviors. Since the world might not have played that role – since vixens could have behaved differently – it is contingent that vixens are shy.

A.J. Ayer (1936: 31) gives a representative statement of this thought: "I allow [logical and mathematical truths] to be necessary and certain only because they are analytic . . . The reason why these propositions cannot be confuted in experience is that they do not make any assertion about the empirical world." But here is a worry about Ayer's thought. Grant that Ayer has given us a conventionalist explanation of the necessity of analytic truths. What is the explanation of the contingency of synthetic truths? Ayer says that analytic truths cannot be confuted in experience *because* they do not make assertions about the empirical world. The presumption seems to be that synthetic propositions *can* be confuted in experience, because they *do* say something about the empirical world. But why does S saying something about the empirical world suffice to ensure that S could be confuted in experience? That follows only if we assume that our experiences could have been otherwise, so as to confirm a different set of synthetic claims. On the face of it, then, this sounds like there *is* some worldly modality: namely, contingency. That while the necessity of analytic truths is explained by their not saying anything about the world, the contingency of synthetic truths is explained by the fact that they do say something about the world, and that our worldly experiences *really* could have been different. Now, it is a perfectly coherent view to say that contingency is a worldly phenomenon, but that necessity is not. The idea being that contingency is the default – the world itself could always have been otherwise – and that necessity arises only when truths are somehow exempt from the tribunal of worldly experience; as Ayer puts it, by not making an assertion about the empirical world. But while this is a consistent position, it is not obviously compatible with the original motivations for conventionalism. For just as it is puzzling how the way things are could render a truth necessary, so is it puzzling how the way things are could render some falsehood possible. Just as we only appear to experience things being the case and not their necessarily being the case, so it is that we do not appear to experience something's merely possibly being the case. The motivations that push us in the direction of seeing necessity as a linguistic rather than a worldly phenomenon seemingly push us in the direction of seeing *all* modality as a non-worldly phenomenon.

Of course, one could attempt to give a conventionalist account of this contingency: we could say that the only sense in which synthetic truths *could* have been confuted in experience – the only sense in which our experiences could have been other than they are – is that our linguistic conventions don't *rule out* our having such alternative experiences. But is this satisfying? Contingency has consequences, for our intellectual and practical endeavors. It is not worth investigating whether there are any female bachelors, nor would it be rational to plan on there being such, because that is not a way things could be: female bachelors are impossible. By contrast, it *is* worth discovering whether I have female students, and planning on that eventuality, because this *is* a way things might be.[2] But if all there is to contingency is not being ruled out by convention, why should contingency matter so? The text of *Pride and Prejudice* doesn't rule out my having female students, but this doesn't give me any reason at all to plan on that eventuality. Why would not being ruled out by linguistic convention give me such a reason? The text of *Pride and Prejudice* and our linguistic conventions don't rule out my having female students for pretty much the same reason: they simply don't speak to the issue. *Pride and Prejudice* is not at all about the gender of my students, and neither are our linguistic conventions. What relevance, then, could the fact that they don't rule out this scenario – a scenario they are simply silent about – have as to whether it is a scenario worth investigating and planning for? Not being ruled out by convention only seems modally relevant if there is genuine *worldly* contingency that all synthetic claims are susceptible to, and that necessary truths are only not susceptible to because their being true by convention means that they are not saying anything about the world, and thus are not sensitive to the genuinely contingent ways the world is.

But let's put this worry aside. Our first version of conventionalism is:

Crude Conventionalism: S is necessary iff it is analytic.

Crude Conventionalism promises to make necessity metaphysically tractable by locating its source in our descriptions of the world rather than the world itself, and epistemologically tractable by reducing knowledge of necessity ultimately to knowledge of meaning. But while Crude Conventionalism was a popular doctrine during the heyday of logical positivism,[3] it has few contemporary adherents, largely due to Quine's attack on the very notion of truth by convention.[4]

Conventions, argued Quine, cannot *create* truths. Conventions can only *transform* one truth to another. For example, since a convention is in place that 'vixen' means 'female fox', we can transform 'All vixens are female' into 'All female foxes are female'. Our linguistic conventions guarantee that the former sentence has exactly the same truth-conditions as the latter. But convention, argues Quine, plays no role in *making* either claim true. Of course, it's a mere matter of logic that all female foxes are female: an instance of the general logical principle that all Fs which are G are F. But we didn't *decide* that this principle of logic holds. It would be incredible indeed if humans had the power to legislate what laws of logic obtain by choosing what linguistic conventions to adopt! It is a matter of how the world is that all Fs which are G are F – a matter of what the correct logic is. Convention lets you assimilate 'All vixens are female' to a logical truth (for without knowing the relevant conventions in play, you would take this to be the logically contingent claim that all Fs are G). But its truth-value is still sensitive to worldly facts: namely, what the laws of logic are. And we don't get to determine the laws of logic by choosing what conventions to adopt. Likewise with the notion of analyticity. How could anything be true solely in virtue of meaning? Doesn't the meaning of a sentence determine merely what it says about the world? Then it will be a matter of how the world is, whether that sentence is true: whether what the sentence says is the case *is* the case.[5]

Many, including myself, take this problem to be utterly damning for Crude Conventionalism. But notions like analyticity and truth by convention still have their defenders, so let us grant their legitimacy *pro tem* and move on to another problem.

The problem of synthetic necessities

Another problem for Crude Conventionalism is that it seems that the necessary truths outrun any plausible list of the analytic truths. Saul Kripke's (1980) seminal work *Naming and Necessity* convinced most philosophers that the notions of necessity, analyticity and *a priority* are not extensionally equivalent. Relevant to our current discussion is Kripke's argument that there are many synthetic (and a posteriori) necessary truths, such as:

Water is H_2O.
Emilio Estevez is the progeny of Martin Sheen.
Michelangelo's *David* is made of marble.

These truths, says Kripke, could not have been otherwise. H_2O is the deep structure of the substance water, and anything that failed to have this structure, he argues – however similar in surface respects – would not be that very substance. The origin of a person is part of what makes them that very person, and so someone not born of Martin Sheen – even if qualitatively identical to Emilio Estevez – would not be that very person. What constitutes an object

is part of what makes it what it is, so if Michelangelo had decided to make a sculpture out of clay, say, it would not be the very statue *David*, no matter how similar it turned out in other respects. So these three claims are, Kripke argues, necessary. But they are not plausibly analytic, or true by convention. After all, it was a *discovery* that water is H_2O. Prior to this discovery, speakers of English could have been perfectly competent with their usage of the terms 'water', 'is', and 'H_2O', but they would have been in no position to know that water is H_2O. No amount of reflection on the meanings of the constituent terms, or on the conventions governing our language, can lead one to know that water is H_2O. It is a paradigmatically worldly truth – a discovery of empirical science. Similarly, it is surely not true by convention that *David* is made of marble. When would we have decided to use 'David' so? The only decision we made governing that name was the act of baptism: it is to pick out *that* statue. We then have a philosophical argument that *that* statue could not have failed to be made of marble – an argument that does not turn at all on the meanings or conventions governing our words.

The existence of synthetic necessary truths shows that Crude Conventionalism is not even extensionally adequate. *Even if* some necessary truths are true by convention, one cannot say that this is what necessity in general consists in. Now, of course it is open to the Crude Conventionalist to simply reject the phenomena and deny the necessity of such claims. After all, there is much disagreement amongst philosophers as to what is necessary or contingent, and certainly not everyone has been convinced by Kripke's arguments for the necessity of origin, constitution, etc. But it would be inadvisable for the conventionalist to commit to this strategy. While philosophers disagree about the cases, the idea that there are *some* synthetic a posteriori necessary truths is overwhelmingly popular. At the very least, the claim that there are such truths hardly seems *incoherent*. To dig one's heels in and reject the very notion of synthetic necessary truths invites the charge that you are simply talking past your opponents. Perhaps the Crude Conventionalist has a good account of *logical* necessity, but we are looking for an account of *metaphysical* necessity, and it is no part of that notion that metaphysically necessary truths be analytic. Better, I think, for the conventionalist to abandon Crude Conventionalism.

That is the route taken by Alan Sidelle (1989). Sidelle retains the conventionalist's focus on analyticity but abandons the Crude Conventionalist's claim that the necessary truths just are the analytic truths. Instead, he says merely that the necessity of any necessary truth must be *explained by* our linguistic conventions. He says (Sidelle 1989: 29–30), "The basic claim of the conventionalist is that it is our decisions and conventions that explain and are the source of modality." This gives us:

> *Sophisticated Traditional Conventionalism*: For every necessary truth, S, the necessity of S is explained by our linguistic conventions.

The mere existence of synthetic necessary truths does not threaten Sophisticated Traditional Conventionalism, but of course we are owed a story about how the explanation goes. How can convention be the source of the necessity of a synthetic, empirical claim?

Sidelle's answer is that every synthetic necessary truth has its basis in a 'principle of individuation' that is itself analytic. So consider the fact that water is H_2O. While this is certainly not analytic, what *is* analytic, argues Sidelle, is that *whatever* the ultimate structure of water happens to be, water necessarily has that very structure. So while prior to the discovery of water being H_2O, competent language users were in no position to know that water is H_2O, they *were* in a position, thinks Sidelle, to know each of the following conditionals:

If water is H_2O, then, necessarily, water is H_2O.

If water is H_3O_4, then, necessarily, water is H_3O_4.

If water is made out of two parts ectoplasm and one part unicorn heart, then, necessarily, water is made out of two parts ectoplasm and one part unicorn heart.

And while no convention was ever put in place to determine that *David* be made of marble, what is true as a result of our linguistic conventions, thinks Sidelle, are conditionals such as:

If *David* is made of marble, then, necessarily, *David* is made of marble.

If *David* is made of clay, then, necessarily, *David* is made of clay.

If *David* is made of cheese and marshmallow fluff, then, necessarily, *David* is made of cheese and marshmallow fluff.

Because it is a matter of convention that *whatever David* is made from, it is necessarily made from *that*.

With this stock of analytic conditionals, once we discover the relevant synthetic a posteriori truth, we can infer the resulting modal claim from the relevant conditional. So, e.g., once we discover empirically the synthetic truth that water is H_2O, we can reason as follows:

1) Water is H_2O. (Synthetic a posteriori truth)
2) If water is H_2O, then, necessarily, water is H_2O. (Analytic truth)

Therefore,

3) Water is necessarily H_2O.

And Sidelle (1989: p. 37) concludes that "all the modal force of this conclusion [3] will be derived from our general principle [of which premise 2 is an instance], which we are supposing to be analytic."

Here are some problems with Sidelle's account. Just how plausible is it that claims like 'Whatever a statue is made out of, it is made out of that necessarily', or 'Whatever the ultimate structure of water is, it necessarily has that structure', are analytic? Someone who proclaims the existence of male vixens can, perhaps, be accused of lacking mastery with the relevant terms: you must not know what 'vixen' *means* if you think there are male vixens. Claiming that water has its ultimate structure contingently does not seemingly invite the same charge. Especially when it comes to things like the essentiality of origin or constitution, these cases are hotly debated. Kripke mounts *arguments* for the necessity of origin and constitution[6] – he does not simply ask us to accept them on the basis of reflection on how we use the terms – and metaphysicians in turn provide counterarguments. None of this has the feel of one group being linguistically incompetent, or of their talking past one another.

Even if the relevant conditionals *are* analytic, it is not obvious that this is enough for the conventionalist, as Stephen Yablo argued in an influential review of Sidelle's book.[7] Let's suppose that it is analytic that whatever water's ultimate structure, it has that structure necessarily. A natural way to understand this is that we have decided to use the term 'water' only to pick out a substance that has its ultimate structure, whatever it is, necessarily. We can certainly decide to use the word 'water' so. But that wouldn't render the essential properties of water a matter of convention. Rather, it would render whether 'water' refers hostage to fortune: there will only be water if there is a substance with such modal features. The modal features of the substance

themselves, for all that's been said, are a matter of how the world is. Instead of explaining away modal features of the world, all that seems to have been accomplished is that we've said there is only water if there are modal features of the world.

Gillian Russell (2010) has recently raised a further problem. Any attempt to explain necessity by appeal to convention has to grapple with the fact that conventions seem to be concerned with the meaning of *sentences*, whereas necessity seems to be a status of the *proposition* expressed. In order to bridge this gap, the conventionalist apparently needs some kind of bridging principle: a way of moving from the claim that the sentence is true by convention to the claim that the proposition expressed is thereby necessary. But one sentence can express different propositions in different contexts. What if a single sentence, S, expressed a contingent truth in one context, C1, and a necessary truth in another context, C2? We have two propositions: the contingent proposition, expressed by S in C1, and the necessary proposition, expressed by S in C2. The conventionalist has to hold that the necessity of the latter proposition is explained by the conventions governing the sentential vehicle by which it is expressed. But how can this be, given that exactly the same conventions govern the sentential vehicle that expresses the contingent proposition – after all, it is the very same sentence. There must be a difference between the two cases, since in one case we say something contingent and in the other something necessary, but the difference cannot lie at the level of the sentence uttered, since it is the same sentence in each case. Thus, it is very hard to see how the difference can lie in a difference in convention. Rather, it seems to lie in something worldly: a difference in the modal status of the different propositions expressed.

Are there such sentences? Russell offers as an example 'I am here now' as said by you and by God. When you utter this sentence, you say something contingent: you could have been elsewhere. But God – a necessary existent who is necessarily omnipresent (let us suppose) – says something necessary with the same sentence. There is no equivocation here. It's not that God means something different by 'here'.[8] God's conventions governing the use of these terms are the same as ours. The difference between the two cases doesn't appear to be a difference in convention, but rather a difference in modal facts: God's relation to the world and the locations in it is not modally flexible in the way our own is. It is hard to see a conventionalist explanation of this.

And of course, it is open to the Sophisticated Traditional Conventionalist to simply reject the example and deny that God is necessarily here now. But to resist all examples like Russell's, she would have to deny that we can ever have a sentence that expresses a contingent truth in one context and a necessary truth in another. But dialectically, this seems objectionable: akin to the Crude Conventionalist's denial of the phenomenon of synthetic necessary truths. The advantage of Sophisticated Traditional Conventionalism over Crude Conventionalism was its ability to take on board the range of metaphysical necessities that contemporary philosophers countenance. It would be strange now to close the door on a phenomenon that certainly seems intelligible just because they cause trouble for the view.

Neo-conventionalism

Recently, a different form of conventionalism about necessity has been defended.[9] *Neo-Conventionalism* abandons the notion of truth solely by convention, or solely in virtue of meaning. Instead of taking it to be a matter of convention that certain claims are true, as traditional conventionalism did, the Neo-Conventionalist grants that all truths are true because of what the world is like, and instead says that it is a matter of convention what the extension of 'is necessary' is. The idea here is that while there is no truth by convention – there are just all the truths, true because of how things are – it is a matter of convention that we deem some of these truths necessary and others contingent.

Consider the predicate 'is good-mannered'. It applies to certain behaviors – writing a thank-you note to your aunt if she gets you a birthday present, saying 'please' if you ask your waiter for more water, etc. – and it doesn't apply to other behaviors – texting during a performance of a play, pouring yourself a drink without offering your guests any, etc. There is clearly a sense in which the extension of 'is good-mannered' is a mere matter of convention. There is no deep feature in common between all the good-mannered acts that is lacked by all the non-good-mannered acts. We could have deemed a different set of actions good-mannered, and we would not be somehow misdescribing reality. (Cultures that hold that burping after a meal is good-mannered are not making a *mistake*; they've just deemed different behaviors good-mannered.) To say that the extension of 'is good-mannered' is fixed conventionally is to say that we are not *tracking* some feature of behavior when we deem it good-mannered – there is no objective state of good-mannered-ness, possession of which makes an attribution of the predicate *correct* – rather, we ourselves are *determining* what counts as good-mannered by *decreeing* what behaviors the predicate applies to.

The Neo-Conventionalist views attributions of 'is necessary' to propositions in much the same way. There is, she thinks, no deep feature that all the necessary truths have that the contingent truths lack, possession of which makes an attribution of 'is necessary' correct. We do not *track* modal status with our attributions of modal terms; rather, we *determine* modal status by *decreeing* what truths count as necessary.

What does it mean to say that there is no deep feature that all the necessary truths have that the contingent truths lack? In the terminology popularized by David Lewis, it is to say that the distinction between the necessary and the contingent is not a *natural* one: it is not a distinction that carves the world at its joints.[10] Some distinctions are natural – they reflect genuine divisions in reality. Consider, for example, the distinction between electrons and everything else. This isn't some arbitrary distinction: reality really is divided in such a way as to single out these fundamental particles, electrons, as *different* from other things. By contrast, consider the things on this list: all electrons within one meter of a goat, the first burrito purchased today, my dog's favorite frisbee. There is a distinction between the items on this list and everything else, of course, but the distinction is entirely *arbitrary*. This is not a distinction that reality itself draws; this division is not carving reality at its joints. If we failed to mark the distinction between the things on that list and everything else, we wouldn't be missing out on anything: the division does not track a genuine division in reality. By contrast, if we lacked the resources to distinguish the electrons from the non-electrons, we would be missing something important about the structure of reality: we'd be missing a genuine division amongst the things that there are. That is because the distinction between electrons and non-electrons is a *natural* one, and the distinction between the items on that arbitrary list and everything else is a highly gerrymandered, or unnatural, one. Natural distinctions are ones that reflect objective similarity. Any two electrons are alike in an objective respect – they belong together – and they are unlike any non-electron in an objective respect. Whereas there is nothing objectively in common between an electron within one meter from a goat and the first burrito purchased today – at least, nothing that is not also had by things not on that list. No objective feature *unites* the items on that list, which is why it is gerrymandered or unnatural – a mere collection, rather than a genuine worldly unity.

The Neo-Conventionalist thinks the necessary truths are a relatively gerrymandered collection. On her view, there is no objective similarity between the necessary truths that 2 + 2 is 4, that water is H_2O, that *David* is made of marble, etc., that is lacked by the contingent truths that there are eight planets, that *David* is in Rome, that I am a philosopher, etc. For the Neo-Conventionalist, there is no worldly, natural distinction between the necessary and the contingent truths to track; there is only the gerrymandered, unnatural distinction that we ourselves carve.

Let us return to Yablo's objection to Sidelle's view. Of course we can decide to use 'water' such that it applies only to a substance that has its ultimate structure necessarily, but as Yablo says, all this accomplishes is that we've decided to let this term refer only if the world comes furnished with things with the appropriate essences. But as Sider (2011: 283) says, we can think of Sidelle's principles of individuation as encoding not our decision how to use 'water', etc., but rather how to use the term 'necessary'. The idea is that when we say 'Whatever the ultimate structure of water is, it has that structure necessarily', we are indicating that whatever of the following claims turns out to be true,

> Water is H_2O.
> Water is H_3O_4.
> Water is made out of two parts ectoplasm and one part unicorn heart.

It is to count as a necessary truth – not on the basis of some special objective property that this truth has that 'Water fills this glass' (e.g.) lacks, but simply because we conventionally determine the extension of 'is necessary' to be so.

For the Neo-Conventionalist, there is no sense at all in which the truth of 'All vixens are female' is due to convention. It is true because all female foxes are female, and that is true because of what the world is like. The Neo-Conventionalist does not think we can decree what laws of logic hold. Thus, they are not threatened by Quine's objections to truth by convention or analyticity. Nor does the Neo-Conventionalist face any problem in the face of synthetic necessary truths: we can decree that 'Water is H_2O' and '*David* is made of marble' are to fall under the extension of 'is necessary' just as much as we can decree that 'All bachelors are unmarried' and '2 + 2 = 4' are to so fall. And because what is settled by convention is what *propositions* count as necessary, the Neo-Conventionalist also avoids Russell's problem. The Neo-Conventionalist can decree that the proposition God expresses when He says 'I am here now' is necessary, but that the proposition I express by uttering the same sentence is not necessary. There is no problem with the same sentence being able to express (in different contexts) a necessary and a contingent proposition, because the necessity of the proposition expressed is not being said to be a result of the meaning of that sentence, or of the conventions governing use of the terms in that sentence.

A big question faces the Neo-Conventionalist. I can introduce a new term 'is squarky' and tell you that it is to be used as follows: the number four, the Scottish Parliament and all presidents are squarky, and nothing else is. I have now fixed the extension of 'is squarky'. You now know that the successor of three and Barack Obama are squarky, but the square root of nine and Taylor Swift are not. The only puzzle about any of that is, what's the point? We've settled the extension of the term, but it is a wholly useless addition to the language. Likewise, we can lay down some rules concerning how 'is necessary' is to be used – what (kinds of) propositions it is to apply to – but if modalizing is to have a point to it at all, we had better have a story as to why drawing this division amongst true propositions is an interesting or useful practice.

Those who see necessity as a deep feature of reality have an easy answer to the question of why it's useful or interesting to draw modal distinctions: to describe reality! If there is a deep metaphysical distinction between the necessary and the contingent truths, then we have a reason to want to describe that distinction, because we are interested in how the world is. But if we are not tracking the distinction between the necessary and the contingent, but are rather determining it, then we need to say why the distinction is of interest. We need an account of why this distinction *matters* – an account that, true to conventionalist motivations, emphasizes its importance to *our* intellectual life and practices. The question of the function and role of modal

judgments has fallen somewhat off the radar in recent philosophical discussions of modality, as the mood has trended towards a more realist treatment of modal truths. But if less realist approaches like Neo-Conventionalism are to be satisfactorily developed, the question of why we modalize becomes crucial once again.[11, 12]

Notes

1 I will treat 'is analytic', 'is true solely in virtue of meaning' and 'is true solely in virtue of convention' as interchangeable here. There are conceptions of analyticity that abandon the idea of truth solely in virtue of meaning/convention (see e.g. Boghossian 1997; Russell 2008; Sider 2011: 191–195), but these are not the notions our conventionalist needs for her goals.

2 Admittedly, many philosophers today would want to at least qualify this claim about the link between contingency and what is worth investigating or planning for, in light of Kripke's (1980) examples of necessary truths whose negation is nonetheless epistemically possible, and so might (as far as we know) be true. But (i) I think something like this claim is still going to be true, and is not going to be explainable if contingency is merely not being ruled out by convention; and (ii) Ayer can hardly offer this defense, since by his lights what is necessary, what is certain and what is *a priori* are all one and the same.

3 See e.g. Ayer (1936).

4 See primarily Quine (1936), but also Quine (1951). Cf. Boghossian (1997), Hale (2002), and Sider (2003).

5 See especially Boghossian (1997).

6 See e.g. footnotes 56 and 57 of Kripke (1980).

7 Yablo (1992).

8 Or more carefully: there is at least one sense of meaning – what Kaplan (1989) calls *character* – that is constant over the two uses of 'here'. And as Russell says, it is character that is surely relevant to the conventionalist's claim that necessity is ultimately explained by meaning, since it is knowledge of character that is required for mastery of a term. To know how to use 'here', I don't need to know where any speaker is, I only need to know how 'here' interacts with the location of the speaker to determine the proposition they thereby express.

9 See Cameron (2009, 2010a, 2010b) and Sider (2003, 2011: ch. 12).

10 Lewis (1983). As applied to necessity, see Sider (2003), Cameron (2009) and Nolan (2011). For a detailed treatment of the notion of naturalness as applied to many domains, see Sider (2011).

11 For some recent approaches to the question, amenable to less realist conceptions of modal truth, see Divers (2010), Divers and Elstein (2012) and Thomasson (2007, 2013).

12 Thanks to Elizabeth Barnes, Kris McDaniel, Chris Menzel, Amie Thomasson, Jason Turner and Robbie Williams for helpful comments.

References

Ayer, A.J. (1936). *Language, Truth and Logic*. London: Victor Gollancz Ltd.

Boghossian, P. (1997). Analyticity. In: B. Hale and C. Wright, eds., *A Companion to the Philosophy of Language*. Oxford: Blackwell, pp. 331–368.

Cameron, R. (2009). What's Metaphysical about Metaphysical Necessity? *Philosophy and Phenomenological Research*, 79(1), pp. 1–16.

Cameron, R. (2010a). The Grounds of Necessity. *Philosophy Compass*, 4(4), pp. 348–358.

——— (2010b). On the Source of Necessity. In: B. Hale and A. Hoffmann, eds., *Modality: Metaphysics, Logic and Epistemology*. Oxford: Oxford University Press, pp. 137–152.

Divers, J. (2010). Modal Commitments. In: B. Hale and A. Hoffmann, eds., *Modality: Metaphysics, Logic, and Epistemology*. Oxford: Oxford University Press, pp. 189–219.

Divers, J. and Elstein, D.Y. (2012). Manifesting Belief in Absolute Necessity. *Philosophical Studies*, 158(1), pp. 109–130.

Hale, B. (2002). The Source of Necessity. *Philosophical Perspectives*, 16, pp. 299–319.

Kaplan, D. (1989). Demonstratives. In: J. Almog and J. Perry and H. Wettstein, eds., *Themes from Kaplan*. Oxford: Oxford University Press, pp. 481–563.

Kripke, S. (1980). *Naming and Necessity*. Cambridge, MA: Harvard University Press.

Lewis, D. (1983). New Work for a Theory of Universals. *Australasian Journal of Philosophy*, 61, pp. 343–377.
Nolan, D. (2011). The Extent of Metaphysical Necessity. *Philosophical Perspectives*, 25(1), pp. 313–339.
Quine, W.V. (1936). Truth by Convention. In: O.H. Lee, ed., *Philosophical Essays for A.N. Whitehead*. New York: Longmans, pp. 90–124.
——— (1951). Two Dogmas of Empiricism. *Philosophical Review*, 60, pp. 20–43.
Russell, G. (2008). *Truth in Virtue of Meaning: A Defence of the Analytic/Synthetic Distinction*. Oxford: Oxford University Press.
——— (2010). A New Problem for the Linguistic Doctrine of Necessary Truth. In: C.D. Wright and N.J.L.L. Pedersen, eds., *New Waves in Truth*. New York: Palgrave Macmillan, pp. 267–281.
Sidelle, A. (1989). *Necessity, Essence, and Individuation: A Defense of Conventionalism*. Ithaca, NY: Cornell University Press.
Sider, T. (2003). Reductive Theories of Modality. In: M. Loux and D. Zimmerman, eds., *The Oxford Handbook of Metaphysics*. Oxford: Oxford University Press, pp. 180–208.
——— (2011). *Writing the Book of the World*. Oxford: Oxford University Press.
Thomasson, A. (2007). Modal Normativism and the Methods of Metaphysics. *Philosophical Topics*, 35, pp. 35–160.
Thomasson, A. (2013). The Nancy D. Simco Lecture: Norms and Necessity. *Southern Journal of Philosophy*, 51(2), pp. 143–160.
Yablo, S. (1992). Review of Alan Sidelle's *Necessity, Essence, and Individuation: A Defense of Conventionalism*. *Philosophical Review*, 101, pp. 878–881.

Chapter 13
NORMS AND MODALITY

Amie L. Thomasson

Claims about necessity and possibility play a central role in metaphysical debates. Consider classic puzzles about material constitution: a statue may be made of a lump of clay, but it seems that the lump of clay *could survive* certain changes in shape that the statue *could not* survive. But how could they differ in these 'modal properties' when the statue and clay are otherwise identical? Or consider the ancient Ship of Theseus problem: if the planks of a ship are gradually replaced with new ones, and then the old planks are reassembled, which is the original ship: the one with the new planks, or the old? Can a ship survive changes in all of its parts? Can it survive disassembly and reassembly? Other puzzles arise about personal survival and identity, when we ask whether persons could survive the loss of memories, brain transplants, or teleportation.[1]

Answering metaphysical questions like these requires us to determine which statements about what is metaphysically possible or necessary are true. But how can we do that? If we think of our modal claims as describing the world, it's natural to think that we have to find out whether there are the needed *truthmakers* for our claims of metaphysical necessity or possibility. However, as I will argue in this chapter, the search for truthmakers for metaphysical modal claims leads to a morass of ontological and epistemological problems. Here I will argue for a different approach to understanding metaphysical modal claims: thinking of them not as serving to *describe* modal features of the world (nor as describing other possible worlds), but rather serving a *normative* function of conveying semantic rules and their consequences. Understanding metaphysical modal claims in this way, I will argue, enables us to demystify the ontology and epistemology of modality, and to clarify the epistemology of metaphysics.

13.1 The search for modal truthmakers

It has become standard to think that, as Tony Roy puts it, "the problem of modality is a problem about truthmakers for modal propositions" (2000, 56). But what could serve as the truthmakers for our claims of metaphysical necessity or possibility? Are they modal features of this world—modal facts, or modal properties? Call a view a form of 'heavyweight modal realism' if it holds that there are distinctively modal properties or facts that *explain* what it is that *makes* some modal statements true. It is tempting to think that when we say: "A person can survive a transplant of her brain into another body", or "A painting cannot survive a process that involves replacing 40% of the original surface paint" these statements simply *describe* what modal properties people

or paintings have (the modal property of (not) possibly-surviving-this-change), just as we describe the person or painting as having other properties (weight, color, etc.). Alternatively, we might think of modal statements as attempts to describe modal *facts* in the world, and as holding true if those modal facts indeed obtain.

But heavyweight modal realism faces what Huw Price (2011) has called the 'placement problem': how are these modal facts or properties supposed to fit into the natural world? They do not seem to be physical properties or facts like those investigated by the empirical sciences—so what are these modal facts or properties supposed to be, and why should we think there are such things? How are they related to the non-modal facts or properties studied by the natural sciences? As the case of the clay and the statue makes clear, this relation is pretty mysterious—because objects like the statue and clay can have all the same physical properties and yet differ in their modal properties. As a result, it seems like we can't think of the modal properties as 'higher level' properties that are somehow fixed by an entity's more basic physical, or other non-modal properties.

If we think that, to know which metaphysical statements are true, we have to come to know their truthmakers, we face even more difficult epistemological problems. For how can we hope to discover what modal properties an object has, or what modal facts obtain? Hume observed long ago that there seems no prospect of giving an empirical account of our knowledge of modality—so what sort of account can we give? Even if one thinks, with Barbara Vetter, that one can acquire knowledge of dispositions or potentialities empirically (having observed many glasses shattering, we infer that glasses are *disposed* to break on sharp impact) (Vetter 2015, 11–13), the same methods cannot ground knowledge of distinctively *metaphysical* modality. For the same observations of the statue/lump of clay lead to different modal conclusions: that the statue would not survive a squashing while the lump would. The so-called 'grounding' problem arises precisely because no non-modal/non-sortal properties that *can* be empirically known seem capable of grounding the difference in the *metaphysical* modal properties (identity and persistence conditions) attributed to the statue and the clay.[2]

One prominent alternative to heavyweight modal realism is David Lewis's (1986) possible worlds realism. On this view, we need not accept that there are any distinctively *modal* features of this world. Instead, we accept that there are many other concrete worlds causally and spatio-temporally disconnected from our own—call these the (merely) possible worlds (our actual world is also a possible world). Lewisian possible worlds realism still enables us to understand modal statements as descriptions, but they are seen as attempted descriptions not of *modal* features of *our* world, but of *non-modal* facts in one or more of the possible worlds (including our actual world).[3]

However, few have been willing to accept that there are such possible worlds, or that they could adequately serve as non-modal truthmakers for our modal propositions (Jubien 2007; Divers and Melia 2002). Even if we are content to let the ontological issues slide and allow that there *are* possible worlds, a massive problem of *relevance* remains—a problem that was raised in reviews of *The Plurality of Worlds* by Nathan Salmon (1988), Graeme Forbes (1988), William Lycan (1988), and Allen Stairs (1988). The problem is that our modal statements just don't seem to be *about* what goes on in other worlds, even if there are such worlds (Stairs 1988, 344); these don't seem like *relevant* truthmakers.[4] The fact that even if we accept the ontology, its *relevance* to the modal question remains in doubt, might already be taken to suggest that the modal claims aren't aiming to *describe* facts about other worlds at all.

Lewis's possible worlds realism also faces daunting epistemic problems like those that trouble the heavyweight realist. For given the causal isolation of the other possible worlds from our own, it remains unclear how we could know anything about them, or about what claims they make

true. Lewis (1986, 104–115) does suggest that we commonly come to have the modal beliefs we do by way of engaging in imaginative experiments guided by a principle of recombination (113–114). But it is not at all clear, on his view, why this sort of procedure *should* give us anything that counts as modal *knowledge*: why should imagination guided by this principle give us any information about what is going on in causally and spatio-temporally disconnected concrete worlds? Lewis himself makes it clear that he does not take himself to be answering that question.[5]

13.2 Exposing the descriptivist assumption

As we have seen, a daunting array of puzzles and problems arises from the attempt to find truthmakers for our metaphysical modal statements—whether we think of the truthmakers as modal properties, modal facts, or possible worlds. What I want to call attention to in this chapter is that they all arise from an assumption so common that it has become almost invisible: the assumption that modal statements are *descriptive* or *representational* in character—that is, that they function to describe or track certain 'features of reality', which can then serve to make modal claims true.[6]

I will argue that this assumption should be brought to light and reexamined. Before we start asking what the truthmakers are for modal claims, we should step back to ask a more basic question: what *function* does it serve to make metaphysical modal claims? Do they serve a kind of describing or world-tracking function, or do they have another function entirely—in a way that might make the search for truthmakers otiose?

In one sense it might seem just obvious that metaphysical modal statements are descriptive. If I say 'Mary is necessarily human', this sounds parallel to saying 'Mary is unusually tall'; both are indicative in form and are naturally thought of as 'describing' (or 'representing') Mary in certain ways. I certainly do not wish to deny these obvious truths.[7] The philosophical assumption of descriptivism, however, goes beyond these truisms: it involves an assumption about *function*. Many of the basic terms in our language seem to serve the function of tracking certain features of our environment, with which they are meant to co-vary, enabling us to get around better. So, for example, it is plausible that terms such as 'wolf' and 'river' serve such a descriptive tracking function. To assume that a term is descriptive in our sense is to assume that it serves that function. Huw Price (2011) calls terms that serve this function 'e-representations', those whose job "is to co-vary with something else—typically, some *external* factor, or environmental condition" (20). Where terms serve this kind of function, it is natural to think of them as aiming to correctly represent what there is (and is not) in the world—and answerable to the world in the sense that we look to the world to determine if what we say using those terms is true or false. When discourse is descriptive in this way, it seems natural to look to the world to find 'truthmakers'[8] for our claims about wolves or rivers, features of the world that 'explain how sentences about the real world are made true or false' (Mulligan et al. 1984, 288).

But there are plausible and interesting philosophical accounts of some central, philosophically interesting terms that treat them as serving very different functions from describing or tracking elements of our environment. For example, Paul Horwich (1998) treats the function of the truth predicate as serving as a device of generalization; Stephen Yablo (2005) suggests that nominative vocabulary for numbers enables us to simplify our expression of scientific laws; and moral expressivists like Simon Blackburn (1993) argue that moral discourse serves not to describe moral facts, but rather to enable us to coordinate our attitudes in certain useful ways.

Even where modal discourse is concerned, the descriptivist assumption wasn't always so invisibly dominant. In the early days of analytic philosophy, it was common to deny that modal statements of various kinds serve a descriptive function. The approach was suggested by early

conventionalists like Schlick (1918),[9] who argued that necessary statements of mathematics and logic are not descriptive statements, but serve as implicit definitions of concepts. It was developed in a new way by Wittgenstein in the *Tractatus* (1922/1933), who treated the propositions of logic as tautologies which say nothing about the world.[10] The approach reappeared in a more sophisticated vein in the work of the later Wittgenstein, and in Ryle's work on statements of scientific laws, which he took to serve not to describe the world but rather to serve as 'inference tickets' (1950, 121). Sellars (1958) develops a similar treatment of statements of scientific laws, which he treats as having the function of *justifying* or *endorsing* inferences from something's being an A to its being a B. More recently, approaches to modality along these lines have been developed by Simon Blackburn (1987/1993) and Robert Brandom (2008).

In the remainder of this chapter, I aim to lay out a view in that tradition, on which *metaphysical* modal statements fundamentally serve not to *describe* features of this or other worlds, but rather serve a basically *normative* function. On reflection, it shouldn't be surprising that modal terms serve a normative function. As a grammatical group, modal terms include not only the metaphysician's 'possible' and 'necessary', but also such terms as 'can', 'may', 'must', and 'shall', which are characteristically used in issuing requirements and permissions, and in stating commands and rules in an impersonal indicative form. These different sorts of modal terms (for alethic, deontic, and epistemic modalities) tend to come together across a wide range of languages (Papafragou 1998, 371), and children tend to learn to use modal terms for obligation, necessity, and possibility at around the same time (about age three) (Wells 1985, 159–160, 253). So it would make sense to think that they have something in common—perhaps that that they all enable us to convey norms in useful and perspicuous ways.

13.3 What function does modal vocabulary serve?

But why might it be useful to have modal terms convey rules and norms? There are surely other ways of communicating rules and norms, such as in non-modal imperatives, or with a stick. Let me begin by discussing why modal terms might be useful in conveying rules in general, and then turn to the particular role of *metaphysical* modal terms in conveying *semantic* rules.

To see what functions modal terms might serve in conveying rules, think about the way rules of games are expressed in language. Some rules may be expressed in imperatives, for example, for checkers:

- "Move only on the dark squares"

But rules can also be expressed in declarative sentences (in the indicative mood), as:

- "Players move their pieces only on the dark squares", or
- "Black always moves first"

Putting rules in the indicative mood has certain advantages over the imperative. For starters, one can't easily state "Black moves first" in an imperative form without knowing who the black player is, and addressing him or her directly. Secondly, expressing rules in the indicative mood enables us to make explicit our ways of reasoning with rules, so that we can, for example, embed them in conditionals and say "If black always moves first, then red never moves first", whereas we cannot put an imperative in the antecedent of a conditional.

But there are also dangers of expressing rules in the indicative form, since they might be mistaken for *descriptions* of what *does happen* or *has happened*—making it hard to distinguish the

expression of the *rule* that black always moves first from the red player's misguided complaint that black *always* moves first.

We can, however, add a modal verb, and say instead:

- "Players *must* move their pieces only on the dark squares" or
- "The black player *must* move first"

Expressing rules in this modal form preserves our ability to make explicit our ways of reasoning with rules—since it is still in the indicative. But it also brings other advantages. First, it clearly distinguishes these statements of rules from mere *descriptions* of what *does* happen. Second, it enables us to express permissions as well as requirements. Neither the imperative form nor the simple indicative enabled us to do that. If we ask, "Does each player take a turn every round?" the only way to give a *negative* answer to this question, as Ryle (1950/1971, 244) pointed out, is to add a modal verb, and say "No, a player *may* choose to skip a turn". These observations lead to the hypothesis that at least one function it serves to have modal terminology in our language is to give us a way of expressing *rules* or *norms* in the indicative mood, in a way that makes the regulative status more explicit, enables us to make explicit our ways of reasoning with rules, and enables us to express permissions as well as requirements.

13.4 The function of metaphysical modal claims

But if modal terms have as their function conveying rules or norms, what rules or norms might be at stake in metaphysical modal claims? The heart of the normativist view of modality is to see metaphysical modal claims as functioning to convey *semantic* rules. The interesting and tricky feature, however, is that they do not do so by *describing* what the semantic rules are. If we said that they *describe* what the rules are—that the adoption of semantic rules is a truthmaker for our metaphysical modal claims—then we would still be stuck with the old descriptivist assumption that we need truthmakers for our modal claims. Moreover, we would fall into the problems of classical conventionalism—our necessary claims would only hold contingently, as it is a contingent matter that we adopt some rather than other semantic rules.

But nor do metaphysical claims of necessity convey semantic rules by *stating* these rules in a metalanguage: metaphysical modal claims are object-language claims. Instead, they *convey* these rules by *using* those very terms—remaining in the object language. Consider the following dialogue:

- Child: "Mom, is Aunt Sophie always going to be a bachelor?"
- Mom: "Bachelors must be men, dear".
- (Pretentious philosophical mom: "It is necessary that bachelors are men, dear".)

What is going on in this dialogue? In the response, Mom just uses the term 'bachelor' and states a necessary truth: It is necessary that bachelors are men. But what she is *doing* thereby is communicating a rule that *could* be stated in the metalanguage, as "The term 'bachelor' may only be applied to men".

Now consider a more philosophical dialogue.

- Question: "Can a ship survive having all of its parts gradually replaced?"
- Response: "Yes, as long as the replacement process is gradual. For all that is essential to artifact identity over time is a continuous history of maintenance, not the retention of any particular material part".

Again, here we have a dialogue conducted in the object language, about ships and other artifacts. But what is being *done* through this dialogue, on the normativist view, is not describing some modal properties of ships that the philosopher pretends to have discovered, but rather communicating some rules of use for ship names (and names for other artifacts): that we are permitted to say this is 'the same ship' as before, or to re-apply the name 'The Queen Mary', as long as there has been a continuous history of maintenance.

Although the primary function of metaphysical modal claims, on this view, is to convey semantic rules rather than to report metaphysical discoveries, it is still useful to do so by just *using* the terms, in the object language. We often use terms as a way of demonstrating or implicitly commenting on how the term is (to be) used, or whether it should be used at all. Chris Barker (2002) calls these 'metalinguistic' uses as contrasted with 'descriptive' uses, where metalinguistic uses are those in which a term is used to "communicate something about how to use a certain word appropriately", rather than to communicate (other) information about the world.[11] For example, we engage in what has become known as 'metalinguistic negation', when one speaker says, "The performance was good tonight", and another replies, "It wasn't good, it was spectacular!" In a case like this, the second speaker apparently *uses* language in order to *show* what choice of words she thinks was appropriate. In other cases, we may sometimes demonstrate how vague terms are appropriately used in a context by using them in certain ways. In some contexts, for example, one might communicate what the standards for tallness are around here by pointing to a man (whose height is not in doubt) and saying "*Jones* is tall" (see Barker 2002, 1–2)—in which we are not adding information about Jones's height, but rather using the term 'tall' in a way that communicates information about how it is appropriately used in this context.

In short, it is not unusual or idiosyncratic for us to communicate standards for language use by *using* it in certain ways. That is exactly what the normativist thinks is going on with claims about what is metaphysically possible and necessary: they are claims in the object language, and so in that sense are world-oriented, 'about the world', not *about* language (just as the aforementioned claims are about the performance or about Jones). Yet their *function* is to convey how the terms *ought to be used*—to convey norms. In the case of metaphysical modal claims, these are semantic norms, typically concerning actual and hypothetical cases in which the term should be applied and refused, or applied again 'to the same thing'. What the addition of the modal verb does, over the simple indicative, is to 'flag' this regulative function, making it more explicit, and enabling us to convey permissions as well as requirements.

So, in sum, on the analysis given here, there are two odd features of metaphysical modal claims. First, like other modal *statements*, though they fulfill a *normative* function, rather than being expressed in imperatives they are expressed in *indicative* form—for good, functional reasons. Second, though simple utterances of claims of metaphysical necessity are in the *object* language, they involve implicitly *metalinguistic* uses of the terms—as ways of conveying something about how the relevant terms are *to be* used. Both of these features can lead us astray into thinking of modal statements as if they are worldly descriptions in need of truthmakers. But once we've noticed the commonalities with other cases in which normative modal language is expressed in indicatives, and object-language claims are used to serve a metalinguistic purpose, we can see that they are not so very strange after all, and then can see our way clear to a more plausible, and less problematic, analysis of modal discourse that does not begin from the drive to seek modal truthmakers in this or other worlds.

But although the normativist doesn't think of modal claims as needing truthmakers, a normativist can nonetheless allow that our modal claims are true or false. Given the rules that (on this view) govern the use of our modal terms themselves, we are entitled to add 'necessarily' onto any object-language expression of an actual semantic rule. So we can begin from "All bachelors

are men" and (since that claim is an object-language expression of a semantic rule) add 'necessarily' and assert "Necessarily, all bachelors are men". We then need only adopt a deflationary understanding of truth (see, e.g. Horwich 1998), according to which the concept of truth is simply governed by the equivalence schema: <p> is true iff p, to recognize the equivalence of this with "<Necessarily, all bachelors are men> is true". The uncontroversial equivalence schema applies just as well to modal as to non-modal indicatives, so there is no problem in allowing that modal claims may be true, stated in propositional form, and used in reasoning.

13.5 Remaining challenges and hopes

The road to developing a full modal normativist approach is long and full of challenges. Some of the challenges I have discussed elsewhere[12]—including showing how it avoids the problems that plagued classical conventionalism, and showing how to avoid the Frege-Geach problem by giving the *meaning* of modal discourse, not just its function or use. Another major and familiar challenge is showing how this view can accommodate not only simple *a priori* necessities like "Necessarily, all bachelors are men", but the sorts of *a posteriori* and *de re* necessities Kripke (1980) famously called attention to: such as 'Water is necessarily H_2O' and 'Elizabeth Warren is necessarily human'. The first key to handling such necessities is to accept that even names and kind terms have *some* conceptual content.[13] The second key is to note the varied and often ostensive and world-deferential forms semantic rules can take. There is no space to develop that solution here, but the original approach was developed in Sidelle (1989) and is applied specifically to the normativist approach in Thomasson (2020).

At any rate, if we can make a normativist view along these lines work, it will be very attractive. One advantage of this view is ontological, and comes from not thinking of modal properties, modal facts, or possible worlds as things our modal claims aim to describe or track, and that are capable of explaining what makes them true. The normativist may allow that there are modal facts and properties—and even other possible worlds—but not in the sense of 'positing' them to 'explain' the truth of our modal claims. Instead, the ontological entitlement to say that there are such things is explanatorily 'downstream' from such truths. That is, we can start by making metaphysical modal claims (which have another function entirely); for example, "Necessarily, all bachelors are men" or "Obama is necessarily human". From these modal truths, we can make trivial inferences to the existence of modal facts (that it is a fact that it is necessary that all bachelors are men) and properties (that Obama has the property of being necessarily human). But we arrive at talk of these modal facts and properties by hypostatizations out of our modal expressions; we do not need to start by 'discovering' these worldly modal features to figure out which modal claims to accept.

The most important advantage of a normativist approach is the epistemological advantage of avoiding the notorious difficulties heavyweight realist views and possible worlds realist views face in explaining our knowledge of modal facts. If we think of ourselves as trying to track and describe the modal properties, modal facts, or possible worlds (the presence of which could make the claims true), it is very difficult to see how we could have any (quasi-)perceptual or intellectual access to these modal facts or properties or worlds. The normativist demystifies modal knowledge by considering the move from *using* language to knowing *basic* modal facts to be a matter of moving from *mastering* the rules for properly applying and refusing expressions (as a competent speaker), to being able to explicitly *convey* these constitutive rules and their consequences in the object language and indicative mood.

A third, related, advantage of the normativist view is its helpfulness in clarifying the methodology of metaphysics and justifying the use of intuition in modal debates. If we think of metaphysicians as trying to 'detect' modal properties of objects, the hopes of adjudicating debates

about the modal features of persons or works of art seem slim: for no one, it seems, has any useful answer to the question of how they are supposed to be detected. Moreover, although it is common to rely on intuitions to support metaphysical views, it's not clear how to justify *why* intuition should be thought a reliable guide to the modal features of the world (when it certainly fails to be a reliable guide about most other features) (see Sosa 2008, 233). But on the normativist view, we have good reason for thinking that intuitions of competent speakers may play a useful role in revealing and making explicit the *actual* semantic rules, and thereby in coming to express modal truths in the object language—and signaling this with the addition of modal verbs.

Nonetheless, as I have argued elsewhere, some uses of metaphysical modal claims may be fruitfully seen as engaged in what David Plunkett and Tim Sundell (2013) have called 'metalinguistic negotiation': as ways of *advocating for* changes in the rules—whether to precisify them or alter them in other ways, in order to serve various purposes, rather than simply as ways of conveying the rules there are.[14] The function of the metaphysical modal claims may still be normative, but it may have to do more with pressing for changes in the rules than with communicating or enforcing the extant rules. The fact that, in the object language, we may often engage in this kind of metalinguistic negotiation of what how our terms *should be* used enables us to account for the fact that debates about metaphysical modality are often enduring and hard to resolve—even among competent speakers. For what is at issue is not just what rules *do* govern the terms (though these, too, may be imprecise, open-ended, contextually variable), but what rules *should* govern our terms—where this is sensitive to a range of other issues about what we should value and how we should live.

But whether they are used with the force of communicating those semantic rules there are, or of pressing for the rules the speaker thinks there ought to be, seeing metaphysical modal claims as having a *normative* function, to do with conveying and enforcing semantic rules, promises to do a great deal to demystify the epistemology of modality—and with it, the methodology of metaphysics.

Notes

1 While there are many other sorts of modal claim (e.g. asking what is physically possible or logically possible), I will focus here on claims like these—about what is *metaphysically* possible or necessary—since those play a central role in many metaphysical debates.
2 For discussion of the grounding problem, see Burke (1992), Zimmerman (1995), Bennett (2004), and Thomasson (2007, ch. 4).
3 I do not mean to suggest that Lewis himself thought of his possible worlds as truthmakers for modal propositions, only that possible worlds realism along Lewis's lines is capable of supplying the truthmakers that the truthmaker theorist needs.
4 In the latter form this is the famous 'Humphrey' objection Kripke (1980, 45 n. 3) raises against Lewis.
5 For further critical discussion of Lewis's reply to the knowledge problem, see Bueno and Shalkowski (2004).
6 Huw Price uses the terminology 'representational' and puts the point in terms of denying that all discourse is 'e-representational'. See Price (2011) for the original formulation and criticisms of the representationalist/descriptivist assumption.
7 Price similarly notes that those he calls 'non-facutalists' (our non-descriptivists) may accept that moral claims, for example, are 'statements in some minimal sense' ('Semantic Deflationism and the Frege Point', p. 3 in pdf).
8 Though I do not mean to endorse truthmaker theory here—only to point out that, however plausible it is in descriptive cases, it leads us astray in others.
9 Schlick, in turn, was developing ideas originating in Hilbert's *Foundations of Geometry* and attempting to generalize them to the cases of logic and mathematics. See Baker (1988, 187ff).
10 A neo-conventionalist view has also been developed by Alan Sidelle (1989).
11 Barker distinguishes when adjectives have a 'descriptive' use from when they have a 'metalinguistic' use (2002, 1). The thought here is that modal terms signal that the assertion in question has a metalinguistic rather than a descriptive use.

12 See my (2020).
13 I (Thomasson 2007) and others (Devitt and Sterelny 1987) have argued for this position elsewhere.
14 See Plunkett (2015) and Thomasson (2020) for developments of the idea that certain metaphysical debates may be seen as engaged in metalinguistic negotiation.

References

Baker, Gordon. 1988. *Wittgenstein, Frege and the Vienna Circle*. Oxford: Blackwell.

Barker, Chris (2002). "The Dynamics of Vagueness". *Linguistics and Philosophy* 25: 1–36.

Bennett, Karen (2004). "Spatio-Temporal Coincidence and the Grounding Problem". *Philosophical Studies* 118: 339–371.

Blackburn, Simon (1993). *Essays in Quasi-Realism*. New York: Oxford University Press.

Brandom, Robert (2008). *Between Saying and Doing: Towards an Analytic Pragmatism*. Oxford: Oxford University Press.

Burke, Michael. (1992). "Copper Statues and Pieces of Copper: A Challenge to the Standard Account". *Analysis* 52(1): 12–17.

Devitt, Michael and Kim Sterelny (1987). *Language and Reality*. Cambridge, MA: MIT Press.

Divers, John and Joseph Melia (2002). "The Analytic Limit of Genuine Modal Realism". *Mind* 111(441): 15–36.

Forbes, Graeme (1988). "Critical Study: The Plurality of Worlds". *Philosophical Quarterly* 38(151): 222–240.

Horwich, Paul (1998). *Truth*. 2nd edn. Oxford: Oxford University Press.

Jubien, Michael (2007). "Analyzing Modality". *Oxford Studies in Metaphysics* 3: 99–139.

Kripke, Saul (1980). *Naming and Necessity*. Oxford: Blackwell.

Lewis, David K. (1986). *On the Plurality of Worlds*. Oxford: Blackwell.

Lycan, William (1988). "Review of 'On the Plurality of Worlds'". *Journal of Philosophy* 85(1): 42–47.

Mulligan, Kevin, Peter Simons, and Barry Smith (1984). "Truth-Makers". *Philosophy and Phenomenological Research* 44(3): 287–321.

Papafragou, Anna (1998). "The Acquisition of Modality: Implications for Theories of Semantic Representation". *Mind and Language* 13(3): 370–399.

Plunkett, David and Tim Sundell (2013). "Disagreement and the Semantics of Normative and Evaluative Terms". *Philosopher's Imprint* 13(23) (December): 1–37.

Price, Huw (2011). *Naturalism without Mirrors*. Oxford: Oxford University Press.

Roy, Tony (2000). "Things and *De Re* Modality". *Nous* 34: 56–84.

Ryle, Gilbert (1950/1971). "'If', 'So', and 'Because'". In *Collected Papers*, Vol. 2. London: Hutchison.

Salmon, Nathan (1988). "Review of 'On the Plurality of Worlds'". *Philosophical Review* 97(2): 237–244.

Schlick, Moritz (1918). *Allgemeine Erkenntnislehre*. Berlin: Springer.

Sellars, Wilfrid (1958). "Counterfactuals, Dispositions and the Causal Modalities". In Herbert Feigl, Michael Scriven, and Grover Maxwell, eds., *Minnesota Studies in Philosophy of Science, Vol. 2: Concepts, Theories and the Mind-Body Problem*. Minneapolis: University of Minnesota Press, 225–308.

Sidelle, Alan (1989). *Necessity, Essence and Individuation: A Defense of Conventionalism*. Ithaca: Cornell University Press.

Sosa, Ernest. 2008. "Experimental Philosophy and Philosophical Intuition". In Joshua Knobe and Shaun Nichols, eds., *Experimental Philosophy*. Oxford: Oxford University Press, 231–240.

Stairs, Allen (1988). "Review Essay: 'On the Plurality of Worlds'". *Philosophy and Phenomenological Research* 49(2): 333–352.

Thomasson, Amie L. (2007). *Ordinary Objects*. New York: Oxford University Press.

——— (2020). *Norms and Necessity*. New York: Oxford University Press.

Vetter, Barbara (2015). *Potentiality*. Oxford: Oxford University Press.

Wells, Gordon. 1985. *Language Development in the Preschool Years*. Cambridge: Cambridge University Press.

Wittgenstein, Ludwig (1922/1933). *Tractatus Logico-Philosophicus*. Translated by C. K. Ogden. London: Routledge.

Yablo, Stephen (2005). "The Myth of the Seven". In Mark Eli Kalderon, ed., *Fictionalism in Metaphysics*. Oxford: Oxford University Press.

Zimmerman, Dean (1995). "Theories of Masses and Problems of Constitution". *Philosophical Review* 104: 53–110.

PART 4

Epistemology of modality


Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Chapter 14

THE INTEGRATION CHALLENGE

Sonia Roca-Royes

14.1 Introduction

It is increasingly common to formulate and address epistemological questions in terms of the *integration challenge*. Roughly, this is the challenge of, for a given domain, providing an epistemology of how we know truths in that domain that is adequate for its metaphysics.

Peacocke characterizes the integration challenge as a problem of reconciliation:

> We have to reconcile a plausible account of what is involved in the truth of statements of a given kind with a credible account of how we can know those statements, when we do know them.
>
> (Peacocke 1999: 1)

The integration challenge as formulated by Peacocke is the generalization of what has been received as Benacerraf's dilemma (Benacerraf 1973) in the case of mathematics. In its shortest form, which I take from Hart, the dilemma can be expressed as the problem that "what is necessary for mathematical truth makes mathematical knowledge impossible" (Hart 1991: 87).

In this chapter, I will focus on the putative elements that are involved in the integration challenge, from both an inter-domain and an intra-domain perspective. The elements to be identified in what follows should not be taken as *by definition constitutive* of the integration challenge, partly because there doesn't seem to be a clear consensus (or explicitness) on what exactly the integration challenge amounts to; nor, equivalently, on what would count as having met it. Rather, the entry aims at exploring what *might* be involved in the integration challenge, and it can be taken as a call for further reflection on it.

14.2 The integration challenge and its precursor in mathematics

Formulating Benacerraf's observations in the form of a dilemma might suggest an insurmountable difficulty. According to the dilemma, it would seem, if one has a credible metaphysics in mathematics, one will lack a credible epistemology, and vice versa. Despite this, there is theoretical room to get out of the dilemma. Some options involve challenging the presuppositions in its very formulation. When assessed against Benacerraf's observations, and for the case of

mathematics, the presuppositions include that mathematical truth requires mind-independent mathematical objects that cannot but be causally inert, and that mathematical knowledge requires causal-affection from the mathematical objects.

It is common ground that those two presuppositions are jointly inconsistent with the claim that we have mathematical knowledge. Let $T^{\mathbf{math}}_{causally\ inert}$ be a theory about mathematical discourse according to which mathematical discourse is about mind-independent, causally inert mathematical objects, and let $E^{\mathbf{math}}_{causal\ affection}$ be an epistemology of mathematics according to which we acquire mathematical knowledge by being causally affected by mathematical objects. If the seeming existence of mathematical knowledge is taken at face value, the pair $\left\langle T^{\mathbf{math}}_{causally\ inert}, E^{\mathbf{math}}_{causal\ affection} \right\rangle$ is certainly untenable because it doesn't leave room for mathematical knowledge. Endorsing one such pair would amount to being committed to mathematical scepticism, and this position is widely taken as a non-starter. As Lewis once wrote, "mathematics is better known than any premise of sceptical epistemology" (Lewis 1986, 4).

Yet, this does not rule out the possibility of tenable pairs $\left\langle T^{\mathbf{math}}_{x}, E^{\mathbf{math}}_{y} \right\rangle$. In the case of mathematics, they will need to be pairs that leave room for mathematical knowledge while at the same time deny at least one of the two presuppositions in Benacerraf's dilemma. For instance, an epistemology of mathematics, $E^{\mathbf{math}}_{imagination}$, according to which we come to know those mathematical propositions whose negations we fail to be able to imagine after reflective attempts, might be coupled (plausibility considerations aside) with a view, $T^{\mathbf{math}}_{imagination}$, according to which mathematical truth is *constitutively* dependent (in the way suggested) upon our imaginative capacities. This pair would explain mathematical knowledge without much difficulty. How plausible, however, is its account of mathematical truth? Presumably not much: mathematical truth is arguably independent of our imaginative capabilities. We need to keep on searching for a suitable pair. Taking the seeming existence of mathematical knowledge at face value, the task is then to find a pair $\left\langle T^{\mathbf{math}}_{x}, E^{\mathbf{math}}_{y} \right\rangle$ whose elements are credible while not ruling out mathematical knowledge. This is, at least partly, what the integration *challenge* is about.

One might wonder how far one can generalize the challenge, and distinguishing between the *integration challenge* and what we might call 'the *integration requirement*' might help us answering this. For any given domain, φ, such that there are (at least pre-theoretically) truth-apt φ-claims, the claim that there is an integration *challenge* for φ suggests at the very least that there is an integration *requirement*. The requirement can be roughly characterized, as Peacocke has in mind in the aforementioned quotation, as that of providing a credible epistemology of φ-truths that makes justice to the kind of facts φ-truths are taken to be about. But the *requirement* is not all that is involved in the *challenge* claim. We should not rule out in advance that, for some domains, satisfying the requirement might be a relatively easy task. For those domains—if there are any—no integration *challenge* would arise. To a first approximation—to be made more informative later in this chapter—therefore, we can say that there is an integration *challenge* in those domains where satisfying the integration requirement is not an easy task. This is—as Benacerraf noted and is widely accepted—the case in mathematics. Other domains for which the existence of an integration challenge has been explicitly acknowledged include ethics, logic, modality, morality, the past, self-knowledge, politics and freedom. And the list can surely continue.

14.3 The integration challenge in modality

I shall here briefly survey the case of *modality*—the topic of this *Handbook*—for which the integration challenge has been acknowledged to be very vivid. David Lewis (1986) has offered a reductive account of modal truth in terms of facts at possible worlds[1] that, Lewis forcefully

defends, is explanatorily superior to any alternative. Yet, those worlds are (like mathematical objects) causally isolated from us. The objection here is that, despite the merits of Lewis's reductive modal metaphysics, it can hardly be coupled with an epistemology of modality *with which to meet the integration challenge satisfactorily* because it makes modal knowledge mysterious.[2] For many, modal knowledge is—like mathematical knowledge—non-negotiable and, as a result, it requires an explanation. No resulting pair, $\left\langle T^{\textbf{modality}}_{\textit{Lewisian worlds}}, E^{\textbf{modality}}_{???} \right\rangle$ has been deemed fit to meet the integration challenge.[3]

In an attempt to meet the integration challenge in modality, Peacocke (1999) has reconceived the metaphysics of modality in terms of *facts according to his ersatz possible worlds*, where the set of those worlds is determined by *what* our modal concepts are. This metaphysics of modality has been designed so as to weaken Lewis's claim that his reductive account is explanatorily superior to any alternative. In turn, such metaphysics is coupled with an epistemology of modality—in terms of reasoning from the implicit conceptions associated to our modal concepts—that goes very naturally with the suggested metaphysics. The resulting pair $\left\langle T^{\textbf{modality}}_{\textit{Peacockian worlds}}, E^{\textbf{modality}}_{\textit{implicit conceptions}} \right\rangle$ does leave room for, and explains, modal knowledge. A salient problem of the account, however, is that the account doesn't guarantee that its subject matter is the pre-theoretical one: namely, metaphysical modality, and so the metaphysical re-conception might have gone a bit too far.[4]

Also as an explicit reaction to the bad reputation of Lewis's account—vis-à-vis the goal of meeting the integration challenge—Hale (2013) offers, in what he calls "the essentialist theory of metaphysical necessity", an account of metaphysical modal truth as grounded in essence. Compared to that of Peacocke's, the account is much less at risk of suffering from a change of subject matter.[5] The challenge, then, is to leave room for modal knowledge and to provide an epistemology of modality that goes well with this metaphysics. Hale aims to meet this challenge by making modal knowledge dependent on knowledge of essence, resulting in a pair $\left\langle T^{\textbf{modality}}_{\textit{essence-grounded}}, E^{\textbf{modality}}_{\textit{essence-based}} \right\rangle$ whose satisfactoriness deserves much scrutiny. As he acknowledges:

> It may well be, and I suspect it often is, thought that it is equally mysterious how we might have knowledge of essence, so that the essentialist theory of necessity is equally vulnerable to the charge of epistemological bankruptcy. I shall try to show that this charge is unjustified.
>
> (Hale 2013, 252)

Historically, other theorists have received the pressure of the challenge in modality as recommending a more radical revision on the pre-theoretical subject matter. As Peacocke (1999) reviews, both Craig (1985) and Blackburn (1993) take modal truth to be mind-dependent because constitutively dependent (to different extents) on our imaginative/conceivability capabilities. Assuming—as seems reasonable on the intended notion of conceivability—that our epistemic access to conceivability facts is transparent, such metaphysical accounts have a readily available epistemology for their modal truths so understood, the resulting pair looking like $\left\langle T^{\textbf{modality}}_{\textit{conceivability-constituted}}, E^{\textbf{modality}}_{\textit{conceivability}} \right\rangle$. As with Peacocke's account—and even more straightforwardly so—the problem with their accounts is "whether it will deliver not necessity, but only some notion of a distinctive kind of truth in the actual world" (Peacocke 1999, 178).

Van Inwagen (1998), to give a further example, combines a mind-independent metaphysics of modality with (partial) modal scepticism. His modal scepticism rests on the fact that conceivability, while having "some very attractive features, and [being] certainly more sophisticated than any other account of modal knowledge" (1998, 76), is an inert route to modal knowledge because the relevant conceivings—in the way van Inwagen understands them—are beyond our

reach: not that conceiving facts are not transparent to us, but, rather, that we simply fail to be able to *conceive* in the sense that would give us access to possibility.

The list of examples presented in this section is far from exhaustive, but it is illustrative of the paths one can take with the aim of meeting the integration challenge in modality. The next section articulates more explicitly, and at a general level, the different paths one can take to meet the integration challenge.

14.4 Ways to meet the integration challenge

For any domain, φ, for which the integration challenge arises, Peacocke (1999, 7–11) identifies seven ways to meet it, which he classifies thus:

Non-revisionary

(1) Reconceiving the metaphysics
(2) Reconceiving the epistemology
(3) Reconceiving the relation between the metaphysics and the epistemology

Revisionary

(4) Slimming-down the truth-conditions
(5) Non-factualism
(6) Scepticism
(7) Neglecting the problem by neglecting the meaningfulness of domain at issue in the first place

By having a look at the strategies that Peacocke considers to be revisionary, one realises that this classification is a matter of what Peacocke thinks one should be aiming at, given (alleged) antecedent beliefs to the effect that the φ-discourse is meaningful, that we have φ-knowledge and that φ-truths are objective. There is plenty of scope for discussion already at this point. One can look for instance into the exhaustiveness of Peacocke's seven categories, into the adequacy of how he classifies them into revisionary/non-revisionary, or into the presumed absoluteness of the classification—as opposed to its being domain-specific (or even domain-specific *and* theorist-relative). Might it not be, for instance, that what is to count as revisionary in the case of modality—for instance, option (5)—need not be so revisionary when it comes to other domains—for instance, aesthetics?

One might also worry that if those seven ways are effective ways of *meeting* the challenge, then there is little room—if any—for failing to meet it; in which case 'challenge' would be a misnomer. Yet, to alleviate this worry, we should receive the very idea of classifying the options into revisionary and non-revisionary as reflecting the assumption that some ways of approaching it are more desirable—at least pre-theoretically—than others. For instance, meeting the challenge in modality by means of Peacocke's strategy (1) might indeed be more satisfactory than meeting it by means of Blackburn's strategy (4) or by means of (6).

The absoluteness of the classification is worthy of attention, and I shall leave it to the reader to explore the issue further. For now, it suffices to note that, when it comes to identifying the elements potentially involved in the integration requirement, the existence of different paths invite reflection on the following *inter-domain* question. Should we aim at meeting the integration requirement in analogous ways for all domains? The suggestion that the classification might not be adequate if intended as *absolute* easily amounts to the suggestion that different domains

might call for different desirable ways to meet the challenge. Indeed, a case can be made that we should not aim at uniformity at this level. Suppose that the challenge for the case of mathematics can be met in a non-revisionary way, and let $\left\langle T^{\textbf{math}}_{\textit{causally inert}}, E^{\textbf{math}}_{\textit{proofs}} \right\rangle$ be a pair that retains (the pre-theoretical?) Platonism about mathematical discourse together with the claim that mathematical truths are known a priori by (deductive) proofs from some basic axioms. Presumably, we don't want our endorsement of $\left\langle T^{\textbf{math}}_{\textit{causally inert}}, E^{\textbf{math}}_{\textit{proofs}} \right\rangle$ to put pressure on what we think are good (i.e., challenge-meeting) pairs in other domains. We seem to have knowledge of the material world around us. Taking this at face value, a pair $\left\langle T^{\textbf{material world}}_{\textit{causally powerful}}, E^{\textbf{material world}}_{\textit{causal affection}} \right\rangle$ according to which material world propositions are about mind-independent and causally powerful objects knowledge of which is partly grounded on their causal affection on us is a good pair for the domain at issue. Arguably, in addition, this pair is more credible—so let us grant—than alternative pairs that would take, for instance, knowledge of the material world to be independent of causal affection, or that would drastically revise the metaphysics of the (alleged) external world.

To intend that our best account of knowledge of, and truths about, the material world remains unthreatened by our best account of knowledge of, and truths about, mathematics—no matter what that might turn out to be—amounts, therefore, to not having inter-domain uniformity aspirations, at least at this level; a qualification will be suggested in Section 14.5.1. Indeed, Peacocke's *Being Known* suggests different strategies for the different domains it deals with.

This is to say that some degree of independent variation across domains on what we take to be credible good pairs should be tolerated. It is not to say, however, that any degree of independence is tolerable. The cases of *modality* and *logic* offer a vivid illustration of why this might not be so. As Wright (1986) emphasizes, if *logical truths* are *necessary truths of a certain form*, then logical truths are objective only if (the relevant) modal discourse is so as well. The identity also requires the epistemology of modality and the epistemology of logic not to be (completely) independent of one another. In addition, it might be defended that, if modal knowledge is best seen as arrived at deductively from (epistemically) basic necessities,[6] acquisition of non-basic modal knowledge will be blocked unless some (the relevant) logical truths are knowable. This latter point is generalizable to any domain, so long as we concede that logical inferences are *fundamental* ways of expanding our knowledge. We should not, therefore, rule out in advance the existence of pairs of domains, φ and ψ, both of which at least partly epistemically accessible, such that what we take to be a good pair $\left\langle T^{\varphi}_{x}, E^{\varphi}_{\gamma} \right\rangle$ might (perhaps mutually) depend on what we take to be a good pair $\left\langle T^{\psi}_{x}, E^{\psi}_{\gamma} \right\rangle$.

14.5 Intra-domain components of the integration challenge

The remaining considerations are *intra-domain*. By means of these, we should have material for reflection on what exactly meeting the integration challenge might be taken to involve, despite the fact that, as Schechter (2010) notes, there is lack of clarity on this. What we want, at a minimum, is some guidance on how to assess a (presumed) pair $\left\langle T^{\varphi}_{x}, E^{\varphi}_{\gamma} \right\rangle$ vis-à-vis the demands of the integration requirement. The following three components are submitted as each potentially involved in the integration requirement. The main aim of this section is to prompt the reader's own assessment on them.

14.5.1 We need a plausible pair

According to the quotation by Peacocke from Section 14.1, merely getting a (or *the*) correct pair is not all that is involved in meeting the integration challenge. There, Peacocke already asks for more: namely, each of the members in the integrated pair must be *plausible* (or credible).

Let us focus on logic as our running example. Suppose that we know that logical truths are objective and that we know that Modus Ponens (henceforth MPP) is valid. Suppose further that we answer the question about how we know the validity of MPP (and other basic logical rules) along the lines of Boghossian's *blind reasoning* approach.[7] That is, by saying that we know MPP by deductively arriving at 'If P, then, if (if P then Q), then Q' from the discharged premises 'P' and 'if P then Q', and that this deductively valid reasoning is justificatory despite being blind, and despite being rule-circular. Suppose further that this answer is accompanied by what is in fact a correct account of how blind, rule-circular reasoning can be justificatory. There might be no trace of falsity in any of our beliefs about the matter. We seem, therefore, to be in front of what is in fact a correct and stable pair $\left\langle T^{\mathbf{logic}}_{objective}, E^{\mathbf{logic}}_{blind\ reasoning} \right\rangle$. Yet, that's presumably not all that is involved in meeting the integration challenge. As just seen, Peacocke suggests, as a component of the integration requirement, that each of the elements in the stable pair be plausible. Depending on the domain at issue, envisioning a plausible metaphysics or envisioning a plausible epistemology might be quite a challenging task in itself. And, typically, areas for which the *metaphysical challenge* is pressing will be areas for which the *epistemic challenge* is pressing, too, let alone the *integration challenge*. This leaves us with the need of trying to articulate what exactly the plausibility demand amounts to. At a minimum, it would seem to require that we have *reasons* for them, but it would be good to be in a position to provide something more substantial that would also help us as a guide.

Schechter (2010) and Wright (2004) both have offered insights about what is involved in the *epistemic challenge* for logic that will prove very helpful to answer our current question—about how best to understand the plausibility demand. Ultimately, we are looking for an answer that applies also to the *metaphysical challenge*. Yet, in what follows, I shall focus on Schechter's and Wright's insights on the epistemic one and leave it to the reader to explore how they might transfer to the metaphysical challenge.

Schechter and Wright agree that merely coming up with an answer—even a correct one—to how we know the truths in a given domain would not suffice. Yet, they formulate different further components. According to Schechter, we need further to address the *Etiological Question*, whereas, according to Wright, what we need to address further is a *Second Order Need*. These are Schechter's two components—of which the first one amounts to correctness:

> *The Operational Question*: How does our cognitive mechanism for deductive inference work such that it is reliable?
>
> *The Etiological Question:* How is it that we have a cognitive mechanism for deductive inference that is reliable?
>
> (Schechter 2010: 444)

And these are Wright's—of which, as with Schechter's, the first one parallels correctness:

> The first-order problem is: if we do indeed have the knowledge of the validity of basic logical laws that we think we do, how might this knowledge be achieved? . . . the second-order problem: that of explaining with what right we claim it.
>
> (Wright 2004: 174)

Generalizing from the case of logic—nothing suggests that the two components are peculiar to its case—I will follow Wright in characterizing the *epistemic challenge* in terms of a first- and a second-order demand. The first one amounts to the requirement that we provide an answer to the question 'How do we know?' and, the second, to the requirement that we provide a

justification for our claim that we are in possession of knowledge. The fact that I follow Wright on this is not to neglect the relevance of Schechter's *Etiological Question*. On the contrary, I think that this question is closely related to Wright's second-order requirement in that a satisfactory answer to Schechter's question might (in certain cases) be what constitutes a satisfactory answer to the second-order demand.[8] I think, however, that the (generalized) Etiological Question suffers from lack of neutrality in a way that Wright's (generalized) Second Order Problem does not. For, on the one hand, it is not to be assumed that the Etiological Question—especially when we are asked to understand it phylogenetically, as Schechter does (2010, 459 n. 29)—will be pertinent in all cases; it will not, for instance, when non-objectivism is the view about the nature of φ-truths (for a given domain).[9] On the other hand, it seems to me that Wright's second-order demand is no less a reasonable demand in the cases where the Etiological Question is not pertinent than it is in the cases where the Etiological Question is pertinent.

Now, Peacocke's own requirement that an epistemology E_x^{φ} be credible can be naturally interpreted along the lines of Wright's second-order component. For, if we are in a position to claim that E_x^{φ} is the correct *epistemology*—as opposed, crucially, to its being the correct *doxastology*[10]—our reasons for this will likely have to involve reasons for thinking that the methods mentioned in E_x^{φ} are good methods for coming to *know* the φ-propositions delivered by the method. And whatever those reasons are, they must arguably include reasons for claiming knowledge of those delivered propositions. There are, in effect, two ways of phrasing the second-order element: as the requirement to justify our claims to knowledge, and as the requirement to justify the proposed epistemology as an *epistemology*. (The reader might at this point go back to the *metaphysical challenge* and try to determine how the second-order component would look like in this case.)

To conclude this section, we might now want to qualify our *inter-domain* claim in Section 14.4—namely, that we need not (should not?) aim at a uniform epistemology across domains. The qualification takes into account the first- and second-order demands and reduces the scope of the claim to only the first-order demand. In other words, the claim is that we need not aim at meeting the first-order demand uniformly. Indeed, the reasons there were neutral as to whether we should aim at meeting the second-order demand uniformly or not. With a view to merely providing material for further exploration, I shall note that the cross-domain applicability of Wright's notion of *entitlement to cognitive project*—applied, partly, precisely to vindicate our knowledge-claims in a given domain—might fuel the belief that, at the second level, a uniform treatment might be available. Also, the cross-method applicability of Enoch and Schechter's (2006, 2008) *pragmatic-based account of epistemic justification of belief forming methods*—in some respects congenial to Wright's account—might fuel that belief too.

14.5.2 The strength of 'integration'

Assuming that we are aiming for a pair that leaves room for φ-knowledge—that is, assuming we are aiming to escape the most revisionary strategies (6) and (7)—how are we to understand, in this context, what we mean by 'integration' itself? Peacocke asks not just that the epistemology E_y^{φ} of a given pair $\langle T_x^{\varphi}, E_y^{\varphi} \rangle$ be credible, but also that it be credible *as an epistemology of the kind of truths* T_x^{φ} *commits us to*. But there is a weak way and a strong way of understanding this requirement. Roughly, the weak reading would simply amount to a compatibility requirement: that the relevant T and E are compatible with the claim that we do have knowledge of the T-facts acquired in the E-way. On this reading, the integration requirement would be satisfied by pairs $\langle T_x^{\varphi}, E_y^{\varphi} \rangle$ even in the case that E_y^{φ} did not rule out substantially different and yet competing alternatives to T_x^{φ}. By contrast, on a stronger reading, the two elements would need to roughly call for one another. An example here would be again the pair $\langle T^{\textbf{material world}}_{\textit{causally powerful}}, E^{\textbf{material world}}_{\textit{causal affection}} \rangle$

. For $E^{\textbf{material world}}_{\textit{causal affection}}$ says that we acquire knowledge of the material world by being causally affected by the objects, and this *requires* a metaphysics with causally powerful objects. Another example—which adds to the list in Section 14.3—would be $\left\langle T^{\textbf{modality}}_{\textit{conventionalism}}, E^{\textbf{modality}}_{\textit{conceptual analysis}} \right\rangle$. According to it, modal truth is a matter of linguistic conventions.[11] It is quite natural to couple this metaphysics with an epistemology according to which we access modal truth by conceptual analysis. Compared to the material world example, however, this pair, despite integrated, suffers from a controversial metaphysics, risking the satisfaction of the requirement in Section 14.5.1.

I shall not firmly answer the question whether we should understand 'integration' in the strong or the weak sense. Yet, I want to offer some reflections that might help us diagnose the current state of certain domains and that might suggest that we are working with integration in the strong sense as an ideal. The material world offers a very well-understood model: we have a dominant type of metaphysics and a dominant type of epistemology, and they seem to be integrated in the strong sense. To some extent, finding pairs in other domains with somewhat credible elements is not particularly difficult. For instance, a pair $\left\langle T^{\textbf{math}}_{\textit{Platonism}}, E^{\textbf{math}}_{\textit{proofs}} \right\rangle$ according to which mathematical truth involves mind-independent mathematical objects coupled with an epistemology of mathematics that has *proofs* at its centre (with a suitable account of epistemic access to axioms) is arguably a pair compatible with knowledge, and its two components are credible even if not free of controversy. The problem is that so is a pair with the same sort of epistemology but according to which mathematical discourse is to be given instead a fictionalist interpretation: $\left\langle T^{\textbf{math}}_{\textit{fiction}}, E^{\textbf{math}}_{\textit{proofs}} \right\rangle$. The Platonists might find $T^{\textbf{math}}_{\textit{Platonism}}$ credible, the fictionalists might find $T^{\textbf{math}}_{\textit{fiction}}$ credible and everyone in this debate might find $E^{\textbf{math}}_{\textit{proofs}}$ credible. But however credible $E^{\textbf{math}}_{\textit{proofs}}$ is, the existence of those two pairs, both equally compatible with mathematical knowledge, suggests that $E^{\textbf{math}}_{\textit{proofs}}$ falls short of being, in a strong sense, an epistemology *of the kind of truths* $T^{\textbf{math}}_{\textit{fiction}}$ *or* $T^{\textbf{math}}_{\textit{objective}}$ *commits us to*. Rather, it is an epistemology of mathematical truth *however truth is to be understood*. For reflection: does this suggest that none of these two pairs are integrated enough to meet the integration requirement satisfactorily? If so, and given how dominant an epistemology along the lines of $E^{\textbf{math}}_{\textit{proofs}}$ is, what are the prospects of meeting the integration challenge in mathematics (or any other domain that shares with mathematics the existence of analogous pairs of pairs)? Will it all boil down to the story given about our epistemic access to the axioms?

Modality offers again another good example. Conceivability accounts claim, roughly, that modal truth is epistemically accessible by means of conceivability exercises. We know—the thought goes—that there could be seven apples in my fridge by conceiving seven apples in my fridge. But when a conceivability epistemology is coupled with a metaphysics of modality according to which modal truth is mind-independent, the resulting pair, $\left\langle T^{\textbf{modality}}_{\textit{mind-independent}}, E^{\textbf{modality}}_{\textit{conceivability}} \right\rangle$, can be said to suffer from lack of integration. Why should there be such (lucky) alignment between our mental (conceiving or imaginative) capabilities and the mind-independent modal realm? An alternative pair, $\left\langle T^{\textbf{modality}}_{\textit{conceivability-constituted}}, E^{\textbf{modality}}_{\textit{conceivability}} \right\rangle$, along the lines of Craig's or Blackburn's suggested accounts, according to which modal truth is constitutively dependent on our imaginative capabilities, is arguably more integrated; although we have seen that it suffers from lack of metaphysical plausibility. Are the domain for which the integration challenge is pressing domains that face difficulties precisely to meet *both* plausibility and integration simultaneously? This is another issue that calls for further reflection.

14.5.3 Integration and disagreement

Questions about disagreement in relation to the integration challenge emerge at two levels. As with most of the issues previously mentioned, I shall content myself with raising two of the questions I find most urgent; and I shall in passing provide some material to guide the reader's own assessment. First, suppose that, for a given domain φ, there is a correct integrated pair

$\left\langle T_x^{\varphi}, E_y^{\varphi} \right\rangle$. How much agreement that this is a correct, integrated pair should we require before we can claim that the integration challenge for φ has been met? Areas for which the integration challenge is pressing are areas notably infected with widespread disagreement both about what the metaphysics of the domain is and about what its epistemology is; let alone about which pair of these is overall more plausible and integrated. Are the prospects of meeting the integration challenge in these domains poor for as long as we don't solve those disagreements?

Second, should we require of our having met the integration challenge for φ that the *application* of the methods mentioned in E_y^{φ} resolves disagreements about what the φ-truths are? Let me note that this is not the unreasonable demand that E_y^{φ} should in itself tell us what the φ-truths are. Take again $\left\langle T^{\textbf{material world}}_{\textit{causally powerful}}, E^{\textbf{material world}}_{\textit{causal affection}} \right\rangle$. $E^{\textbf{material world}}_{\textit{causal affection}}$, by itself, doesn't tell us what the truths about the material world are. It doesn't tell us, for instance, that there are no apples in my fridge now. But if endorsement of $\left\langle T^{\textbf{material world}}_{\textit{causally powerful}}, E^{\textbf{material world}}_{\textit{causal affection}} \right\rangle$ were judged not to be satisfactory to meet the integration challenge, it would not be for this reason. What $E^{\textbf{material world}}_{\textit{causal affection}}$ does have the power to do is the following: if we engage in the project of finding out how many apples there are in my fridge right now, *application* of the methods mentioned in $E^{\textbf{material world}}_{\textit{causal affection}}$ will leave little room for disagreement (among those subjects whose perceptual systems are working properly). We open the fridge's door, apply the methods (i.e. *look*) and settle the issue. Might it be that when $\left\langle T^{\textbf{material world}}_{\textit{causally powerful}}, E^{\textbf{material world}}_{\textit{causal affection}} \right\rangle$ is judged to be an integrated pair, it is partly due to how little room for disagreement $E^{\textbf{material world}}_{\textit{causal affection}}$ leaves? Again, another issue, further reflection on which might help us measure the right scope of the integration requirement.

14.6 Conclusion

This chapter has been devoted to characterizing the *integration challenge* in general terms with a view to exploring, and prompting reflection on, what might be involved in it. By way of illustration, I have reviewed some ways in which the challenge has been approached in the case of modality (Section 14.3). For most of the chapter, however, the discussion has been conducted at a rather general level for, as Peacocke made explicit, the reasons behind Benacerraf's dilemma in the case of mathematics are generalizable to any domain.

The suggestion of this chapter is that there's reason to think that how best to meet the epistemic challenge is domain-relative. Granting that a given domain, φ, is meaningful and φ-knowledge possible (even if perhaps only partially), what is common to the challenge in all domains is the need of finding plausible metaphysics and plausible epistemologies that are *integrated*, at the very least, in the (weak) sense of not making knowledge impossible. It might be that we require more of integration, along the lines of the dependence exhibited by $\left\langle T^{\textbf{material world}}_{\textit{causally powerful}}, E^{\textbf{material world}}_{\textit{causal affection}} \right\rangle$. It might also be that we require little disagreement about what is a plausible metaphysics and a plausible epistemology. And it might also be that we require that the methods mentioned in the suggested (plausible) epistemology leave little room for disagreement about what the truths in that domain are. More reflection is needed about what exactly the scope of the integration challenge is. The more we require, the harder it will be for it to ever be met. Yet, independently of how we fix the semantics of the 'integration challenge', the issues presented here will likely continue to bother us, whether as part of or aside of the integration challenge.

Notes

1 On this account, necessities reduce to truth at all possible worlds and possibilities, to truth at some world (Lewis 1986).

2 For expressions of this complaint, see for instance Hale (2013, 252) and Peacocke (1999, 3–4).

3 Although Lewis did address the epistemological challenge (Lewis 1986, §2.4).
4 See for instance Wright (2002) and Roca-Royes (2010).
5 There are *independent* reasons—coming from Kit Fine's *Essence and Modality* (1994) and on which Hale rests—that suggest that metaphysical necessity is so grounded.
6 See a review in Vaidya (2015).
7 See Boghossian (2003).
8 This might be especially plausible in the case of an evolutionary-based answer to the phylogenetic understanding of the Etiological Question—of the kind Schechter envisions for the case of objective logical truth. (Provided, however, that one can plausibly link the usefulness of a given cognitive mechanism and its being truth-tracking; perhaps by claiming that the latter is the best explanation of the former.)
9 This much seems to be acknowledged by Schechter in a footnote where he says, when commenting on how to answer the Etiological Question in the case of non-objective domains, that: "a different potential answer to the etiological question is that the facts of the domain are constituted by our opinions, and so we would have possessed a reliable cognitive mechanism no matter what" (Schechter 2010, 459 n. 29).
10 By 'doxastology', as an enquiry about *us*, I mean a mere description of our belief-forming methods. An epistemology, as an enquiry about us, takes on board, at a minimum, the presumption of knowledge in the question 'How do we know?'
11 See Sidelle (1989) for an example.

References

Benacerraf, Paul (1973). 'Mathematical Truth'. *The Journal of Philosophy*, 70/19: 661–679.

Blackburn, Simon (1993). 'Morals and Modals'. Repr. in his *Essays in Quasi-Realism*. New York: Oxford University Press.

Boghossian, Paul (2003). 'Blind Reasoning'. *Proceedings of the Aristotelian Society, Supplementary Volume*, 77: 225–248.

Craig, Edward (1985). 'Arithmetic and Fact'. In I. Hacking (ed.), *Exercises in Analysis*. Cambridge: Cambridge University Press.

Enoch, David and Schechter, Joshua (2006). 'Meaning and Justification: The Case of Modus Ponens'. *Noûs*, 40/4: 687–715.

——— (2008). 'How Are Belief-Forming Methods Justified?' *Philosophy and Phenomenological Research*, 76/3: 574–579.

Fine, Kit (1994). 'Essence and Modality'. *Philosophical Perspectives*, 8, Logic and Language: 1–16.

Hale, Bob (2013). *Necessary Beings: An Essay on Ontology, Modality, and the Relations between Them*. Oxford: Oxford University Press.

Hart, W.D. (1991). 'Benacerraf's Dilemma'. *Crítica, Revista Hispanoamericana de Filosofía*, 23/68: 87–103.

van Inwagen, Peter (1998). 'Modal Epistemology'. *Philosophical Studies* 92: 67–84.

Lewis, David (1986). *On the Plurality of Worlds*. Oxford: Blackwell.

Peacocke, Christopher (1999). *Being Known*. Oxford: Oxford University Press.

Roca-Royes, Sonia (2010). 'Modal Epistemology, Modal Concepts, and the Integration Challenge'. *Dialectica*, 64/3: 335–361.

Schechter, Joshua (2010). 'The Reliability Challenge and the Epistemology of Logic'. *Philosophical Perspectives*, 24, Epistemology: 437–464.

Sidelle, Alan (1989). *Necessity, Essence, and Individuation: A Defense of Conventionalism*. Ithaca, NY: Cornell University Press.

Vaidya, A. (2015). 'The Epistemology of Modality'. In Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2011 edition). Available at: http://plato.stanford.edu/archives/win2011/entries/modality-epistemology/.

Wright, Crispin (1986). 'Inventing Logical Necessity'. In Jeremy Butterfield (ed.) *Language, Mind and Logic*, Cambridge: Cambridge University Press: 187–209.

——— (2002). 'On Knowing What Is Necessary: Three Limitations of Peacocke's Account'. *Philosophy and Phenomenological Research*, 64/3: 655–662.

——— (2004). 'Intuition, Entitlement, and the Epistemology of Logical Laws'. *Dialectica*, 58/1: 155–175.

Chapter 15

THE EPISTEMIC IDLENESS OF CONCEIVABILITY

M. Oreste Fiocco

The world comprises all things. 'Thing' here should be understood with the utmost generality, referring to any existent in any category. Some claims about these things are *necessary* simpliciter, in that what they represent *must*—in the broadest sense—be true. Some claims are *possible* simpliciter, in that what they represent *might* or *could*—again, in the broadest sense—be true. Thus, what is necessary indicates the principles of the world, and what is possible indicates its limits. Call knowledge of what is necessary or possible simpliciter *modal knowledge*.[1] Such knowledge may also include what knowledge of necessity or possibility provides. Knowing $\sim p$ is necessary yields knowledge that p is *impossible* (i.e., could not be true). Knowing $\sim p$ is possible, though what p represents is actually so, yields knowledge of the *contingency* of p.

If one knows anything, one is supposed to be in the position to have at least some modal knowledge. If one knows p, it is uncontroversial that one is able to infer justifiably and know that what p represents is possible. Some knowledge of necessity and impossibility is, with minimal logical ability or conceptual capacities, likewise uncontroversial; for example, it must be that if p *and* q is true, p is as well, and it is impossible that there is a round square. Plausibly, one has more modal knowledge than these apparently trivial examples. I can know, it seems, that the stroller before me could fit in the back of my car, that I could not ride a bicycle up my own nose, that water must be H_2O. Many claim to have even more modal knowledge than such mundane examples. Some claim to know, for instance, that there could be a perfect being, that one could exist without one's body, that there could be a physical duplicate of a (conscious) person that lacks consciousness, that one must originate from the sperm and egg from which one actually did. Such esoteric modal claims are often central to arguments taken to show what in fact exists and what or how certain phenomena actually are, and so are central to many philosophical issues of perennial interest.

One's involvement with the world seems to be limited merely to things as they are; hence, modal knowledge—trivial, mundane, or esoteric—should be perplexing. A thoroughgoing epistemology should, however, account for all of it. Traditionally, the notion of *conceivability* has been regarded as crucial to an account of modal knowledge. The conceivability of, say, a proposition is supposed to provide at least some evidence that what that proposition represents is possible. This idea is prominent in seminal writings of Anselm and Descartes, and Hume later articulated a well-known explicit connection between what is conceivable and what is possible.[2] Conceivability is regarded as no less crucial in contemporary discussions of modal epistemology.

Indeed, Stephen Yablo remarks, "if there is a seriously alternative basis for possibility theses [i.e., claims about what is possible] philosophers have not discovered it."[3] Despite the assumed significance of conceivability to acquiring modal knowledge, there is no received account of what exactly it is to conceive a proposition.

I believe one has a good deal of modal knowledge, though perhaps less than others presume one has. However, I think conceiving is utterly idle in acquiring such knowledge: the conceivability of a proposition can provide no evidence whatsoever that what it represents is possible. To show this, I first examine the basis of modal knowledge. I consider what conceivability in general is supposed to be and argue, in light of the preceding considerations, that conceivability is not epistemically efficacious, in the sense of providing evidence, on any proposed specific account. I then maintain that there could be no account of conceivability on which it is epistemically efficacious, that the very idea of the conceivability of a proposition being evidential is misguided. I conclude with some brief recommendations for pursuing a satisfactory modal epistemology.

15.1 The world and the ontological basis of modal knowledge

Again, modal knowledge is knowledge of what is necessary or possible *simpliciter*. Clearly, such knowledge pertains to the world—all knowledge does—but if the world is just all the things there are, it is by no means obvious how one can acquire knowledge of what *must be* or of what *could be* from what merely *is*. Without some account of how necessity and possibility inhere in or arise from the things there are, there seems little hope of illuminating modal epistemology, in general, or assessing the role of conceivability in acquiring modal knowledge, more specifically. An account of the ontological basis of modal knowledge would provide, then, some insight into with what one needs to engage in order to have such knowledge and how accessible it is.

Given that the world is nothing more nor less than all things (including the relations in which things stand), modal knowledge can be understood to be knowledge of the *constraints* among things. Knowledge of necessity is knowing that a thing must be as it is (at least in some respects) or that certain things must relate as they do. Knowledge of possibility is knowing that something could be different than how it actually is or that it could relate to others in ways it does not. It might also be knowledge that some of the things that exist could fail to, or even that there could be different things than there in fact are.

If modal knowledge is knowing such constraints, a satisfactory modal epistemology turns on them, on their source(s) and means. There can be, it seems, just two alternative accounts of these constraints. On one, constraints are present in each thing itself: a thing must be as it is (at least in some respects) and could be certain other ways simply in existing, in being what it is. On the other account, constraints arise only through the interaction between some privileged kind(s) of *constraining* thing and other *unconstrained* things. The most familiar version of the latter is one on which minds constrain the other things in the world via their activity. (A variation of this, also familiar, is an account on which a single divine mind constrains all other things.) Note that any mixed view, on which things constrained in themselves are also constrained by minds, is irrelevant to present purposes. Of interest here is necessity and possibility *simpliciter*; the constraints imposed by minds on inherently constrained things would yield only a secondary necessity and possibility.

The first, robustly realist account, on which things are necessarily or possibly as they are independently of any conscious being, has been present since the outset of Western philosophy and has long been associated with Aristotle. The second account has had many guises in the history of Western philosophy. In the modern era, it is familiar from Hume, who maintains that necessary connections are projected onto things by one's expectations acquired through

experience, and from Kant, who can be understood as maintaining that one's mind supplies the constraints on the world that make any of it knowable. These positions underlie many in the analytic tradition from those of W.V. Quine (1953: 22; 1951: 176) and Rudolf Carnap (1947), who deny that things are necessarily or possibly ways independently of how they are described or referred to, to contemporary conventionalist views, like those of Alan Sidelle (1989) and Amie Thomasson (2007), on which the basis of necessity and possibility is the means conscious beings have of classifying things.

If one is to have modal knowledge, one must ultimately engage with different things depending on which of these two accounts of the source of the constraints among things is correct. Either one must engage with things that are in themselves necessarily or possibly certain ways, or one must engage with the means that conscious beings use to describe or refer to or classify things. The latter seem accessible via reflection. If, however, things themselves are the source of the constraints in the world, mere reflection obviously would not suffice for modal knowledge.

In light of these considerations, if conceiving a proposition is presumed to be instrumental to modal knowledge, the account on which the ontological basis of such knowledge is the activity of minds seems more promising, for whatever it is precisely to conceive something, doing so is a mental activity. This account, however, is in the end untenable; the ontology it requires is incoherent. This bold claim requires more thorough justification than can be provided here, yet two considerations can be adduced that should render it quite compelling.

First of all, if all the constraints that yield necessity and possibility *simpliciter* arise from the activity of minds, then the constraints on any mind itself, and there surely are some, must arise from the activity of a mind. This seems ultimately to require that some mind constrain itself. Yet anything that constrains must be sufficiently determinate—that is, constrained—to act at all and, a fortiori, to act in a way that is constraining. An unconstrained thing coming to constrain itself is incoherent; therefore, this account, which requires such a thing, is inconsistent. So there must be some constraints among things independent of the activity of any mind (or any other privileged constraining kind).

A second, independent consideration corroborates the inconsistency of the account in question. On this account, minds are supposed to impose constraints on a world lacking any. This world might be characterized as amorphous *stuff*, or as a teeming hodgepodge of unconstrained "things". Yet a truly amorphous stuff or an utterly unconstrained thing is incoherent. Such stuff, or such a "thing", is not any way necessarily, so it need not be unconstrained. It could, then, be constrained: it could be an unconstrained constrained "thing". For that matter, since what is possible for an utterly unconstrained "thing" is posterior to a constraint imposed upon it (in this case, by some mind), prior to such imposition, that "thing" could not be any way, not even unconstrained. Thus, prior to being constrained, it would be an unconstrained "thing" that is not unconstrained. This is contradictory.

I conclude that any account on which the constraints that yield necessity and possibility *simpliciter* are not inherent to each thing itself is unacceptable. Thus, there are constraints in the world independent of any mind (except, of course, those constraints inherent to minds). If one is to have modal knowledge, it must come, in the first instance, by engaging with these constraints and, hence, theirs sources, viz., things themselves. Nothing else would suffice.

15.2 What, in general, conceivability is supposed to be

As noted in the introductory section, the conceivability of propositions has long been regarded as crucial to discussions of modal epistemology. Although there is a good deal of controversy regarding what the apt, specific account of conceiving on which the act is epistemically

efficacious is—and whether there is such an account at all—what conceiving in general is supposed to be is clear enough from its many appearances in various philosophical contexts over the centuries and the multitude of examples presented therein.

To conceive a proposition, *p*, requires one to at least perform a mental act that results in one being in a mental state with content *p*. Conceiving *p* is, then, an essentially representational act, one that does not require one to have ever perceived or have had immediate cognitive access to all the things represented by that proposition. Moreover, conceiving *p* is a non-factive act, given that it is supposed to be compatible with the world actually being such that ~*p* is true, so one can successfully conceive *p* although what *p* represents is not the case.

Conceiving *p* and, consequently, the susceptibility of this proposition to the performance of this act, is supposed to be epistemically efficacious in the sense that doing so provides some defeasible evidence that what *p* represents is possible. Some have maintained that the insusceptibility of *p* to the performance of this act provides some evidence that ~*p* is necessary. However, this claim is much less widely accepted than the former. Many recognize that there are numerous reasons, based on one's own cognitive limitations, that one might fail to conceive *p* (e.g., its complexity) and, hence, failure to do so might be more about these limitations than the necessity of what ~*p* represents (see, for instance, Tidman 1994: 297) So I focus herein on the claim that the conceivability of a proposition is (defeasible) evidence for the possibility of what it represents.

If conceiving *p* is supposed to be the evidence by which one comes to know that the arrangement of things *p* represents is possible, and one has such modal knowledge, then, obviously, one must be able to perform this act of conceiving *p*. Were one unable to do so, then the conceivability of *p* clearly could play no (direct) epistemic role in one's knowing that what it represents is possible. Moreover, one must be able to cognize when one is conceiving *p* (or when one has done so), otherwise the evidence supposedly provided by doing so would be unavailable. Thus, conceiving *p*, if it is to be epistemically efficacious with respect to modal knowledge, must be accessible to one, in being both performable and consciously so. Finally, and significantly, conceiving *p* is not supposed to complement other evidence one might have for accepting that what *p* represents is possible. Rather, it is supposed to be one's *only* evidence and, a fortiori, one's primary and basic evidence. If one knows that what *p* represents is possible, it is only via conceiving *p*. For those, then, who regard conceivability as crucial to modal knowledge, a plausible account of what it is to conceive is indispensable to a satisfactory modal epistemology.

What has been said so far about conceiving is accepted by anyone who regards the conceivability of a proposition as epistemically efficacious. However, this account certainly does not suffice for a satisfactory modal epistemology. There are ever so many consciously performable representational activities whereby one comes to be in a mental state with a given propositional content—speaking, thinking, believing, supposing, considering, entertaining, etc.—and this general account provides no insight into which of these specifically is the one that is epistemically efficacious with respect to modal knowledge. Still, this general account does present some conditions on the apt specific account. The earlier discussion regarding the ontological basis of modal knowledge presents more. Most importantly, that discussion presents the condition that the apt specific account of conceiving must make it perspicuous how, in conceiving *p*, one is engaged not merely with *p* but with those things it represents. For it is the constraints inherent to these (non-representational) things in the world that determine what is possible (and necessary), and if conceiving *p* is one's only evidence for the possibility of what *p* represents, this evidence must be of and, hence, come from those inherent constraints.

In the next section, I show that no heretofore proposed account of conceiving meets all these conditions; in the one following, I show that no account of conceiving could.

15.3 There is no apt specific account of conceivability

In this section, I canvass many specific accounts of what it is to conceive a proposition, *p*, in order to show that on none is doing so plausibly epistemically efficacious with respect to knowledge of the possibility of what *p* represents. Earlier, at the outset, I stated that I think one has a good deal of modal knowledge. Here, for the purposes of argument, I assume some uncontroversial examples of it, in particular, knowledge of certain propositions representing what is impossible. My primary strategy is to show that on a given specific account of what it is to conceive *p*, one is able to conceive a proposition that represents an arrangement of things that is impossible. Showing this indicates that that account should be rejected as one on which conceiving is epistemically efficacious with respect to knowledge of possibility. If, on a given account, one were able to conceive both propositions representing what is possible and those representing what is impossible, one would need some further means of distinguishing among the conceivable propositions the ones that represent what is possible. The conceivability of a proposition is, however, supposed to be the sole source of evidence regarding the possibility of what it represents.

Some might be unmoved by the successful implementation of this strategy. If I am able to show that on a given account of what it is to conceive a proposition, one is able to conceive both those representing what is possible and what is impossible, then, one might claim, all I have shown is that that account is *fallibly* epistemically efficacious, that in some cases it might provide adequate evidence for the possibility of what is represented by *p*, but in other cases not. Unless there is evidence for the impossibility of what *p* represents, the conceivability of that proposition should be taken as evidence for the possibility of what it represents. Given the widespread acceptance of fallibilistic views of knowledge in contemporary epistemology, such an account of conceiving might be deemed good enough for modal knowledge.

This sort of response is mistaken. The conceivability of a proposition, *p*, does not provide fallible (or prima facie) evidence for the possibility of what *p* represents; it provides no evidence at all. In each case in which a specific account of conceivability is shown to be compatible with conceiving both propositions that represent what is impossible and ones that represent what is possible, this laxness can be accounted for in terms of a lack of engagement with those constraints inherent to things that are the ontological basis of necessity and possibility simpliciter. Thus, conceiving on that account has, literally, nothing to do with the basis of modal knowledge and so cannot be epistemically efficacious with respect to it.[4]

There have been many proposals for what, specifically, it is to conceive a proposition. These can be distinguished as *negative* accounts or *positive* ones.[5] On a negative account of conceiving, a proposition, *p*, is conceivable if, in considering *p*, one does not discern it to be contradictory. On a positive account, *p* is conceivable if one is able to perform consciously an act that presents what is represented by *p* as being the case.

Any negative account of what it is to conceive a proposition faces immediate and seemingly conclusive objections. Consider *p*, the proposition that the man next door is not the man I saw at the grocery store, where the man I saw at the grocery store is in fact, yet unbeknown to me, the man next door. Given that the man next door *is* the man I saw at the grocery store, what is represented by *p* is impossible; no thing could fail to be itself. Nevertheless, I can come to be in a mental state with content *p*, consider that content, and not be able to discern a contradiction in it. Thus, *p* is conceivable, on this account, though what it represents is not possible. There are many similarly problematic propositions. Consider *g*, Goldbach's conjecture, that every even number greater than two is the sum of two primes. Goldbach's conjecture is a mathematical claim and, as such, is plausibly either necessary or impossible; it has not been proven true, nor has

a counterexample been produced.[6] One can consider *g* and not discern a contradiction, but one can also consider ~*g* and not discern a contradiction. Both *g* and ~*g* are, then (negatively) conceivable, yet what is represented by one is impossible. Consider, as well, the propositions that water is not H_2O, that an object is entirely red and entirely green (at the same time), that a man metamorphoses overnight into a giant sentient insect. In considering these, one can discern no contradiction, so each proposition is (negatively) conceivable. However, what each represents is, presumably, impossible.

A negative account of conceivability is compatible with conceiving what is in fact impossible. But perhaps this is not surprising, for this sort of approach is misguided in principle. First of all, one might fail to discern a contradiction in a given proposition, *p*, because of one's own cognitive limitations or because of the complexity of *p* rather than because *p* is free from contradiction; in which case, *p* would be conceivable though what it represents is impossible. This casts doubts on the reliability and scope of this approach. Much more importantly, though, when considering whether a proposition is contradictory, one reflects on the concepts associated with the terms that express that proposition. These concepts, the representational means one has to cognize and recognize and classify (non-representational) things in the world need not be—and most often are not—based on or attuned to those features of things that can plausibly be regarded as the inherent constraints that are the ontological basis of necessity and possibility. It is to be expected, then, that there is a lack of correspondence between how one conceptualizes things and how those things must be or could be in themselves.

Indeed, it is this lack of correspondence that enables one quickly to find many examples of (negatively) conceivable propositions that represent what is impossible. One can conceptualize a particular thing, e.g., a man, in ever so many ways that are independent of the constraints inherent to that thing, and so there are propositions about it that are not contradictory, yet nonetheless represent an impossible arrangement of things. Likewise, water need not be conceptualized as H_2O, although that stuff itself—water, i.e., H_2O—cannot fail to be H_2O, and so the proposition that water is not H_2O is not contradictory, yet nonetheless represents what is impossible. Similar points can be made with respect to the other examples of negatively conceivable propositions that represent what is impossible.

Given the gap between how one conceptualizes a thing and the constraints inherent to that thing—the basis of what is necessary and possible for it—any negative account of conceivability, which depends on discerning contradictions and, hence, the concepts representing a thing, is abortive. Still, some attempt to defend the epistemic efficacy with respect to modal knowledge of a negative account of what it is to conceive. Thus, to this end, Chalmers introduces the general distinction between *prima facie* and *ideal conceivability* (Chalmers 2002: 147–149; Menzies 1998). This distinction presupposes some specific account of conceivability. Supposing, then, a negative account like the one just considered, a proposition, *p*, is prima facie (negatively) conceivable if it is conceivable to one on first appearances or even some consideration; in this case, *p* is prima facie negatively conceivable if, after some consideration, one does not discern *p* to be contradictory. A proposition, *p*, is ideally (negatively) conceivable, if it is conceivable on "ideal rational reflection", that is, if a perfect rational agent fails to discern a contradiction in *p*.

The distinction is thought to be helpful to those who wish to defend the epistemic efficacy of conceivability, for if one maintains that in order for the conceivability of a proposition, *p*, to provide evidence for the possibility of what *p* represents, *p* must be ideally conceivable, then one is able to dismiss at least some of the examples of (negatively) conceivable propositions that represent what is impossible. Both Goldbach's conjecture and its negation are not ideally conceivable. A perfectly rational agent would, presumably, eventually identify some contradiction in

the false proposition and would also discern a contradiction in any other false proposition from an apriori domain. Maintaining that it is only ideal (negative) conceivability that is epistemically efficacious with respect to knowledge of possibility would also address the concern, noted earlier, about the reliability and scope of negative conceivability in light of cognitive limitations and the complexity of some propositions.

But this attempts to respond to the relatively trivial objections to negative conceivability while leaving unaddressed the most telling one, and by means that are ineffectual. It is by no means clear what "ideal rational reflection" is, but it is clear that human persons do not reflect in any ideal way, nor could they given their many unavoidable shortcomings. One of the conditions on the epistemically efficacious account of conceiving is that one be able to perform the act, but if what it is to conceive a proposition is to ideally conceive it, one cannot. So the notion of ideal conceivability is really of no use in a satisfactory modal epistemology.[7] Moreover, the most telling objection against negative conceivability is that one can (negatively) conceive a proposition—like that water is not H_2O—that represents what is impossible yet is not contradictory. It is not as if this proposition, or the others presented to this point, might contain a hidden contradiction; it is straightforward enough to see that it does not. Thus, it is undeniably negatively conceivable, though it represents what is not possible.

Whatever it is to conceive a proposition, it seems hard to deny that a proposition like that water is not H_2O is conceivable. It is too simple, too ordinary. So propositions like this have vexed those who maintain that (negative) conceivability is epistemically efficacious with respect to modal knowledge since Kripke and Putnam and others began discussing so-called *aposteriori necessities*; propositions like that water is H_2O, that gold has atomic number 79, that Hesperus is Phosphorus, that this wooden lectern is not made out of ice. The negations of all these propositions, which represent impossible arrangements of things, are negatively conceivable.[8] In response, defenders of the epistemic efficacy of negative conceivability have proposed another way of distinguishing acts of conceiving. One can consider a proposition (say, for the purposes of determining whether it is contradictory), focusing merely on the concepts associated with the terms expressing that proposition and ignoring how the world actually is. Or one can consider a proposition (say, again, for the purposes of determining whether it is contradictory), taking into consideration the concepts associated with the terms that express that proposition, but also considering to what those concepts in fact apply. Chalmers call the former *primary conceivability* and the latter *secondary conceivability*.[9]

The negation of a necessary aposteriori proposition, such as that water is not H_2O, is not secondarily (negatively) conceivable, for when one takes into consideration that water is indeed H_2O, one immediately discerns a contradiction. There is a sense, then, in which one can deny that a proposition like that water is not H_2O is conceivable. Still, there are other propositions—like that an object is entirely red and entirely green (at the same time) and that a man metamorphoses overnight into a giant sentient insect—that are secondarily conceivable and that represent what is impossible. So secondary conceivability does not address the most telling objection to negative conceivability. When one primarily conceives a proposition, *p*, one ignores the world, considering only the concepts associated with the terms expressing *p*. Such an account of conceiving is certainly of no use in acquiring modal knowledge, for such knowledge must come by engaging with things in the world. It is the constraints inherent to these that are the basis of necessity and possibility. The distinction between primary and secondary conceivability is, therefore, a red herring with respect to modal knowledge.

I conclude that the immediate objections to any negative account of what it is to conceive a proposition, in a way that is supposed to provide evidence for one's knowledge of possibility, are indeed conclusive. Despite the proposed distinctions (viz., prima facie v. ideal, primary v.

secondary), there remain propositions that are negatively conceivable and yet represent what is impossible. No positive account, though, is any more useful to a satisfactory modal epistemology.

On a positive account, one will recall, *p* is conceivable if one is able to perform consciously an act that presents what is represented by *p* as being the case. This characterization does little to specify the general account of conceivability articulated in the previous section. A number of suggestions have been made over the years, though, as to what the specific representational act is that presents what is represented by *p* as being the case. It has been suggested, then, that what it is to conceive a proposition, *p*, is to *understand p* or *believe p* or *entertain p* (in the sense of including *p*, with other propositions it entails, in a coherent description)[10] or *conjecture p* or *suppose p* or *mentally simulate p*, etc. (see Gendler and Hawthorne 2002: 7–8 for further suggestions.) However, even those who accept that the conceivability of a proposition provides some evidence for the possibility of what it represents recognize that these suggestions are futile. One can readily perform these acts with a proposition that represents an impossible arrangement of things. One can understand that an object is entirely red and entirely green (at the same time); one can believe that Hesperus is not Phosphorus; one can entertain the proposition that a man metamorphoses overnight into a giant sentient insect, etc.

A more stringent account of what it is to conceive is needed, one that does not permit one to conceive a proposition that represents what is impossible. One proposal, present since at least Hume, is that what it is to conceive a proposition, *p*, is to *imagine p*. Understanding conceivability in terms of imaginability has come to be standard in contemporary investigations of modal epistemology through the work of Stephen Yablo (1993) and subsequent highly influential discussion (see, e.g., van Inwagen 1998; Chalmers 2002). On this sort of account, for a proposition, *p*, to be imagined, it is not necessary that one form a visual mental image of what *p* represents. Imagining, in the relevant sense, need not involve any sensory image, for one is supposed to be able to imagine propositions that are not sensible at all, such as that there exists an invisible being that leaves no trace on perception (Chalmers 2002: 151) or that God is omnipotent or that there is now a sound beyond the range of one's hearing.[11] To imagine a proposition, *p*, then, one must merely call to mind a scenario in which *p* is true (see Yablo 1993: 29 and, following him, Chalmers 2002: 150). Doing so is thought to provide some evidence that what *p* represents is possible.

If this is all there is to imagining in the pertinent sense, however, it seems one can easily imagine propositions, such as that water is not H_2O or that Hesperus is not Phosphorus, that represent what is impossible. Famously, though, Kripke and many following him maintain that one *cannot* imagine these propositions. With respect to, at least, the negations of aposteriori necessities, one can only imagine propositions that represent qualitatively indiscernible arrangements of things (see Kripke 1980: 103–4). Despite one's efforts, one does not imagine Hesperus, i.e. Phosphorus, when attempting to imagine that Hesperus is not Phosphorus; rather, one imagines distinct planets that merely look like Hesperus (Phosphorus). According to Kripke and his followers, then, one can be mistaken about what it is that one takes oneself to be imagining.

This raises the question of what exactly one must do to imagine a proposition in the way that provides evidence of the possibility of what that proposition represents. The question becomes even more pressing in light of Peter van Inwagen's discussion of the matter. van Inwagen accepts that the imaginability of a proposition, *p*, provides some evidence of the possibility of what *p* represents. However, since the basis of possibility is in things, in accepting the possibility of what *p* represents, one is committed to a coherent reality incorporating that possible arrangement of things among a world of others. Recognizing this limits what one is supposed to be able to imagine. To imagine, say, that there is a naturally occurring purple cow—a proposition that most

would not hesitate to accept as representing a possibility—one would have to be competent with a great many details of chemistry or bovine genetics that no one currently is. This is because, according to van Inwagen, in order to really imagine that there is a naturally occurring purple cow, one would have to have some idea how such a thing is in the world, how, for instance, purple pigment is encoded in cow DNA (see van Inwagen 1998: 78). Some find this an unreasonably high standard for imaginability, prompting the proposal of less demanding accounts. For example, Heimir Geirsson proposes one on which to imagine a proposition, *p*, one must enhance with sufficient detail the scenario called to mind in which *p* is true, enough to show that it contains no obvious feature that would undermine the plausibility of the possibility of what *p* represents, but not so much as to demonstrate how it is possible (see Geirsson 2005).

I think that any question about what one must do to imagine a proposition, *p*—how much detail one must call to mind in order to do so and, relatedly, whether one is, in fact, imagining *p* when one takes oneself to be—indicates confusion. These questions are simply inconsistent with one's familiar intentional capacities. To call to mind a scenario in which proposition *p* is true, one must merely bring *p* before one's mind. To do so, one need only consider *p*. If there is any question as to what exactly one is considering (and it is hard to see, setting aside the present context, how any such question could ever arise), one need only *stipulate* what is before one's mind. Hence, when I imagine that there are carnivorous rabbits on Mars, I simply bring to mind, via this very proposition, a scenario in which there are carnivorous Martian rabbits, that is, one in which it is true that there are these rabbits.[12] I think that the strongest grounds for what I am maintaining here can be obtained just by calling to mind a proposition, any proposition, and considering how farfetched it would be for another to claim that one has failed to call precisely that proposition to mind.[13] There is, therefore, no room for doubt regarding what one calls to mind and whether one can call to mind a scenario in which a proposition is true.

If this is so, one can imagine any proposition one can express; as I have said elsewhere, the imagination is utterly promiscuous.[14] So I can imagine any number of propositions that represent impossible arrangements of things: that I am a pastrami dip, that I am a pastrami dip that is not identical to itself, that 2 + 2 = 5, that there are round squares. Some, however, have suggested that there are indeed limitations on what one can imagine or that imagining, despite being stipulative, is not entirely idle with respect to modal knowledge. One is supposed to be unable to imagine "morally deviant" propositions, such as that hurting a child for the fun of it is not wrong, and other outré propositions similar to the sort I just listed (see Gendler 2000; Weatherson 2004; Kung 2010: 629). Yet I seem to have no difficulty imagining such propositions. The suggestion (or assertion) that I cannot is inconsistent with the intentional capacities with which I am so familiar. I am dubious, then, that there are considerations that could render the suggestion even remotely plausible. Peter Kung recognizes that what one imagines is largely stipulative (see Kung 2010). Nonetheless, he maintains that what one imagines, when it is accompanied by a mental image, contains a qualitative core—which is not stipulated—that can provide some evidence regarding the possible arrangement of things. But with any mental image, its interpretation requires stipulation; one must stipulate the lighting in which the image is supposed to be viewed, the perspective, the curvature of space, etc. Kung, then, seems to underestimate the extent to which what one imagines is stipulated. Therefore, I see no limits on the utter promiscuity of the imagination, nor any part of an act of conceiving qua imagining that is epistemically efficacious with respect to modal knowledge.

Accounts of what it is to conceive a proposition, *p*, in a sense that is epistemically efficacious with respect to knowing that what *p* represents is possible are either negative—requiring one to fail to discern a contradiction in *p*—or positive—requiring one actively to present what is represented by *p* as being the case. The foregoing discussion shows that any proposed specific

account, negative or positive, is consistent with one conceiving a proposition that represents what is impossible. There is, therefore, no specific account of conceivability that is relevant to acquiring modal knowledge. This failure of all accounts is some indication that the very idea that conceivability is pertinent to a satisfactory modal epistemology is mistaken.

15.4 There could be no apt account of conceivability

To some, the preceding section might have seemed like a gratuitous, even frustrating, interlude. The insuperable problems for any specific account of what it is to conceive a proposition, on which its conceivability is supposed to provide evidence for the possibility of what it represents, might have been apparent given merely what conceivability in general is supposed to be and the discussion of the ontological basis of modal knowledge. Indeed, I think direct and compelling arguments for the epistemic idleness of conceivability can be made on these grounds. But the idea that conceivability is epistemically efficacious with respect to modal knowledge is so prevalent and so entrenched in philosophical tradition that any such direct argument would likely be judged facile and dismissed. Now, in light of the futility of a host of specific accounts of what it is to be conceivable in an epistemically efficacious way, I hope the force of these direct arguments can be appreciated. The considerations they raise illuminate why the specific accounts fail.

The ontological basis of modal knowledge is (and must be) the actual constraints inherent to the things in the world, the constraints whereby these things are necessarily or possibly certain ways. Any evidence for the possibility of what some proposition, *p*, represents must indicate those actual constraints inherent to the things represented by *p* (whereby those things are possibly certain ways). A mental state that is epistemically efficacious with respect to knowing that what *p* represents is possible must, then, indicate those actual constraints. Conceiving *p* is an essentially representational act, one that is non-factive. One can successfully conceive *p* without what *p* represents being the case. The things *p* represents might not be as *p* represents them, or they might not even be at all. Of course, the conceivability of *p* is not supposed to provide evidence that what it represents is in fact the case; it is supposed merely to provide evidence that what it represents is possible. However, the possibility of what *p* represents is determined by the actual constraints inherent to those things *p* represents. If one can conceive *p* even in the absence of the things *p* represents, then conceiving *p* is not sufficiently grounded in those things to indicate *any* of their features, neither their mundane qualities nor those inherent constraints whereby they possibly are certain ways. Therefore, if conceiving *p* is non-factive with respect to what *p* represents, it is also non-factive with respect to the possibility of what *p* represents. In which case, the conceivability of *p* provides no evidence for that possibility.

A further and perhaps deeper problem with the idea that the conceivability of a proposition, *p*, provides evidence for the possibility of what *p* represents is that conceiving is an essentially representational and crucially propositional act. The general issue here is that a state that represents propositionally, that is, presents some thing(s) as being a certain way, cannot itself be evidence that what it represents is so, because there is *always the question of whether the things it presents are in fact as presented*.[15] Thus, even if, somehow, despite conceiving *p* being non-factive, in conceiving *p* one comes to be in a representational state that presents the things that *p* represents as having those inherent constraints whereby they possibly are as *p* represents them, this would not suffice as evidence for those things in fact having those inherent constraints. This is because the question would remain whether those things do in fact have the inherent constraints they are presented as having. To answer this question requires direct, non-representational engagement with those things themselves—engagement that reveals that they indeed have those inherent constraints. Yet conceiving *p* is supposed to be the *only* evidence one has that what *p* represents

is possible and, hence, the only evidence one has that the things *p* represents have the inherent constraints whereby they are possibly as represented by *p*. Conceiving *p*, therefore, presents no evidence that what *p* represents is possible.

These two arguments illuminate why all the specific accounts of conceivability as epistemically efficacious with respect to modal knowledge fail: in conceiving, one is not appropriately engaged with the things in the world whose inherent constraints are the source of necessity and possibility. Consequently, on any specific account, one can conceive what is impossible. I see these arguments as giving expression to—and vindicating—a basic concern that nags anyone who reflects at all critically on the supposed connection between conceivability and modal knowledge. One wonders how the performance of a mere representational act can provide insight into how things in the world, existing independently of any mind, could be (or even are). What I have shown is that it cannot.

15.5 A better way

Conceiving a proposition, *p*, can provide no evidence for the possibility of what *p* represents. Conceivability, therefore, is epistemically idle with respect to modal knowledge. The notion does not deserve its long-standing centrality in discussions of modal epistemology; indeed, it seems not to deserve any role in any philosophical discussion. Given that one has modal knowledge, it must come by means other than conceiving propositions. Considering why conceivability is epistemically idle provides some guidance on how to pursue a satisfactory modal epistemology.

The actual constraints inherent to the things in the world are the source of necessity and possibility, so modal knowledge must begin with them. Since necessity and possibility have a common source, one should expect, and attempt to elaborate, a unified modal epistemology, one that accounts for one's knowledge of necessity and possibility in largely the same way. This account should make clear how one engages with the constraints inherent to things in themselves directly—relationally rather than representationally—and so should not rely crucially on one's conceptual or apriori capacities. It seems, then, that a satisfactory modal epistemology will share some similarities with the apt account of perceptual knowledge. This indicates that one can expect the intentional relation of *acquaintance* to play a pivotal role.[16]

Notes

1 For the purposes of the present discussion, I take modal knowledge to be propositional.

2 See Hume (2007: 26). For historical discussion of the putative connection between conceivability and necessity and possibility, see Boulter (2011), Alanen (1991), and Alanen and Knuuttila (1988).

3 Yablo (1993: 2). Yablo's paper has become a contemporary *locus classicus* in discussions of modal epistemology.

4 I should note, in this connection, that the problem with an account of conceivability on which one is able to conceive both propositions that represent what is possible and ones that represent what is impossible is not that on such an account one would not be able to know, merely by conceiving a proposition, *p*, that one knows that what *p* represents is possible (because conceiving *p* is compatible with what *p* represents being impossible). In other words, the problem with such a lax account of conceivability is not that it violates some so-called *KK Principle*, according to which, in order to know *p*, one must know that one knows *p*. Rather, the problem with an account of conceivability that is lax in this way is that this laxness indicates that conceiving is not engaged with the basis of modal knowledge in the requisite way for it to be epistemically efficacious.

5 This distinction is introduced in Chalmers (2002). Chalmers notes a connection here to James van Cleve's distinction between *weak* and *strong conceivability* (Chalmers 2002: 156), according to which *p* is weakly conceivable if one does not see that *p* is impossible and *p* is strongly conceivable if one sees that *p* is possible (see van Cleve 1983). The connection is tenuous. Whereas Chalmers's distinction is meant

to characterize different specific accounts of what it is to conceive, van Cleve's is meant to be an articulation of two such specific accounts. Moreover, van Cleve elaborates these accounts in terms of *intuition*: to see that *p* is possible is to intuit the truth of *possibly p*. A modal epistemology in terms of intuition is usually regarded as a competitor to any relying crucially on conceivability, and it raises distinct epistemic issues. Hence, I set aside van Cleve's discussion of weak and strong conceivability.

6 This much-discussed example was introduced, for the same use to which it is put here, in Kneale (1949: 79–80).

7 See Geirsson (2005: 291) for a similar point.

8 Indeed, Putnam makes the challenge explicit: "it is conceivable that water isn't H_2O. It is conceivable but it isn't logically possible [i.e., possible simpliciter]! Conceivability is no proof of logical possibility [i.e., possibility simpliciter]" (see Putnam 1975: 233).

9 See Chalmers (2002: 156–159). These two ways of considering what is expressed by a sentence is the basis of so-called *two-dimensional semantics*. See Jackson (1998: 47–52) for additional introductory exposition of this framework.

10 See Tidman (1994) for the sources of these suggestions. See, as well, van Cleve (1983: 36) and Yablo (1993: 9–12).

11 See Tidman (1994: 299). Yablo makes clear that he does not take imagining to require sensory images at Yablo (1993: 27 n. 44).

12 The example comes from Seddon (1972), who can be construed as denying that one can conceive such rabbits. Geirsson concurs. See Geirsson (2005).

13 Wittgenstein makes a similar point with respect to imagining King's College on fire. See Wittgenstein (1958: 39).

14 See Fiocco (2007). For similar animadversions regarding taking conceivability to be imaginability, see Byrne (2007).

15 In Fiocco (2019), I employ similar considerations to argue for a naïve realist account of perception; hence, perception must be *relational* rather than *representational*.

16 I would like to thank Yuval Avnur for extremely helpful written comments on a draft of this paper. Duncan Pritchard also provided helpful written comments, for which I am grateful.

References

Alanen, L. (1991) "Descartes, Conceivability, and Logical Modality," in T. Horowitz and G. Massey (eds.), *Thought Experiments in Science and Philosophy*, Savage, MD: Rowman and Littlefield.

Alanen, L. and Knuuttila, S. (1988) "The Foundations of Modality and Conceivability in Descartes and his Predecessors," in S. Knuuttila (ed.), *Modern Modalities: Studies of the History of Modal Theories from Medieval Nominalism to Logical Positivism*, Dordrecht: Kluwer.

Boulter, S. (2011) "The Medieval Origins of Conceivability Arguments," *Metaphilosophy* 42: 617–641.

Byrne, A. (2007) "Possibility and Imagination," *Philosophical Perspectives* 21: 125–144.

Carnap, R. (1947) *Meaning and Necessity*, Chicago: University of Chicago Press.

Chalmers, D. (2002) "Does Conceivability Entail Possibility?" in T. Gendler and J. Hawthorne (eds.), *Conceivability and Possibility*, Oxford: Oxford University Press.

van Cleve, J. (1983) "Conceivability and the Cartesian Argument for Dualism," *Pacific Philosophical Quarterly* 64: 35–45.

Fiocco, M.O. (2007) "Conceivability, Imagination and Modal Knowledge," *Philosophy and Phenomenological Research* 74: 364–380.

Fiocco, M.O. (2019) "Structure, Intentionality and the Given," in C. Limbeck-Lilienau and F. Stadler (eds.), *The Philosophy of Perception: Proceedings of the 40th International Wittgenstein Symposium*, Berlin: De Gruyter.

Geirsson, H. (2005) "Conceivability and Defeasible Modal Justification," *Philosophical Studies* 122: 279–304.

Gendler, T. (2000) "The Puzzle of Imaginative Resistance," *Journal of Philosophy* 97: 55–81.

Gendler, T. and Hawthorne, J. (2002) "Introduction: Conceivability and Possibility," in T. Gendler and J. Hawthorne (eds.), *Conceivability and Possibility*, Oxford: Oxford University Press.

Hume, D. (2007) *A Treatise of Human Nature: A Critical Edition*, D.F. Norton and M.J. Norton (eds.), Oxford: Oxford University Press. (Original work published 1739/40)

van Inwagen, P. (1998) "Modal Epistemology," *Philosophical Studies* 92: 67–84.

Jackson, F. (1998) *From Metaphysics to Ethics: A Defense of Conceptual Analysis*, Oxford: Oxford University Press.

Kneale, W. (1949) *Probability and Induction*, Oxford: Oxford University Press.
Kripke, S. (1980) *Naming and Necessity*, Cambridge, MA: Harvard University Press.
Kung, P. (2010) "Imagining as a Guide to Possibility," *Philosophy and Phenomenological Research* 81: 620–663.
Menzies, P. (1998) "Possibility and Conceivability: A Response-Dependent Account of Their Connections," in R. Casati and C. Tappolet (eds.), *Response-Dependence*, European Review of Philosophy, Vol. 3, Stanford, CA: CSLI Press.
Putnam, H. (1975) "The Meaning of 'Meaning'," in *Mind, Language, and Reality*, Cambridge: Cambridge University Press.
Quine, W.V.O. (1951) "Two Dogmas of Empiricism," in *From a Logical Point of View*, Cambridge, MA: Harvard University Press.
Quine, W.V.O. (1953) "Three Grades of Modal Involvement," in *The Ways of Paradox and Other Essays*, Cambridge, MA: Harvard University Press.
Seddon, G. (1972) "Logical Possibility," *Mind* 81: 481–494.
Sidelle, A. (1989) *Necessity, Essence, and Individuation*, Ithaca, NY: Cornell University Press.
Thomasson, A. (2007) *Ordinary Objects*, New York: Oxford University Press.
Tidman, P. (1994) "Conceivability as a Test for Possibility," *American Philosophical Quarterly* 31: 297–309.
Weatherson, B. (2004) "Morality, Fiction, and Possibility," *Philosopher's Imprint* 4: 1–27.
Wittgenstein, L. (1958) *The Blue and Brown Books*, New York: Harper and Row.
Yablo, S. (1993) "Is Conceivability a Guide to Possibility?" *Philosophy and Phenomenological Research* 53: 1–42.

Chapter 16

EPISTEMOLOGY, THE CONSTITUTIVE, AND THE PRINCIPLE-BASED ACCOUNT OF MODALITY

Christopher Peacocke

The purpose of this chapter is to address some challenges to the principle-based account of metaphysical necessity that I offered in Peacocke (1999: ch. 4, 2002). The challenges arise for the epistemology of the account and for its associated theory of modal understanding. I begin with a very brief summary of the principle-based account.

The principle-based account aims to steer a middle course between the modal realism of David Lewis (1986) and apparently mind-dependent accounts of the kind found in Simon Blackburn (1993). The overarching idea of the principle-based account is that in elucidating metaphysical necessity, we should not start with some independently understood notion of a possible state of affairs and then go on to say that a necessary proposition is one that holds in all possible states of affairs. Rather, we state a set of principles to which a specification must conform if it is to count as possible. We then see necessary propositions as consequences of the set of principles that collectively define what makes something possible.

A technical development of this idea can be given, and that technical development is required if we are to show how the account establishes uncontroversial metaphysical necessities. For the purposes of formulating and addressing the epistemological and understanding-theoretic issues, however, the following informal and simplified outline should suffice.

A specification is a triple of assignments. A specification assigns extensions to concepts; it assigns n-place properties to n-place atomic concepts; and it assigns n-place extensions to n-place properties. Truth, according to a given specification, of a Thought built up from concepts, or of a Russellian proposition built from objects and a property or relation, is defined by the standard recursions. A specification *s* is then said to be admissible if it meets the following three conditions:

(1) *s* is admissible only if for any concept, the extension assigned to the concept is the result of applying the same rule as is applied in the determination of the actual semantic value of the concept. ('The Modal Extension Principle')
(2) *s* is admissible only if for any entity, whether an object or a property, the assignment respects what is constitutive of that entity. ('The Constitutive Requirement')
(3) An assignment is admissible if its admissibility is not excluded by (1) and (2). ('The Principle of Plenitude')

For the issues involved in providing a less simplified account, and in particular the issue of variable domains of assignments, see Rosen (2002) and Peacocke (2002).

According to the principle-based account, a Thought or proposition is possible if it holds in some admissible assignment. It is necessary if it holds in all admissible assignments. To illustrate with a maximally simple example, consider a Thought of the form (A&B)→A, where '→' is the material conditional. If an admissible assignment counts A&B as true, it must by the Modal Extension Principle also count A as true. Otherwise, the assignment would not be in accordance with the rule that determines the truth-value of conjunctions in the actual world, which requires that in any case in which A&B is true, A is true. It follows that (A&B)→A holds under all admissible assignments, and so is necessary. This account of its necessity involves neither Lewisian possible worlds nor any kind of mind-dependence.

The principle-based account provides a means of integrating the metaphysics and the epistemology of metaphysical necessity. The metaphysics of necessity is provided by the account in terms of admissibility. To grasp the concept of necessity, a thinker has tacit knowledge of (1) through (3) and of the characterization of necessity in terms of admissibility. A thinker can obtain knowledge of modal truths by applying this tacit knowledge. If a thinker judges in accordance with the content of this tacit knowledge, her judgment will in the nature of the case amount to knowledge, for it is reached by those very principles that determine modal truth itself.

Under this elaboration of the principle-based account, the metaphysics of necessity is prior in the order of philosophical explanation to the nature of thought about necessity. We draw upon the account of the metaphysics of necessity in explaining what it is to grasp the notion of necessity. The metaphysics is also similarly prior in the order of explanation of the epistemology of necessity, for the corresponding reason. We draw on the metaphysics of necessity in explaining how knowledge of necessity can be obtained.

Criticism of the distinctive features of the principle-based account could be focused on either the Modal Extension Principle (1), or the Constitutive Requirement (2). In fact the Modal Extension Principle seems in general not to have been found problematic. It is true that radically revisionist and arguably skeptical views of rule-following and concepts could undermine the principle-based account. Rule-following itself has been extensively debated in the literature. My own unavoidably non-neutral view on rule-following and concepts is that the non-skeptical views that are needed for (1) to work in the account have remained standing after these debates. It is rather the Constitutive Requirement (2) on which critical discussion of the principle-based account has focused, both in oral discussion and in the literature – see especially the very thorough discussion in Roca-Royes (2010) and in Vaidya (2015). There are at least two salient issues.

The first issue is the absence to date of any account of what is involved in a theory of understanding that says that there is some kind of tacit knowledge of the Constitutive Requirement in grasping the concept of metaphysical necessity. Peacocke (1999) noted the need for a substantive account of the constitutive and of grasp of it but did not provide such an account. Simply saying that arguments from the best explanation support the principle-based account is not supplying all we should demand in philosophy. Even if arguments from the best explanation are sound, they can show us at most that something is so – for instance, that a principle-based account is correct. They do not answer the question of how it can be so.

The second issue is more specific and concerns the relation between the constitutive and metaphysical necessity. While Kit Fine's work (1994) has made us vividly aware of the importance of distinguishing the constitutive and the modal, it can still be objected against the principle-based account that we have to conceive of the constitutive as constrained by the condition that

anything that is constitutive must hold in any genuinely possible circumstances. If that is so, will not grasp of the constitutive presuppose grasp of the modal? And will that not make the principle-based account circular, both as an account of understanding and as an account of the metaphysics of modality?

I think that the notion of the constitutive is intuitively clear. It is the notion of what makes something the entity it is, whatever category the entity is, whether it be an object, event, property, number, and so forth. But this intuitive clarity does not answer the question of what it is to grasp the notion of the constitutive, nor does it supply a more detailed account of the nature of the constitutive that might aid in answering the question about grasp of the notion, nor does the intuitive clarity provide an elaboration of the role of the constitutive in the principle-based account of metaphysical modality.

The remainder of this chapter is devoted to outlining a response to these questions and criticisms. I note initially that the discussions that have expressed concerns about the Constitutive Requirement (2) have not, to the best of my knowledge, actually questioned the correctness of the principle-based account as a description of the metaphysics of modality. This lends attraction to the project of developing a better account of the constitutive and our grasp of it, an account that can be used in the elaboration of the principle-based account, rather than tossing out the principle-based account, and its various advantages, altogether. I will suggest a way of carrying out that project, and will work up to the introduction of a theoretical concept that, I will argue, we need to identify and apply if we are to give the required account of grasp of the constitutive.

I begin with some observations on the constitutive. The constitutive, like the modal, is also itself principle-based. For each category of entity, there are principles that are constitutive of entities of that category. For continuant material objects, for example, for each such object x there is some sortal kind F under which x falls, and for any object y, for $x = y$ to hold is for x to be the same F as y (Wiggins 2001). That is one such principle constitutive of a material continuant. So is a principle stating that the continuant's origins, suitably specified, are constitutive of it. These principles specify what makes the object the particular individual it is.

For the general category of abstract objects, and for each type of abstract object, there are principles that individuate objects of that type. Each natural number n is individuated by the condition for there to be n objects. The number 0 is individuated by the condition for there to be 0 *F*s is for $\sim\exists z Fz$; 1 is individuated by the condition for there to be exactly 1 *F* is for it to be the case that $\exists z(Fz \,\&\, \forall y(Fy \rightarrow y = z))$; and so forth.

Under a principle-based conception of the constitutive, the force of the Constitutive Requirement (3) in the principle-based account of modality is this: a specification is admissible only if, for any entity, it does not deny the principles that are constitutive of that entity.

So far, maybe, so good, but what does this have to do with a thinker's grasp of the constitutive that is involved in understanding modality on the principle-based account of modality? Here we have to address what can be called the *fitting* question, the issue of how we fit an account of the nature of entities of a given kind onto a thinker's ability to think about entities of that kind. We can start by considering some plausible instances of the fitting relation, before attempting a generalization.

What is involved in thinking about or in representing a particular continuant object x? In the case of continuant material objects, a certain fundamental sensitivity is needed in assessing whether some object y identified as meeting condition H at time t_2 is identical with the given object x, identified as meeting condition G at time t_1. The subject must be fundamentally sensitive, in assessing such identities, to whether x is the same F as y, where F is the sortal mentioned in a constitutive account of the object x. You have a fundamental sensitivity, in assessing, for

instance, whether I am identical with a certain teenager you knew in school, to the condition whether I am the same person as that teenager. You can possess this sensitivity without possessing the concept of the constitutive.

We need to invoke a distinction between fundamental sensitivity and mere de facto sensitivity, which rests on ancillary information. You may in fact use as a test for whether I am the same person as the teenager you knew that I have a certain distinctive style of handwriting you once saw the teenager use. This test is clearly defeasible in the light of information that many people from your school have that distinctive style of handwriting. The test of being the same person is not so defeasible, it is fundamental.

You can also be in the position of knowing that there is some sortal kind under which an object falls, without knowing what kind that is. You can perceive something, and wonder whether it is a small, remotely controlled drone, or an insect. *Being the same drone* and *being the same insect* will in some circumstances determine different answers to the question "Is that the same thing again?" For continuant material objects, it is not a priori which sortal kind it falls under. What would it even mean for it be a priori of a material object that it falls under a certain kind? Something is a priori only under a mode of presentation of an object, and there is no such thing as a canonical or distinguished mode of presentation of any given material object.

Now consider the case of natural numbers. A thinker capable of thinking of the particular natural numbers $0, 1, 2, \ldots$ will be fundamentally sensitive in her judgments about them to their constitutive individuating principles that mention the numerical quantifiers containing them. Again, we need a distinction between fundamental sensitivity and de facto sensitivity. I may judge that there are nine stars because they are at the vertices of a certain shape, and do so wrongly because my belief that there are nine vertices in that shape is mistaken. The norm to which I am fundamentally answerable, in representing there as being nine stars there, is given by the corresponding first-order quantification: that there are entities $x_1 \ldots x_9$ that are stars there, distinct from one another, and they are the only stars there. Again, this fundamental sensitivity can be present in a thinker without the thinker herself having to employ the notion of the constitutive at any point.

We are now in a position to introduce the theoretical notion promised earlier. It is that of *tacitly knowing something in the constitutive mode*. For a subject to tacitly know C in the constitutive mode is for

(1) C to be constitutive of one or more of the entities that C is about;
(2) for the subject to be fundamentally answerable to C in forming judgments or representations about that entity or those entities, in a way that is noninferential, not dependent on any other information; and
(3) for this answerability to be correspondingly a condition for thinking about that entity or those entities in the first place.

So, when C is tacitly known in the constitutive mode, there is no thinking about or representing the entity or entities C is about without already being fundamentally answerable to C.

The condition, formulated in terms of the language of first-order with identity, for there being 9 Fs is something tacitly known in the constitutive mode by subjects capable of thinking about the cardinal number 9 as such. The principle, applied to a given continuant material object x, that there is some sortal kind F such that something is identical with x if and only if it is the same F as x, is something tacitly known in the constitutive mode by those capable of thinking about or representing x.

Tacitly knowing something in the constitutive mode no more requires the thinker to have the concept of the constitutive than knowing something in an a priori way requires the thinker to have the concept of the a priori. In both the case of tacitly knowing in the constitutive mode, and the case of knowing something a priori, we characterize certain mental states of thinkers using various modestly theoretical notions. The subjects who enjoy the mental states must meet the relevant conditions involved in those theoretical notions, but these subjects do not need to conceptualize or represent those conditions, nor do they need to employ those notions. Similarly, we as philosophers may reflect on what is involved in the constitutive, and how to elucidate it, and how it is related to ordinary thought about the entities that constitutive principles concern. No such activity is required of ordinary nonphilosophical thinkers.

To say that some condition C is tacitly known in the constitutive mode does not at all make unnecessary the empirical question of how thinkers acquire the ability to represent the entities that C concerns. The question of acquisition is always genuine, and empirical answers to it are needed. The famous work of Susan Carey (2009: ch. 8) on the acquisition of the ability to represent the cardinal numbers is an answer to the empirical question in the case of that particular subject matter. The empirical question is a question about how a state is acquired that does, if what I have said is along the right lines, involve tacitly knowing some condition in the constitutive mode. The fact that it involves such tacit knowledge constrains what an account of acquisition has to be acquisition of. It neither obviates nor in any way replaces that empirical account.

I offer the theoretical concept of tacitly knowing something in the constitutive mode as an answer to what I earlier called the question of fitting constitutive principles about ontology of any given kind to thinkers capable of representing elements of that ontology.

The notion of tacitly knowing something in the constitutive mode is crucial in addressing the question of whether there is circularity in the epistemology and the theory of understanding associated with the principle-based account of modality. In the examples I have briefly considered, the characterization of what it is for C to be tacitly known in the constitutive mode does not involve the exercise of the notions of metaphysical necessity or possibility on the part of the thinker who knows C in a constitutive way. The capacities involved, in those examples, in tacitly knowing something in the constitutive mode have to do with facts about the thinker's judgments about identity in the actual world, or about judgments involving numerical quantification in the actual world. Here there is an affinity with the Modal Extension Principle (1). Facts about rules and correctness conditions in the actual world contribute to the determination of which specifications are admissible, and so to the conditions for genuine possibility, beyond the actual world.

Grasping modal concepts involves the capacity to move from contents tacitly known in the constitutive mode to corresponding constraints on admissibility, which determine the range of what a thinker takes to be possible. A thinker who does that for each category of entities she represents, and who also respects the Modal Extension Principle and the Principle of Plenitude, will be able to attain modal knowledge.

We can distinguish various degrees to which grasp of the constitutive as such can be present in a thinker's understanding of modality.

At the lowest degree, it seems there can be a thinker who moves from what she tacitly knows in the constitutive mode to constraints on admissibility, for each category in her ontology, without representing the notion of the constitutive at all. This is a case of sensitivity to the constitutive/nonconstitutive distinction without representation of that distinction as such.

At one step up, we can conceive of a thinker whose capacities include those at the previous lowest degree, but who also displays a certain productivity in her understanding of modality. For each new category that may be introduced to her ontology, this thinker at one step up always

reliably makes the right connection between her tacit knowledge involved in the grasp of that new category and the corresponding constraint on admissible specifications. The case could be such that the best explanation of this productive capacity is that she in some way represents the notion of the constitutive, and applies it in moving to corresponding constraints on admissibility when she extends her ontology.

At an even further higher degree, we have a reflective thinker who employs the concept of the constitutive at the personal level in her thought, and who thereby has the capacity to reflect on her judgments about possibility, and the conditions under which she makes those judgments. This reflective thinker, evidently an incipient philosopher, will need to employ the concept of the constitutive to unify the cases in which she is accepting constraints on the possible, and in characterizing the sensitivity of her judgments about possibility.

Somewhat different observations are required for the case of individual essence of material objects, the component of constitutive principles stating that an object's original components (at a suitably selected level) are constitutive of it. It is not true that there is some kind of sensitivity of judgments or representations about the actual world in which one could ground a requirement of holding suitable origins constant in any admissible specification. There is, however, a fundamental sensitivity in a thinker's acceptance of counterfactuals about a specific material continuant that can be regarded as extended to all admissible specifications. We judge, for instance, that you would have been harmed if there had been radioactive damage to the sperm or to the egg from which you originated. We do not think you would have been so damaged under circumstances in which that sperm and egg cell are undamaged. Such counterfactual reasoning displays a grasp of what is required for a specification to be genuinely about you. The requirement that any admissible specification whatever must not deny your actual origins can be seen as an extension – not from a sensitivity about judgments concerning the actual world, but from a sensitivity in such judgments about counterfactuals, to all genuine possibilities. If we wanted to, we could also modify the above discussion to consider two subcases in which a subject tacitly knows C in the constitutive mode. Concerning the entity or entities that C concerns, one case would distinguish the conditions for thinking about that entity or those entities in the actual world, while the other case would distinguish the conditions for thinking about them in nonactual circumstances. That would then avoid any implausible implication that those who dispute essence of origin are to be classified as failing to think about people, chairs, and animals in the actual world.

Is there some kind of illegitimate circularity in the project in which I have been engaging, that of trying to say what is constitutive of sensitivity to the constitutive, and constitutive of grasp of the constitutive? That is indeed what I am aiming to say, but there is no circularity here, any more than there would be circularity in using spatial notions (as we would have to) in saying what it is to be sensitive to spatial properties, and what it is to represent something as spatial. Questions about the nature of the constitutive are good and genuine, and they need to be addressed in metaphysics. But they are to be distinguished from questions about the nature of sensitivity to the constitutive, and from questions about grasp of the concept of the constitutive.

How does the principle-based account bear on discussions of the relation between conceivability and possibility (Chalmers 2002; Evnine 2008; Yablo 2008)? Stephen Yablo (2008) carefully distinguishes a specific kind of conceivability and argues that that kind of conceivability of *p* gives prima facie reason for judging that *p* is possible. As far as I can see, the broad features of Yablo's discussion are consistent with – indeed, dovetail well with – the principle-based account. The principle-based account is an account of what it is that, in Yablo's view, conceivability is prima evidence *for*. In saying that a specification is possible only if it respects constitutive

principles – the Constitutive Requirement (3) noted earlier – the principle-based account also supplies one resource for explaining why Yablo-style conceivability is, as Yablo says, at most prima facie evidence for possibility. As Arnauld in effect insisted in an objection to Descartes, conceiving that *p* is consistent with the impossibility of *p* if the conceiving has overlooked something that is constitutive of the entities mentioned in *p*. Similarly, the principle-based account of modality explains why the best kind of conceivability to support a claim of possibility has to involve what Chalmers (2002) calls "coherent modal imagination", which requires that it be "possible to fill in arbitrary details in the imagined situation such that no contradiction reveals itself" (153). Any contradictions would be violations of the principles that define admissibility and possibility. More generally, the principle-based account can explain why the norms on conceivability, which aims to establish possibility, are as they are.

How does this treatment bear on broadly rationalist and a priori conceptions of metaphysical modality? There are several distinctions to be drawn here. A first element of a rationalist conception to be distinguished is the idea that reasons and justification for a proposition are grounded in the theory of understanding for the concepts in that proposition. This first core element of rationalism is vindicated if good ordinary claims to modal knowledge can be justified by making essential use of the present theory of modal understanding, to the effect that it consists in tacit knowledge of the principles of possibility (1) through (3). I cannot argue for it in the present limited space, but that claim seems highly plausible. It is important to note, however, that this claim about the understanding-based character of reasons and justification does not by itself – or not without considerable further substantial philosophical argument – imply that all modal truth is in principle a priori accessible. The understanding-based character of reasons and justification is prima facie entirely consistent with some truths about the constitutive resisting purely a priori access. So we ought to distinguish the claim that reasons and justification are understanding-based from the thesis that all modal knowledge is fundamentally a priori, a thesis we can label 'the a priori claim'. This a priori claim would be a consequence of the much stronger thesis that all modal truth is fundamentally a priori, which is a formulation of what David Chalmers calls pure modal rationalism (Chalmers 2002: 194–195).

The claim that all modal knowledge is fundamentally a priori is variously motivated. One motivation is the idea that an a posteriori means of knowing can tell us only about the actual world. The other motivation is a plausible treatment of examples of a posteriori necessities, such as that Hesperus is Phosphorus, or that gold has atomic number 79. The plausible treatment of these cases is that their status as a posteriori necessities follows from propositions each of which is either modal but a priori, or else a posteriori but non-modal. The necessity of identity, $\forall x \forall y(x = y \rightarrow \Box x = y)$, is modal but a priori. Similarly, consider the proposition that if gold is a chemical kind, then if it has an atomic number, it necessarily has that number. That too is modal but also a priori. One idea motivating the a priori claim is that there are no a posteriori necessities whose status as such cannot be explained along these lines.

The treatment of the principle-based account of modality for which I have been arguing puts pressure on the a priori claim, or at the very least, it exerts pressure on us to distinguish two versions of the doctrine. One form concerns some notion of ideal or in-principle justification. (Chalmers 2002, in his related but somewhat different position, also sharply distinguishes the ideal from the non-ideal cases.) The ideal form of the a priori claim states that for any a posteriori but necessary proposition *p*, there is always some set of propositions, some of which are essentially modal and in principle knowable a priori, the others wholly non-modal and posteriori, that jointly entail *p*. The other form of the a priori claim states that our actual, non-ideal, knowledge of propositions that are posteriori necessities, and are known to be so, always traces back to such a set of propositions that are actually known.

This second form, about our actual knowledge, is prima facie undermined by the treatment I have outlined here of the principle-based account of necessity, and its associated theory of understanding and epistemology. On this treatment, a thinker may tacitly know something in the constitutive mode, and her use of this knowledge may constrain her modal judgments and knowledge, without her knowing more general purely modal propositions that contribute to the implication of necessary a posteriori status. But even for those who have the conceptual sophistication to be in a position to formulate propositions involving the notion of the constitutive, and to relate them to modality, relevant actual knowledge may be lacking. Consider the proposition that running is a mechanical action in which there is a time at which both the agent's feet are off the surface on which the running occurs. (This is a necessary condition, not a sufficient condition – fast hopping is not running.) We come to know this proposition by various kinds of empirical investigation, both by first-personal experience of our own running, and by third-personal observation of others' runnings. The proposition seems to have a necessary a posteriori status; it contributes to a correct statement of what running *is*. Do we actually know a priori an entirely general modal proposition about properties of events, such as the property of being a runner, which, together with non-modal a posteriori propositions, explain why it has this necessary a posteriori status? I doubt it. Maybe properties are individuated in terms of their causal powers, and that would serve the purpose of defending the in-principle version of the a priori claim in the case of the example of running. But I think many people would rightly be much more confident, on the present empirical evidence, that necessarily, running involves motion including a time at which both feet are off the running surface, than they are confident in any general proposition about the individuation of properties. If that is correct, then the a priori claim in its version about our actual knowledge about what is necessary a posteriori cannot be sustained given this chapter's elaboration of the principle-based account of necessity. Since, however, arguments from the very nature of modal understanding, with its involvement with the notion of the constitutive, have been used to support this last point, this state of affairs should be seen not as a refutation of a general rationalist position about modality. It is rather a by-product of an elaboration of one element of rationalism about modality.

References

Blackburn, S. (1993) "Morals and Modals," in *Essays in Quasi-Realism*, Oxford: Oxford University Press.

Carey, S. (2009) *The Origin of Concepts*, New York: Oxford University Press.

Chalmers, D. (2002) "Does Conceivability Entail Possibility?" in *Conceivability and Possibility*, ed. T. Gendler and J. Hawthorne, Oxford: Oxford University Press.

Evnine, S. (2008) "Modal Epistemology: Our Knowledge of Necessity and Possibility," *Philosophy Compass* 3: 664–684.

Fine, K. (1994) "Essence and Modality," *Philosophical Perspectives* 8: 1–16.

Lewis, D. (1986) *On the Plurality of Worlds*, Oxford: Blackwell.

Peacocke, C. (1999) *Being Known*, Oxford: Oxford University Press.

Peacocke, C. (2002) "Principles for Possibilia," *Noûs* 36: 486–508.

Roca-Royes, S. (2010) "Modal Epistemology, Modal Concepts, and the Integration Challenge," *Dialectica* 64: 335–361.

Rosen, G. (2002) "Peacocke on Modality," *Philosophy and Phenomenological Research* 64: 641–648.

Vaidya, A. (2015) "The Epistemology of Modality," in *The Stanford Encyclopedia of Philosophy*, ed. E. Zalta. Available at: https://plato.stanford.edu/entries/modality-epistemology/, accessed December 28 2016.

Wiggins, D. (2001) *Sameness and Substance Renewed*, New York: Cambridge University Press.

Yablo, S. (2008) "Is Conceivability a Guide to Possibility?" in *Thoughts: Papers on Mind, Meaning, and Modality*, Oxford: Oxford University Press.

Chapter 17

THE COUNTERFACTUAL-BASED APPROACH TO MODAL EPISTEMOLOGY

Timothy Williamson

17.1 Objective modalities

In 'Naming and Necessity', Saul Kripke compellingly separated the metaphysical distinction between necessary and contingent from the epistemological distinction between a priori and a posteriori and the semantic distinction between analytic and synthetic. Although he introduces the distinction as one of metaphysics in (he hopes) a nonpejorative sense, when drawing it he invokes no metaphysical theory. Rather, his diagnostic question is in ordinary English: given a truth, 'might it have been otherwise?' If yes, it is contingent; if no, necessary (Kripke 1972: 261, 1980: 35–36). His treatment of necessity is rooted in vivid examples, described in ordinary language. The point of assigning the distinction to metaphysics is not to mystify or elevate it, but to clarify it by avoiding confusions with what belongs to epistemology or semantics. Once we get the point, Kripke trusts, we can apply the distinction accurately to various cases, by thinking clearly about them. We have epistemological access to it, but it does not belong to epistemology.

Although 'Naming and Necessity' always had its critics, it quickly became the canonical text for the contemporary metaphysical distinction and has remained so ever since. But *how* do we know what is necessary, and what contingent, in Kripke's sense? The question is potentially misleading, since it may suggest that only one way of knowing is available, whereas a robust distinction can usually be detected in several different ways. However, a specific kind of epistemological access to the distinction is particularly relevant to Kripke's discussion, since it is the kind on which he relies, as do many other philosophers past and present. This access is not perceptual, even though we arguably sometimes learn by perception how things could have been otherwise (Strohminger 2015); we do not use perception to think through Kripke's examples. Nor is the access inductive, even though we arguably sometimes learn by induction how things could have been otherwise (Roca-Royes 2016); we do not use induction to think through the examples. Nor does the access always depend on natural science, even though natural science arguably teaches us how things could have been otherwise (Williamson 2016a); we do not use natural science to think through many of the examples. For similar reasons, the access does not depend on knowledge of a developed philosophical theory. Rather, clear-headed but more or less common-sense pre-theoretic reflection on cases is supposed to yield some a priori knowledge of necessity and possibility.[1] When Kripke discusses the distinction between the properties

an object has necessarily (or essentially) and those it has contingently (or accidentally), he insists 'it is very far from being true that this idea . . . is a notion which has no intuitive content, which means nothing to the ordinary man' (1972: 265, 1980: 41).

Excluding epistemological and semantic readings by itself does not narrow down Kripke's diagnostic 'might it have been otherwise?' to just one reading. It is still open to a variety of readings that are, if not metaphysical, at least *objective*, or in linguists' terminology 'circumstantial' or 'dynamic' (Kratzer 1977, Portner 2009). The conversational context may exclude various outlandish possibilities. For instance, the truth-value in everyday contexts of 'It might have fallen on your head' may depend on how easily you could have been standing one metre to the left. Although Kripke ignores such mundane readings when introducing his distinction, a further effect of his label 'metaphysical' is to rule them out as unintended, and so to achieve more or less maximal inclusiveness over objective possibilities; the context of abstract philosophical discussion also contributes to that effect. He does mention physical necessity, itself presumably a kind of objective necessity, but far more demanding than what matters in most everyday contexts. Indeed, he 'suggests that a good deal of what contemporary philosophy regards as mere physical necessity is actually necessary *tout court*', concluding 'The question how far this can be pushed is one I leave for further work' (Kripke 1972: 769, 1980: 164). The phrase 'necessary *tout court*' may indicate a conception of other kinds of objective possibility as restrictions of metaphysical possibility, making the latter the maximally inclusive kind of objective possibility (Williamson 2016a).

If Kripke is drawing on ordinary ways of thinking and knowing about kinds of objective possibility, those ways may well be primarily geared to the restricted kinds of objective possibility mainly at issue in ordinary contexts, like the perceptual, inductive, and some natural scientific modes of access. Such ordinary knowledge of restricted objective modalities includes knowing that one can reach the lower shelves but not the upper ones, or that it can easily snow in January but not in July. That knowledge is paradigmatically a posteriori, not a priori. Thus, common sense a priori knowledge of objective modality looks anomalous.

Kripke shows how a priori and a posteriori knowledge of modality can combine. For instance, we know only a posteriori that this table was not made of ice, but we know a priori that if it is not made of ice, then it is necessarily not made of ice. By deduction, we can come to know (not purely a priori) that it is necessarily not made of ice (Kripke 1971: 153). Perhaps, then, we come to know (a posteriori) restricted modal facts about our local region of modal space by deduction from our a priori knowledge of unrestricted modal facts about the layout of global modal space and our a posteriori knowledge about where we are in that space. But that proposal makes more problems than it solves. How do we know a priori the layout of modal space, for instance, that if the table is not made of ice, then it is necessarily not made of ice? To know our local region of modal space, we need not go on a detour through knowledge of global modal space; instead, we can use cognitive capacities aimed directly at the local region.

A plausible account of the kind of knowledge of metaphysical modality exemplified in 'Naming and Necessity' should anchor it in cognitive capacities that could evolve in species like ours without miraculous luck. If philosophers claim a special faculty for intuiting or conceiving, whose function is to detect the metaphysical distinction between necessity and contingency, one wonders how on earth humans came to be designed with a special gift for metaphysics.

The counterfactual-based approach to the epistemology of metaphysical modality enables us to avoid awarding ourselves such implausible privileges. It starts from ordinary counterfactual conditionals such as 'If she hadn't moved, she wouldn't have been noticed'. Such conditionals play an important role in ordinary thought about causal relations. We use them in learning from past experience and planning for the future. Our sensitivity to such relations has survival value.

The next section explains some logical relations between counterfactual conditionals and metaphysical modality, and, schematically, their potential to underpin corresponding epistemological relations. The final section fills in some of the epistemology.

17.2 Counterfactuals and metaphysical modality

Although counterfactuals can be given epistemic readings (Edgington 2008, Vetter 2016, Williamson 2016b, 2020: 251–253), their salient reading is usually objective; that is their intended reading throughout this chapter.

Henceforth, $\Box\mathbf{A}$ abbreviates 'It's metaphysically necessary that A'; $\Diamond\mathbf{A}$, 'It's metaphysically possible that A'; $\mathbf{A} \mathbin{\Box\!\!\rightarrow} \mathbf{C}$, the counterfactual 'would' conditional with antecedent 'A' and consequent 'C'; $\rightarrow$, the material conditional; $\leftrightarrow$, the material biconditional; $\neg$, negation; $\&$, conjunction; $\bot$, an absurdity; $\forall$ and $\exists$, quantifiers binding 'propositional' variables in sentence position.

Two logical principles link counterfactuals to metaphysical modality:

NECESSITY $\quad \Box(\mathbf{A} \rightarrow \mathbf{C}) \rightarrow (\mathbf{A} \mathbin{\Box\!\!\rightarrow} \mathbf{C})$

POSSIBILITY $\quad (\mathbf{A} \mathbin{\Box\!\!\rightarrow} \mathbf{C}) \rightarrow (\Diamond\mathbf{A} \rightarrow \Diamond\mathbf{C})$

In effect, NECESSITY says that if **A** *couldn't* hold without **C**, then **A** *wouldn't* hold without **C**; POSSIBILITY says that if **A** wouldn't hold without **C**, then if **A** could hold, so could **C**.

These principles are plausible, but not undeniable. One difference between them is that POSSIBILITY is plausible for a wide range of restricted kinds of objective possibility in place of metaphysical possibility, whereas NECESSITY is implausible for any kind of objective necessity more restricted than metaphysical necessity. For instance, when we restrict $\Diamond$ and $\Box$ to easy possibilities, POSSIBILITY still looks fine (if **A** wouldn't hold without **C**, and **A** could *easily* hold, so could **C**), but NECESSITY looks wrong (for instance, when $\mathbf{C} = \neg\mathbf{A}$, if **A** could hold but not *easily*, then **A** couldn't *easily* hold without $\neg\mathbf{A}$ holding, but it doesn't follow that **A** *wouldn't* hold without $\neg\mathbf{A}$ holding). Predictably, even under the intended metaphysical reading of the modal operators, NECESSITY has had more critics than POSSIBILITY. They deny its corollary that a metaphysically impossible antecedent counterfactually implies everything (Nolan 1997; Kment 2006, 2014; Lowe 2012; Brogaard and Salerno 2013; Berto, French, Priest, and Ripley 2018). They even deny that a contradiction counterfactually implies everything, a principle validated by all standard possible world semantics for counterfactuals (Stalnaker 1968; Lewis 1973; Kratzer 1977). For instance, they deny that if this shirt were green and not green, it would be red. To do so, many of them construct an impossible 'world' (a set of sentences) that verifies 'This shirt is green and not green' without verifying 'This shirt is red'. Space does not permit such non-standard theories of counterfactuals to be assessed in detail here. They drift towards an irrelevant epistemic reading of counterpossibles, since they provide no robust sense in which what matters for the truth-value of the conditional are the objective states of affairs expressed by the antecedent and consequent, irrespective of the wording used to express them. Arguably, the supposed counterexamples to the vacuous truth of counterpossibles are artefacts of uncritical reliance on unreflective heuristics for evaluating counterfactuals that work for the most part but fail in some cases (Williamson 2017). In what follows, we assume NECESSITY and POSSIBILITY with the metaphysical reading of the modal operators, to see where they take us.[2]

Given elementary principles of modal logic and the logic of counterfactuals, NECESSITY and POSSIBILITY entail three characterizations each of metaphysical necessity and possibility in terms of the counterfactual conditional (Williamson 2007: 156–159):

(1□) $\Box A \leftrightarrow (\neg A \mathrel{\Box\!\!\rightarrow} \bot)$
A must hold just if it wouldn't fail without absurdity.
(1◊) $\Diamond A \leftrightarrow \neg(A \mathrel{\Box\!\!\rightarrow} \bot)$
A could hold just if it's not that it wouldn't hold without absurdity.
(2□) $\Box A \leftrightarrow (\neg A \mathrel{\Box\!\!\rightarrow} A)$
A must hold just if even if it failed, it would hold.
(2◊) $\Diamond A \leftrightarrow \neg(A \mathrel{\Box\!\!\rightarrow} \neg A)$
A could hold just if it's not that it wouldn't hold without failing.
(3□) $A \leftrightarrow \forall p\ (p \mathrel{\Box\!\!\rightarrow} A)$
A must hold just if it would hold whatever held.
(3◊) $A \leftrightarrow \exists\ p\ \neg(p \mathrel{\Box\!\!\rightarrow} \neg A)$
A could hold just if something isn't such that if it held, **A** would fail.

Several of these biconditionals sound odd, because a normal conversational implicature of uttering a conditional is the possibility of its antecedent (otherwise, why bother uttering it?), but conversational oddity is compatible with truth.

We cannot treat (1□), (2□), and (3□) as all *definitions*, giving the meaning of □, for the right-hand sides of the three biconditionals are not synonymous with each other—they differ in semantic structure—and so cannot all be synonymous with the shared left-hand side. Similarly, we cannot treat (1◊), (2◊), and (3◊) as all definitions, giving the meaning of ◊. Plausibly, none of the six biconditionals pairs synonyms. Nevertheless, they tightly constrain metaphysical modality in terms of the familiar, practically indispensable counterfactual conditional. They prevent the metaphysical distinction from floating free of ordinarily intelligible ways of thinking.

If we know NECESSITY and POSSIBILITY, we can come to know the biconditionals by deduction, and in principle use them to derive knowledge of metaphysical modality from knowledge of counterfactual connections. A natural concern is that even if we are good at knowing about ordinary counterfactual connections, the biconditionals involve extraordinary counterfactuals. The reliability of our cognitive capacities for the former might not extend to the latter (Roca-Royes 2011, Gregory 2017). Our competence with counterfactuals might reach its epistemic limits just when we need it for metaphysical modality.

To see that the concern is exaggerated, just consider any opposed pair of ordinary counterfactuals $A \mathrel{\Box\!\!\rightarrow} C$ and $A \mathrel{\Box\!\!\rightarrow} \neg C$ that our ordinary cognitive capacities *can* handle. **A** and **C** may describe the behaviour of familiar everyday objects. By hypothesis, we know whether each counterfactual holds. Thus, we know which of these four cases obtains:

(i) $A \mathrel{\Box\!\!\rightarrow} C$ and $A \mathrel{\Box\!\!\rightarrow} \neg C$ both hold.
(ii) $A \mathrel{\Box\!\!\rightarrow} C$ holds; $A \mathrel{\Box\!\!\rightarrow} \neg C$ fails.
(iii) $A \mathrel{\Box\!\!\rightarrow} C$ fails; $A \mathrel{\Box\!\!\rightarrow} \neg C$ holds.
(iv) $A \mathrel{\Box\!\!\rightarrow} C$ and $A \mathrel{\Box\!\!\rightarrow} \neg C$ both fail.[3]

In each case, we can reach a non-trivial conclusion about metaphysical modality:

Given (i), by the elementary logic of counterfactuals, $A \mathrel{\Box\!\!\rightarrow} (C \,\&\, \neg C)$ holds. But by elementary modal logic, a contradiction is impossible. Thus, using POSSIBILITY, we can derive $\neg\Diamond A$.
Given (ii), we have $\neg(A \mathrel{\Box\!\!\rightarrow} \neg C)$, so by modus tollens on NECESSITY, we have $\neg\Box(A \rightarrow \neg C)$, and so by elementary modal logic, we can derive $\Diamond(A \,\&\, C)$.
Given (iii), similarly to (ii), we can derive $\Diamond(A \,\&\, \neg C)$.
Given (iv), as in (ii) and (iii), we can derive $\Diamond(A \,\&\, C)$ & $\Diamond(A \,\&\, \neg C)$.

Thus, in each case, we can know something non-trivial about metaphysical modality by deduction from what by hypothesis is knowledge of ordinary counterfactual connections. Ordinary counterfactuals are inextricably linked to extraordinary ones.

Sceptics about metaphysical modality claim that we *cannot* know anything much about metaphysical modality. Given the preceding argument, such a sceptic must also be a sceptic about ordinary counterfactual connections (or about logic), claiming that we cannot know much about them, either. That is presumably a scepticism too far.

Of course, that argument does not show that with our ordinary cognitive capacities we can know the metaphysical modal status of a given philosophically controversial proposition. For instance, it does not show that we can know whether zombies are metaphysically possible. There might be no suitable pair $\mathbf{A} \mathbin{\Box\!\!\rightarrow} \mathbf{C}$ and $\mathbf{A} \mathbin{\Box\!\!\rightarrow} \neg\mathbf{C}$. But that was to be expected. If there were an ordinary route to such knowledge, the proposition would not be so controversial. We have no reason to expect ourselves to be potentially omniscient about metaphysical modality.

Obviously, if we *can* know something in a given way, it doesn't follow that we *do* know it that way (Jenkins 2008). We may know it another way, or not at all. The sort of common-sense knowledge of metaphysical modality invoked by Kripke seems not to use any elaborate formal reasoning—though his account of our knowledge that the table is necessarily not made of ice does posit a subtle deductive step we were probably unaware of taking.

There is a more reasonable line of thought. (1$\Box$)–(3$\Diamond$) show that the cognitive systems with which we gain ordinary knowledge of counterfactual connections are sensitive to the right factors to provide knowledge of metaphysical modality. It is therefore a very uneconomical hypothesis that our actual ordinary knowledge of metaphysical modality comes from some quite different cognitive system. The default hypothesis is that what we have identified as capable of doing the job is what does do the job. It is open to anyone to argue for an alternative explanation, but they had better specify what the alternative cognitive system is, how it can generate ordinary knowledge of metaphysical modality, and why it is natural for us to have such a system; otherwise, the default remains in place. At present, the abductive evidence arguably favours a broadly counterfactual approach to common-sense knowledge of metaphysical modality, by inference to the best explanation.[4]

Connections between the epistemology of metaphysical modality and the epistemology of counterfactuals are close to the surface of 'Naming and Necessity'. For instance, when discussing the idea of transworld identification, Kripke writes (1972: 271, 1980: 50):

> given certain counterfactual vicissitudes in the history of the molecules of a table, *T*, one may ask whether *T* would exist, in that situation, or whether a certain bunch of molecules, which in that situation would constitute a table, constitute the very same table *T*.

Again, discussing the necessity of origin, he says: 'it seems to me that anything coming from a different origin would not be this object' (1972: 314, 1980: 113). We can formalize the claim as a universally quantified counterfactual, with **o** for 'this object', **D** for 'comes from a different origin', and quantification over individuals:[5]

ORIGIN1 $\quad \forall\mathbf{x}\,(\mathbf{Dx} \mathbin{\Box\!\!\rightarrow} \mathbf{x} \neq \mathbf{o})$

Instantiating **o** for **x** gives:

ORIGIN2 $\quad \mathbf{Do} \mathbin{\Box\!\!\rightarrow} \mathbf{o} \neq \mathbf{o}$

Since the consequent is absurd, a version of (1$\Box$) yields:

ORIGIN3 $\quad \Box\neg\mathbf{Do}$

This object could not have come from a different origin.

So far, the discussion has been deliberately neutral over the epistemology of ordinary knowledge of counterfactuals, in order to bring out the structure of the position more clearly. But of course the abductive case for the counterfactual-based approach involves a more substantive epistemology of that ordinary knowledge. The final section sketches such an epistemology and its application to ordinary knowledge of metaphysical modality.

17.3 Knowledge of counterfactuals and knowledge of metaphysical modality[6]

The normal human way to evaluate a conditional with antecedent **A** and consequent **C** is by supposing **A** and then evaluating **C** under that supposition.[7] To do so, one develops the supposition 'offline' in ways that resemble updating on the new information **A** 'online', though with significant differences. By default one uses background information and forms of reasoning, except where **A** somehow makes them inappropriate (online updating sometimes involves belief revision, too). The development sometimes involves deductive reasoning, but it is often an imaginative exercise, though still reality-oriented. What would happen if the boss had a nervous breakdown? To a first approximation, if one accepts **C** under the supposition, one accepts the conditional; if one rejects **C** under the supposition, one rejects the conditional. If one suspends judgment on **C** under the supposition, one may suspend judgment on the conditional, pleading ignorance. Alternatively, if one regards ignorance as not responsible for one's failure to decide **C**, one may treat **C** and $\neg$**C** as each rejected under the supposition, and reject each corresponding conditional, too.[8] If much is at stake, one may repeat the procedure several times, hoping to achieve a robust outcome.

The process of development is subtly different for indicative and subjunctive conditionals. The indicative 'I am thinner than I am' is an outright inconsistent supposition. The indicative 'I was thinner than I am' with an ordinary past tense 'was' is an unproblematic supposition, and indeed true: I have put on weight. But the supposition 'I was thinner than I am' also has a 'subjunctive' reading, on which 'was' takes us not to a past situation but to a counterfactual alternative to the present situation. Both readings of 'I was thinner than I am' require one to keep track of two contrasted situations from within the supposition. The subjunctive reading is the relevant one for what we have called 'counterfactual' conditionals.

Supposing **A** and then developing the counterfactual supposition is a 'more a priori' procedure than learning **A** from experience and then updating on it. Even so, evaluating ordinary counterfactuals is typically a posteriori: 'If you flicked that switch, the heating would go off'. But, unsurprisingly, by putting enough into **A** and not too much into **C**, one can construct cases where one can get from **A** to **C** by what looks like purely a priori reasoning. That is just an extreme case of a tendency built into the suppositional procedure.

How does all this apply to metaphysical modality? Consider the Kripke-inspired example ORIGIN1–ORIGIN3. One makes the counterfactual supposition **Dx** ('**x** comes from a different origin [from its actual one]'), where '**x**' works like an arbitrary name; one distinguishes the counterfactual situation from the actual one. One adds $\mathbf{x} \neq \mathbf{o}$ under the counterfactual supposition. One therefore accepts the open conditional $\mathbf{Dx} \mathrel{\square\!\!\rightarrow} \mathbf{x} \neq \mathbf{o}$. Since '**x**' was arbitrary, one also accepts its universal generalization ORIGIN1. One then derives ORIGIN2 and ORIGIN3 as before. But there was really no need to go through the counterfactual conditionals. Having added $\mathbf{x} \neq \mathbf{o}$ under the counterfactual supposition **Dx**, one already has what it takes for accepting ORIGIN3 itself.

What entitled us to add $\mathbf{x} \neq \mathbf{o}$ to the counterfactual supposition in the first place? If adding $\mathbf{x} \neq \mathbf{o}$ requires already knowing something about origin to be metaphysically necessary, we have

made little or no progress towards clarifying the epistemology of metaphysical modality. This is an instance of the general issue of what to hold fixed under a counterfactual supposition, Goodman's problem of cotenability. Some critics have taken it to undermine the counterfactual approach (Tahko 2012).

Obviously, for proponents of the counterfactual-based approach to posit a black box for assessing metaphysical modal status as a separate module within the suppositional procedure would be to give the game away. But they have no need of such a black box.

An analogy may help here. Imagine a philosopher arguing that we must have innate knowledge of epistemology because, unless we *already* know some epistemology, we can't acquire any knowledge, since we don't know which cognitive methods are epistemologically sound. His argument is clearly bogus. Knowing nothing about epistemology, one can acquire knowledge by unreflectively employing cognitive methods that are in fact epistemologically sound. Many non-human animals do just that. Later, we may reflect on our cognitive methods, and—by using them!—acquire knowledge of epistemology, perhaps even of the soundness of those very methods. Often, practice precedes theory. Such knowledge will not satisfy committed sceptics, but since nothing will satisfy them, that test is unreasonable.

Similarly, knowing nothing about metaphysical modality, we can acquire knowledge about counterfactual connections by unreflectively employing normal human methods for assessing counterfactual conditionals. Later, we may use them to make and apply the distinction between metaphysical necessity and metaphysical contingency, and reflect that they commit one to applying the distinction in some ways rather than others—just as normal human methods for acquiring perceptual knowledge commit one to classifying objects by shape in some ways rather than in others.

There are general patterns in what we hold fixed and what we don't when developing a counterfactual supposition, depending on what the supposition is. One manifestation of those patterns is that developing some suppositions yields a contradiction, while developing others does not. We classify the former suppositions as metaphysically impossible, the latter as metaphysically possible. Obviously, that does not mean that judgments of metaphysical necessity were inputs to the process. It does not even mean that we have, as it were, a separate list in our brains of the parameters to be held fixed between actual and counterfactual situations no matter what supposition is being developed.

Here is another analogy. Imagine a deductive reasoner who starts with a hard-wired system of natural deduction, with standard introduction and elimination rules for the usual logical connectives. Those rules concern reasoning with arbitrary sets of auxiliary assumptions (side premises). Nevertheless, the rules allow some conclusions to be reached from the empty set of assumptions. Those conclusions are the logical truths. Obviously, that does not mean that judgments of logical truth were inputs to the reasoning. It does not even mean that some logical truths play a special role in the rules, presumably as axioms: there are no axioms in a standard natural deduction system (say, for first-order logic without identity). Recognizing logical truths is just a by-product of a system mainly designed for recognizing the logical consequences of non-logical assumptions. Likewise, recognizing metaphysically necessities may be just by-products of a system mainly designed for recognizing the counterfactual consequences of metaphysically contingent suppositions.

Of course, our theoretical understanding of the imaginative process of developing a counterfactual supposition is much less clear than our theoretical understanding of formal systems of natural deduction. We have only a rough idea of the general patterns. But, given the practical constraints under which those patterns evolved, it would be surprising if they reserved special treatment for the practically unimportant class of metaphysical necessities. Their treatment is

likely to be a by-product of more general patterns. That would fit well with the observation in Section 17.2 that classifications of metaphysical modality are logical corollaries of assessments of ordinary counterfactual connections, in cases (i)–(iv).

Similar comments apply to talk of conceivability and inconceivability as guides to metaphysical possibility and impossibility. To treat such conceivability as an isolated *sui generis* mental phenomenon is psychologically implausible. 'Conceiving' is better understood as merely one aspect of the far broader phenomenon of imagining. A good account of conception should emerge as a corollary of a more general theory of imagination.

It has been questioned whether the imaginative process of developing counterfactual suppositions is really needed for evaluating ascriptions of metaphysical modality (Yli-Vakkuri 2013). Sometimes, we can reach the required conclusions by purely deductive reasoning from non-modal premises about identity and distinctness. For the substitution of one atomic term for another with the same reference in the context of a non-epistemic counterfactual conditional should leave its truth-value intact, since counterfactual connections depend on the objects, properties, and relations at issue, not on how they are described.[9] Thus, from $\mathbf{x = y}$ and the trivial truth $\mathbf{x \neq y \mathrel{\Box\!\!\rightarrow} x \neq y}$, we can derive $\mathbf{x \neq y \mathrel{\Box\!\!\rightarrow} x \neq x}$ by substitution, which corresponds to a standard proof of the necessity of identity (compare (1□)). Yli-Vakkuri argues that we can take this process much further, using the necessity of property identity. For instance, a proof in his style of ORIGIN3 (□¬**Do**) starts from the property identity $\mathbf{\lambda x(x = o) = \lambda x(x = o \,\&\, \neg Do)}$ (the property of being this object is the property of both being this object and not coming from a different origin from its actual one). If the properties are identical, they are necessarily identical, so necessarily coextensive, so it is impossible for **o** to come from a different origin. However, the property identity claim is not immediately compelling. The natural human way to test it is by asking oneself whether this object could have come from a different origin, and imaginatively developing the counterfactual supposition that it *did* come from a different origin. Kripke's suggestion that 'anything coming from a different origin would not be this object' (ORIGIN1) is just a slight variation on this theme: it answers the original question by making a counterfactual supposition (**Dx**) overtly neutral on the identity of the thing, but when developed it takes us to the conclusion that the thing is not this object ($\mathbf{x \neq o}$). In practice, without such an imaginative exercise, however rudimentary, we seem unable to access the identity of the properties. Although perception, induction, natural science, and logic also contribute to our knowledge of objective modality, it is very doubtful that they can ever fully take over the cognitive role of the imagination.

Intriguing questions also arise as to how we know the *logic* of counterfactuals and metaphysical modality. Arguably, an abductive methodology is needed to assess candidate laws (Williamson 2013a: 423–429). Such a methodology needs evidence to work on, for instance in the shape of non-general truths expressed with the counterfactual conditional and modal operators. We use our imaginations in generating such data.

There is no reason to limit the cognitive reach of the imagination to counterfactual conditionals and their equivalents. We can also use it to gain knowledge of necessity and possibility in more restricted senses (Williamson 2016c). For instance, through an imaginative exercise, one might work out whether it is possible to get into the bank vaults undetected: metaphysical possibility is not the modality of interest. Facts about both such restricted modalities may in turn have implications for metaphysical modality (Williamson 2017). But counterfactual conditionals are of special relevance to metaphysical modality given the logical relations between them (Section 17.2). By studying the epistemology of counterfactual conditionals, we study a cognitive system of major practical utility that also has the power to draw a key metaphysical distinction. Although psychology is no substitute for epistemology, in the long run, a deeper

psychological understanding of our use of the imagination in assessing counterfactuals may help us to understand its use in assessing ascriptions of necessity and possibility, without resort to fantasies of surveying worlds with the mind's eye.[10]

Notes

1 Williamson (2013b) argues that the distinction between a priori and a posteriori knowledge is epistemologically superficial, on grounds related to the argument of Section 17.3. For brevity, to relate the discussion to Kripke's work, I follow him in using the distinction.

2 Kment (2006) and Berto, French, Priest, and Ripley (2018) discuss accounts of metaphysical modality in terms of counterfactuals that do not assume NECESSITY. Readers should note that the account in Williamson (2007) was written with something like the semantics of counterfactuals given by Lewis (1973) in mind. Much more recently, I have come to the conclusion that counterfactual conditionals are contextually restricted strict conditionals, in English compositionally derived from the material conditional expressed by 'if' and a contextually restricted necessity operator expressed by 'would' (Williamson 2020). This has non-trivial consequences for the logic of counterfactuals. Fortunately, the differences do not impact very much on the epistemological issues discussed in this chapter, though more of the work has to be done by pragmatic constraints on the contextual restriction, rather than by purely logical principles; for the relation between the two accounts with respect to the modal aspect of thought experiments, see Williamson (2020: 229–241).

3 The principle of Conditional Excluded Middle (valid in Stalnaker's logic for counterfactuals, invalid in Lewis's) asserts the disjunction of $\mathbf{A} \mathbin{\Box\!\!\rightarrow} \mathbf{C}$ and $\mathbf{A} \mathbin{\Box\!\!\rightarrow} \neg\mathbf{C}$ and so excludes case (iv) (though it may be held to be 'indeterminate' which disjunct holds). The principle has been coming back into favour (Williams 2010). However, the data in its favour may come from over-reliance on a fallible heuristic (Williamson 2017 and 2020).

4 Kroedel (2012) argues that psychological evidence favours an account more closely related to $(3\Box)$–$(3\Diamond)$ than to the other two pairs.

5 Subtle issues arise over the interaction of quantifiers and conditionals (Williamson 2007: 305–308, 2013a: 127–128).

6 For more details of the account defended here see Williamson (2005, 2007: 141–178, 2020) and, for the connection with imagination, Williamson (2016c). Hill (2006) makes a similar connection between the epistemology of counterfactuals and the epistemology of metaphysical modality.

7 See Evans and Over (2004) and Williamson (2020) for suppositional accounts of the psychology of conditionals.

8 If rejecting a conditional involves accepting its negation, one thereby violates Conditional Excluded Middle.

9 Opponents of NECESSITY may resist this claim; see Section 17.2 for discussion.

10 Further challenges to the counterfactual-based approach and responses include Peacocke (2011), Williamson (2011), Vetter (2016), and Williamson (2016b).

References

Berto, F., French, R., Priest, G., and Ripley, D. (2018) 'Williamson on counterpossibles', *Journal of Philosophical Logic*, 47: 693–713.

Brogaard, B., and Salerno, J. (2013) 'Remarks on counterpossibles', *Synthese*, 190: 639–660.

Edgington, D. (2008) 'Counterfactuals', *Proceedings of the Aristotelian Society*, 108: 1–21.

Evans, J., and Over, D. (2004) *If*, Oxford: Oxford University Press.

Gregory, D. (2017) 'Counterfactual reasoning and knowledge of possibilities', *Philosophical Studies*, 58: 821–835.

Hill, C. (2006) 'Modality, modal epistemology, and the metaphysics of consciousness', in S. Nichols (ed.), *The Architecture of Imagination*, Oxford: Oxford University Press.

Jenkins, C. (2008) 'Modal knowledge, counterfactual knowledge and the role of experience', *Philosophical Quarterly*, 58: 693–701.

Kment, B. (2006) 'Counterfactuals and the analysis of necessity', *Philosophical Perspectives*, 20: 237–302.

———. (2014) *Modality and Explanatory Reasoning*, Oxford: Oxford University Press.

Kratzer, A. (1977) 'What "must" and "can" must and can mean', *Linguistics and Philosophy*, 1: 337–355.
Kripke, S. (1971) 'Identity and necessity', in M. Munitz (ed.), *Identity and Individuation*, New York: New York University Press.
———. (1972) 'Naming and necessity', in D. Davidson and G. Harman (eds.), *Semantics of Natural Language*, Dordrecht: Reidel.
———. (1980) *Naming and Necessity*, Oxford: Blackwell.
Kroedel, T. (2012) 'Counterfactuals and the epistemology of modality', *Philosophers' Imprint*, 12.
Lewis, D. (1973) *Counterfactuals*, Oxford: Blackwell.
Lowe, E.J. (2012) 'What is the source of our knowledge of modal truths?' *Mind*, 121: 919–950.
McCullagh, M., and Yli-Vakkuri, J. (eds.) (2017) *Williamson on Modality*, London: Routledge.
Nolan, D. (1997) 'Impossible worlds: a modest approach', *Notre Dame Journal of Formal Logic*, 38: 535–572.
Peacocke, C. (2011) 'Understanding, modality, logical operators', *Philosophy and Phenomenological Research*, 82: 472–480.
Portner, P. (2009) *Modality*, Oxford: Oxford University Press.
Roca-Royes, S. (2011) 'Modal knowledge and counterfactual knowledge', *Logique et Analyse*, 54: 537–552.
———. (2016) 'Similarity and possibility: an epistemology of *de re* possibility for concrete entities', in B. Fischer and F. Leon (eds.), *Modal Epistemology after Rationalism*, New York: Springer.
Stalnaker, R. (1968) 'A theory of conditionals', *American Philosophical Quarterly Monographs* 2: 98–112.
Strohminger, M. (2015) 'Perceptual knowledge of nonactual possibilities', *Philosophical Perspectives*, 29: 363–375.
Tahko, T. (2012) 'Counterfactuals and modal epistemology', *Grazer Philosophische Studien*, 86: 93–115.
Vetter, B. (2016) 'Williamsonian modal epistemology, possibility based', *Canadian Journal of Philosophy*, 46: 766–795, and in McCullagh and Yli-Vakkuri (2017).
Williams, J.R.G. (2010) 'Defending conditional excluded middle', *Noûs*: 44: 650–668.
Williamson, T. (2005) 'Armchair philosophy, metaphysical modality and counterfactual thinking', *Proceedings of the Aristotelian Society*, 105: 1–23.
———. (2007) *The Philosophy of Philosophy*, Oxford: Blackwell.
———. (2011) 'Reply to Peacocke', *Philosophy and Phenomenological Research*, 82: 481–487.
———. (2013a) *Modal Logic as Metaphysics*, Oxford: Oxford University Press.
———. (2013b) 'How deep is the distinction between a priori and a posteriori knowledge?' in A. Casullo and J. Thurow (eds.), *The A Priori in Philosophy*, Oxford: Oxford University Press.
———. (2016a) 'Modal science', *Canadian Journal of Philosophy*, 46: 453–492, and in McCullagh and Yli-Vakkuri (2017).
———. (2016b) 'Reply to Vetter', *Canadian Journal of Philosophy*, 46: 796–802, and in McCullagh and Yli-Vakkuri (2017).
———. (2016c) 'Knowing and imagining', in A. Kind and P. Kung (eds.), *Knowledge through Imagination*, Oxford: Oxford University Press.
———. (2017) 'Counterpossibles in semantics and metaphysics', *Argumenta*, 4.
———. (2020) *Suppose and Tell: The Semantics and Heuristics of Conditionals*, Oxford: Oxford University Press.
Yli-Vakkuri, J. (2013) 'Modal scepticism and counterfactual knowledge', *Philosophical Studies*, 162: 605–623.

Chapter 18

MODALITY AND A PRIORI KNOWLEDGE

Albert Casullo

The prominence of questions regarding the relationship between a priori knowledge and the modalities in contemporary philosophy is largely due to the influence of Immanuel Kant and Saul Kripke. Although Kant's primary focus is on epistemological questions that pertain to the nature and existence of a priori knowledge, and Kripke's primary focus is on metaphysical and semantic issues that pertain to the modalities, the relationship between a priori knowledge and the modalities plays a prominent role in the work of both. As we shall see, many of the contemporary discussions take as their point of departure either particular views of one of them or apparent disagreements between them.

Kant makes three noteworthy contributions to the contemporary discussion. He introduces a conceptual framework that involves three distinctions: the epistemic distinction between a priori and a posteriori (or empirical) knowledge; the metaphysical distinction between necessary and contingent propositions; and the semantic distinction between analytic and synthetic propositions. Within this framework, Kant poses four questions: What is a priori knowledge? Is there a priori knowledge? What is the relationship between the a priori and the necessary? Is there synthetic a priori knowledge? Kant's distinctions, the questions that he poses, and his responses to them remain at the center of the contemporary discussion.

Kant (1965: 43) maintains that a priori knowledge is absolutely independent of all experience, whereas a posteriori knowledge is possible only through experience. Kant approaches his second question indirectly by seeking criteria, or sufficient conditions, for the a priori. Although he offers two such criteria, necessity and strict universality, his primary argument for the a priori appeals to the first. Kant's claim that necessity is a criterion of the a priori commits him to the following thesis about the relationship between the a priori and the necessary:

(K1) All knowledge of necessary propositions is a priori.

He also appears to endorse:

(K2) All propositions known a priori are necessary.

Although Kant is often portrayed as holding that the categories of the a priori and the necessary are coextensive, the conjunction of (K1) and (K2) do not support that attribution since it does

not entail that all necessary propositions are known or knowable a priori. According to Kant, all propositions of the form "All A are B" are either analytic or synthetic: analytic if the predicate is contained in the subject, synthetic if it is not. Here he contends that

(K3) All knowledge of analytic propositions is a priori, and
(K4) Some propositions known a priori are synthetic.

In support of (K4), Kant argues that the predicate terms of "7 + 5 = 12" and "The straight line between two points is the shortest" are not covertly contained in their respective subjects.

Kripke (1980: 34) contends that, in contemporary discussion, the terms "a priori," "analytic," and "necessary" are often used interchangeably. He credits Kant with distinguishing between "a priori" and "analytic." His focus is on the terms "a priori" and "necessary." Here he stresses, following Kant, that the former expresses an epistemological concept, and that the latter expresses a metaphysical concept. Moreover, he stresses that it is a substantive philosophical thesis, as opposed to a matter of definitional equivalence, that the two concepts are coextensive. Apart from forcefully defending Kant's conceptual framework and stressing that answers to Kant's third and fourth questions require substantive philosophical argument, Kripke makes two noteworthy contributions to the contemporary discussion. First, he rejects (K1) and (K2) by arguing that some necessary propositions are known a posteriori and that some contingent propositions are known a priori. Second, Kripke distinguishes between belief in *de dicto* modality (the view that propositions are necessarily or contingently true or false) and belief in *de re* modality (the view that an object's possession of a property is necessary or contingent), and defends the latter. Putative examples of *de re* modal knowledge underwrite his case against (K1).

18.1 What is a priori knowledge?

Kant's conception of a priori knowledge is not fully perspicuous, since he allows that such knowledge can depend on experience in some respects, but is not explicit about the respect in which it must be independent of experience. Most contemporary theorists agree that a priori knowledge is knowledge whose *justification* does not depend on experience. They endorse the following articulation of Kant's conception of a priori knowledge:

(KAP) S knows a priori that p if and only if S's justification for the belief that p does not depend on experience and the other conditions on knowledge are satisfied.

Since (KAP) does not entail that necessity is constitutive of the a priori, it does not directly underwrite either (K1) or (K2). They require independent support.

There is a remaining source of controversy among contemporary proponents of (KAP), which pertains to the sense of "dependence" relevant to a priori justification. Some, such as Casullo (2003), endorse:

(AP1) S knows a priori that p if and only if S's belief that p is nonexperientially justified (i.e., justified by some nonexperiential source), and the other conditions on knowledge are satisfied.

Others, such as Philip Kitcher (1983), favor:

(AP2) S knows a priori that p if and only if S's belief that p is nonexperientially justified and cannot be defeated by experience, and the other conditions on knowledge are satisfied.

Other contemporary theorists, such as BonJour (1985), reject (KAP) in favor of the traditional rationalist conception of a priori knowledge:

(RAP) S knows a priori that p iff S's justification for the belief that p consists in intuitively "seeing" (directly or indirectly) that p is necessarily true, and the other conditions on knowledge are satisfied.

Since (RAP) entails that necessity is constitutive of the a priori, it directly supports (K2) but not (K1).

Casullo (2003) contends that both (RAP) and (AP2) face serious objections. Kripke (1980) and Kitcher (1983) hold that an adequate conception of a priori knowledge should allow for the possibility that one can know empirically some proposition that one knows a priori. (AP2) precludes this possibility. Suppose, for example, that S knows a priori some mathematical proposition that p and that S can know a posteriori that p. It follows that S's belief that p can be justified either a priori or a posteriori to a degree that is sufficient for knowledge. The empirical sources that have been alleged to justify mathematical propositions, such as counting objects or consulting a mathematician, are defeasible by empirically justified overriding defeaters. Hence, it follows that if S's belief that p is justifiable by experience, then S's belief that not-p is also justifiable by experience. But, according to (AP2), if S's belief that p is justified a priori, then S's belief that not-p is not justifiable by experience. Hence, (AP2) entails that if S's belief that p is justified a priori, then it is not justifiable by experience.

(RAP) faces a different objection. If we assume that the metaphorical sense of "see" shares with the literal sense the following feature: "S sees that p" entails "S believes that p," then (RAP) leads to a problematic regress. It entails that if S's belief that p is justified a priori, then S believes that necessarily p. Either S's belief that necessarily p is justified or not. If not, it is hard to see why it is a necessary condition of having an a priori justified belief that p. If it is, then presumably it is justified a priori. But in order for S's belief that necessarily p to be justified a priori, S must believe that necessarily necessarily p, and the same question arises with respect to the latter belief. Is it justified or not?

(AP1) also faces a challenge. Some theorists have questioned the significance of the a priori–a posteriori distinction. Williamson (2007) distinguishes two roles that experience can play in the acquisition of knowledge: enabling and evidential. According to (AP1), a priori knowledge is incompatible with an evidential role for experience, but it is compatible with an enabling role for experience. He contends that, on his account of knowledge of counterfactuals, experience can play a role that is neither purely enabling nor strictly evidential, and as a consequence, such knowledge is not happily classified as either a priori or a posteriori. Williamson (2013) distinguishes two approaches to introducing the a priori–a posteriori distinction and argues that the distinction introduced by both is superficial.[1]

18.2 Is there a priori knowledge?

(K1) is the leading premise of Kant's primary argument for the existence of a priori knowledge:

(K1) All knowledge of necessary propositions is a priori.
(K5) Mathematical propositions, such as that 7 + 5 = 12, are necessary.
(K6) Therefore, knowledge of mathematical propositions, such as that 7 + 5 = 12, is a priori.

Kant's argument remains central to the contemporary discussion since it is frequently cited and endorsed by proponents of the a priori.

In order to evaluate Kant's argument, we must first recognize that the expression "knowledge of necessary propositions" masks the following important distinctions:

(A) S knows the *truth-value* of p just in case S knows that p is true, or S knows that p is false.
(B) S knows the *general modal status* of p just in case S knows that p is a necessary proposition (i.e., necessarily true or necessarily false), or S knows that p is a contingent proposition (i.e., contingently true or contingently false).
(C) S knows the *specific modal status* of p just in case S knows that p is necessarily true, or S knows that p is necessarily false, or S knows that p is contingently true, or S knows that p is contingently false.

(A) and (B) are logically independent. One can know that Goldbach's Conjecture is a necessary proposition without knowing whether it is true or false. Alternatively, one can know that the Pythagorean Theorem is true without knowing whether it is a necessary proposition or a contingent proposition. Since the specific modal status of a proposition is just the conjunction of its truth value and its general modal status, one cannot know the specific modal status of a proposition unless one knows both its general modal status and its truth value.

Returning to Kant's argument, we can now distinguish two readings of (K1):

(K1T) All knowledge of the *truth-value* of necessary propositions is a priori.
(K1G) All knowledge of the *general modal status* of necessary propositions is a priori.

The argument is valid only if (K1) is read as (K1T). Kant (1965: 43), however, supports the contention that necessity is a criterion of the a priori with the observation that although experience teaches us that something is so and so, it does not teach us that it cannot be otherwise. Taken at face value, this observation states that although experience teaches us that a proposition is true, it does not teach us that it is necessary. Hence, Kant's contention supports (K1G), but not (K1T). Consequently, either Kant's argument is invalid or its leading premise is unsupported.

Recent proponents of the a priori, such as Chisholm (1966), have offered a reformulation of Kant's argument:

(K1G) All knowledge of the *general modal status* of necessary propositions is a priori.
(K5) Mathematical propositions, such as that $7 + 5 = 12$, are necessary.
(K7) Therefore, knowledge of the *general modal status* of mathematical propositions is a priori.

This version of Kant's argument circumvents the problem faced by the original version and can be generalized to knowledge of the general modal status of any necessary proposition.

Radical empiricism is the view that denies the existence of a priori knowledge. Many contemporary radical empiricists draw their inspiration from Quine's (1963) classic paper, "Two Dogmas of Empiricism." The bearing of the paper's argument on the existence of a priori knowledge remains controversial. The primary target of the paper is a variant of Frege's conception of analyticity: an analytic statement is one that is reducible to a truth of logic by replacing synonyms with synonyms. There are two strands to Quine's argument: (1) synonymy cannot be explained in terms of definition, interchangeability *salve veritate*, or semantic rules; and (2) the verification theory of meaning does provide an account of statement synonymy; but the theory

presupposes radical reductionism, which is a failed program. Quine, however, contends that a vestige of that program survives in the view that individual statements admit of confirmation or disconfirmation. This view, he maintains, lends credence to the idea that there are statements confirmed no matter what, which Quine (1963: 43) rejects on the grounds that "no statement is immune to revision." Neither strand of Quine's argument is explicitly directed at a priori knowledge. The first challenges the cogency of semantic concepts such as synonymy. The second challenges the remaining vestige of reductionism. Hence, if Quine's argument does present a challenge to a priori knowledge, some additional premise is necessary that connects one of its explicit targets to the a priori.

The traditional reading of the argument is that Quine's goal is to undermine the central tenet of logical empiricism, (LE), by showing that the analytic–synthetic distinction is not cogent:

(LE) All a priori knowledge is of analytic truths.

Suppose we grant that Quine's goal is to undermine (LE) and that his arguments establish that the analytic–synthetic distinction is not cogent. It follows that (LE) is incoherent. But it does not follow that the claim of logical empiricists that there is a priori knowledge is incoherent since they do not take (LE) to be constitutive of the concept of a priori knowledge, and they do not base their case for the existence of a priori knowledge on a premise, such as (LE), that involves the concept of analytic truth.[2]

One might suggest, in defense of Quine, that although the concept of a priori knowledge does not include the concept of analytic truth explicitly, it does include it implicitly. The only plausible case for maintaining that the concept of a priori knowledge implicitly involves the concept of analytic truth is based on two premises: (1) the concept of necessary truth is constitutive of the a priori, and (2) the concept of necessary truth is analyzable in terms of the concept of analytic truth. Both premises are problematic since (KAP) does not entail that necessity is constitutive of the a priori, and there is no available analysis of the concept of necessary truth in terms of the concept of analytic truth.

Some champions of "Two Dogmas" propose an alternative reading of Quine's argument. Putnam (1983) maintains that the two strands of Quine's argument are directed toward two different targets. The first is directed toward the semantic concept of synonymy, but the second is directed toward the concept of a statement that is confirmed no matter what, which is a concept of apriority. Kitcher (1983: 80) endorses this reading of Quine's argument: "If we can know a priori that *p* then no experience could deprive us of our warrant to believe that *p*." This reading commits Quine to (AP2), which was rejected in Section 18.1.

18.3 What is the relationship between the a priori and the necessary?

Kripke (1980: 100–16) rejects (K1) by offering examples of necessary truths that are alleged to be known a posteriori. His two initial examples involve statements in which an essential property is attributed to some particular physical object and identity statements involving co-referential proper names. He later extends his account of identity statements to theoretical identity statements. Our discussion will focus on his initial examples. Let "a" be the name of a particular desk and F be the property of being made of wood. Suppose that someone knows that Fa—i.e., that this desk is made of wood. Such knowledge is a posteriori. Yet, if Fa is true, it is necessarily true since F is an essential property of a. In any possible world in which a exists, a is F. Similar observations apply to Kripke's example of identity statements involving proper names. Since, according to Kripke, ordinary proper names, such as "Hesperus" and "Phosphorus," are rigid designators, each picks out the same object in all possible worlds in which it picks out any object. Therefore, if both pick out the same object in the actual world, both pick out the same

object in all possible worlds in which they pick out any object. Hence, if "Hesperus is Phosphorus" is true, it is necessarily true. Yet, if someone knows that it is true, such knowledge is a posteriori since it is an astronomical discovery that Hesperus is Phosphorus.

How do Kripke's examples bear on (K1)? Once again, we must distinguish between (K1T) and (K1G). Kripke's examples, if cogent, establish that (K1T) is false. They establish that one knows a posteriori that some necessary propositions are *true*. They establish that one knows that this desk is made of wood and that Hesperus is Phosphorus. Kripke's examples, however, do not establish that (K1G) is false. They do not establish that one knows a posteriori that some necessary propositions are *necessary*. Moreover, Kripke denies that such knowledge is a posteriori. In both cases, Kripke (1971: 153, 1980: 109) is explicit in maintaining that we know by "a priori philosophical analysis" that if the statement in question is true, then it is necessarily true. Hence, he does not take his examples to be cases of a posteriori knowledge of the *general modal status* of necessary propositions. Instead, Kripke (1971: 153) offers a general inferential model for understanding knowledge of a posteriori necessities, where one knows (Kr3) by inference from (Kr1) and (Kr2):

(Kr1) If P, then necessarily P.
(Kr2) P.
(Kr3) Therefore, necessarily P.

On Kripke's model, one knows (Kr1) a priori and (Kr2) a posteriori. Since knowledge of (Kr3) is based on knowledge of (Kr2), which is a posteriori, knowledge of (Kr3) is also a posteriori.

It may appear that Kripke's conclusion that one has a posteriori knowledge that *necessarily* P entails that (K1G) is false. Here we must distinguish between (K1G) and:

(K1S) All knowledge of the *specific modal status* of necessary propositions is a priori.

Kripke's examples establish that (K1S) is false: They establish that one knows a posteriori that some necessary propositions are *necessarily true*. Since knowledge of the specific modal status of a proposition is the conjunction of knowledge of its general modal status and knowledge of its truth value, it follows from the fact that one's knowledge of the truth value of P is a posteriori that one's knowledge of its specific modal status is also a posteriori. However, from the fact that one's knowledge of the specific modal status of P is a posteriori, it does not follow that one's knowledge of its general modal status is also a posteriori.

In conclusion, there is a significant point of agreement and a significant point of disagreement between Kant and Kripke. Kant endorses both (K1T) and (K1G). Kripke's examples of necessary a posteriori knowledge are examples of a posteriori knowledge of the truth of necessary propositions. They are, if cogent, counterexamples to (K1T). Kripke, however, denies that his examples are those of a posteriori knowledge of the general modal status of necessary propositions. He explicitly endorses (K1G), at least with respect to the examples that he considers.

Although Kant and Kripke both endorse (K1G), we are faced with the question: Why accept (K1G)? Kant's observation that experience can teach us only what is the case appears to be rooted in the idea that experience can provide knowledge of only the actual world, but not of other possible worlds. But a good deal of our ordinary practical knowledge and the bulk of our scientific knowledge provide clear counterexamples to the claim. My knowledge that my pen will fall if I drop it does not provide information about what is the case, for its antecedent is contrary to fact. Scientific laws are not mere descriptions of the actual world. They support counterfactual conditionals and, hence, provide information beyond what is true of the actual world. Kripke offers only sketches of supporting arguments for (K1G) with respect to the particular cases that he considers. He does not offer either a general account of knowledge of

general modal status or a defense of the claim that we do not (or cannot) have a posteriori knowledge of general modal status.

Kripke (1980: 56) rejects (K2) by offering an example of a contingent truth that is alleged to be known a priori. His example is based on the observation that a definite description can be employed to fix the reference, as opposed to giving the meaning, of a term. Consider someone who employs the definition description "the length of S at t_0" to fix the reference of the expression "one meter." Kripke maintains that this person knows, without further empirical investigation, that S is one meter long at t_0. Yet the statement is contingent since "one meter" rigidly designates the length which is in fact the length of S at t_0 but, under different conditions, S would have had a different length at t_0. In reply, Plantinga (1974) and Donnellan (1979) contend that, without empirical investigation, the reference fixer knows that the sentence "S is one meter long at t_0" expresses a truth, though not the truth that it expresses.

18.4 Is there synthetic a priori knowledge?

The literature on the a priori is dominated by Kant's endorsement of

(K4) Some propositions known a priori are synthetic.

In addressing that literature, one question immediately arises: Why is the existence of synthetic a priori knowledge *epistemologically* significant? Kant maintains that analytic a priori knowledge requires only possession of the relevant concepts and the principle of contradiction, but synthetic a priori knowledge requires more. In order to know that $7 + 5 = 12$, for example, Kant (1965: 53) maintains: "We have to go outside these concepts, and call in the aid of the intuition which corresponds to one of them." The significance of (K4) is rooted in the assumption that the source of synthetic a priori knowledge is different from, and more problematic than, the source of analytic a priori knowledge.

Kant, however, does not defend this assumption. Although he maintains that knowledge of analytic propositions requires only knowledge of the principle of contradiction and the content of concepts, he does not explicitly address the source of such knowledge. Since he does not explicitly address the source of analytic a priori knowledge, Kant has no basis for claiming that the source of such knowledge is different from the source of synthetic a priori knowledge, let alone that the latter is epistemologically more problematic than the former. Consequently, the epistemological significance of (K4) is presupposed rather than established.

Reactions to (K4) fall into three broad categories. Those in the first endorse (K4) but take issue with some of Kant's examples. Those in the second reject (K4). Those in the third deny the cogency of the analytic–synthetic distinction and, a fortiori, the cogency of (K4). The epistemological import of these reactions is minimal.

Frege endorses (K4), but contends that

(F) All arithmetic truths are analytic.

Frege's (1974: 4e) defense of (F) requires a modification of Kant's definition of an analytic truth: "If, in carrying out this process [of constructing the proof of the proposition from primitive truths], we come only on general logical laws and on definitions, then the truth is an analytic one." Frege's goal is to demonstrate that all arithmetic truths can be proved from primitive truths via general logical laws and definitions. This project faces formidable technical obstacles. But even if they are overcome, such a demonstration, taken by itself, tells us little about knowledge of arithmetic truths since it is silent with respect to the issue of how one knows the primitive

truths, definitions, and logical laws employed in such proofs. In particular, such a demonstration is compatible with the claim that the truths of arithmetic are knowable only via intuition. Hence, it is of limited epistemological significance.

Ayer rejects (K4) and defends

(LE) All a priori knowledge is of analytic truths.

Ayer's (1952: 78) defense also requires a modification of Kant's definition of an analytic truth: "a proposition is analytic when its validity depends solely on the definitions of the symbols it contains, and synthetic when its validity is determined by the facts of experience." Ayer's most explicit defense of (LE) is presented in the context of discussing logical truths. Ayer (1952: 78–9) contends that the proposition "Either some ants are parasitic or none are" is analytic on the grounds that one need not resort to observation to determine that it is true: "If one knows what is the function of the words 'either,' 'or,' and 'not,' then one can see that any proposition of the form "Either *p* is true or *p* is not true" is valid, independently of experience." Here Ayer explains a priori knowledge of logical truths in terms of an ability to "see" that they are true independently of experience. The sense of "see" that he invokes is not the literal sense; it is a metaphorical sense he does not explain. Therefore, Ayer's explanation is of limited epistemological significance.

Quine's rejection of the cogency of the analytic–synthetic distinction has been widely viewed as challenging the existence of a priori knowledge. In Section 18.2, we examined two lines of argument in support of that view and concluded that both fail. There is, however, a third argument that draws its inspiration from Quine. Its leading premise poses an explanatory challenge:

(E1) If a theory of knowledge posits a category of knowledge but cannot explain how that knowledge is possible, then the theory is unacceptable.

If (E1) is conjoined with the leading assumption of logical empiricism:

(E2) The only nonmysterious explanation of how a priori knowledge is possible involves the analytic–synthetic distinction.

Then Quine's contention that the distinction is incoherent leads straightforwardly to the conclusion that

(E3) Therefore, a theory of knowledge that endorses the a priori is unacceptable.

The explanatory challenge is the primary challenge facing proponents of a priori knowledge. The epistemological import of Kant's, Frege's, and Ayer's accounts of analyticity, however, is minimal. Therefore, the explanatory challenge goes beyond the coherence of the analytic–synthetic distinction. A response to the challenge must reject (E2). Casullo (2003) offers a strategy for responding to the challenge that gives a prominent role to empirical investigation.

18.5 New developments

There are two significant developments in the discussion of knowledge and modality, one epistemological and one metaphysical. The traditional debate regarding a priori knowledge focused primarily on knowledge of the truth value of necessary propositions. The contemporary debate gives greater prominence to questions about knowledge of the general modal status of

propositions—i.e., questions about knowledge of necessity and possibility. As we saw in Section 18.3, both Kant and Kripke endorse (K1G), but neither offers a compelling supporting argument. Controversies within metaphysics, especially those regarding dualism, have raised the prominence of questions regarding knowledge of possibility.

Kripke's rejection of (K1) features statements that attribute essential properties to contingent objects. Fine (1994) distinguishes two approaches to explaining essence: the traditional Aristotelian approach in terms of real definition, and the modern approach in terms of *de re* modal attributions. He rejects the latter on the grounds that not all properties that an object has necessarily are properties that it has essentially. For example, although it is necessary that Socrates is distinct from the Eiffel Tower, it is not essential to Socrates that he be distinct from the Eiffel Tower. As a corollary, Fine maintains that metaphysical necessity should be viewed as a special case of essence.

Questions about possibility and necessity are related given the following equivalences:

(N) $\Box P \equiv \neg\Diamond\neg P$; and
(P) $\Diamond P \equiv \neg\Box\neg P$.

Questions about modality and essence are related given the following entailment:

(E) F is essential to a $\rightarrow \Box Fa$.

Nevertheless, an array of questions about the relationship between the a priori and the modalities remain open. These are the core questions that must be addressed to fully articulate that relationship.

The first two questions are *priority* questions. Is knowledge of one modality (i.e., necessity or possibility) epistemologically prior to (or more basic than) knowledge of the other? Is knowledge of essence epistemologically prior to (or more basic than) knowledge of modality? In order to fully address the priority questions, a number of *source* questions must be addressed. With respect to the first priority question, three further questions arise. Is the basic source of knowledge of necessity identical to the basic source of knowledge of possibility? An integrated theory maintains that they are identical; a divided theory maintains that they are different. A related question arises with respect to knowledge of each of the modalities. Is there a single basic source of knowledge for each modality? A single-source account maintains that there is a single source for each. A multi-source account maintains that there is more than one source for one or both modalities. When the basic sources of modal knowledge are identified, a third question can be raised with respect to each of them. Is it a priori, a posteriori, or neither?

With respect to the second priority question, there are two options to consider. For those who endorse the modern approach to understanding essence, the basic source of knowledge of essence is identical to the basic source of modal knowledge and the answer to the second priority question is negative. For those who follow Fine and reject the modern approach, the second priority question remains open. To answer it, one must first address the question: What is the basic source (or sources) of knowledge of real definition? One can then address two further questions. Is that source (or sources) different from the basic sources of knowledge of possibility and necessity, and if so, is it more basic than those sources? Is that source (or sources) a priori, a posteriori, or neither?

The final set of questions pertain to *de re* modal knowledge. Is the basic source of *de re* modal knowledge of contingent objects identical to the basic source of *de re* modal knowledge of necessary objects? This question breaks down into two questions. Is the basic source of knowledge

that a contingent object necessarily has some property identical to the basic source of knowledge that a necessary object necessarily has some property? Is the basic source of knowledge that a contingent object possibly has some property identical to the basic source of knowledge that a necessary object possibly has some property? When the basic sources of *de re* modal knowledge are identified, a third question can be raised with respect to each of them: Are the basic sources a priori or a posteriori?

Notes

1 For a general survey of challenges to the significance of the a priori–a posteriori distinction, including Williamson's, and a defense of its significance, see Casullo (2012). For other challenges to the distinction, see Casullo (2015).
2 For a more detailed discussion of the leading argument of logical empiricists in support of the existence of a priori knowledge, see Casullo (2003).

References

Ayer, A. J. (1952) *Language, Truth and Logic*, New York: Dover Publications, Inc.
BonJour, L. (1985) *The Structure of Empirical Knowledge*, Cambridge, MA: Harvard University Press.
Casullo, A. (2003) *A Priori Justification*, New York: Oxford University Press.
Casullo, A. (2012) "Articulating the A Priori–A Posteriori Distinction," in Casullo, *Essays on a Priori Knowledge and Justification*, New York: Oxford University Press.
Casullo, A. (2015) "Four Challenges to the A Priori–A Posteriori Distinction," *Synthese* 192: 2701–24.
Chisholm, R. M. (1966) *Theory of Knowledge*, 1st ed., Englewood Cliffs, NJ: Prentice-Hall, Inc.
Donnellan, K. S. (1979) "The Contingent *A Priori* and Rigid Designators," in P. French et al. (eds.), *Contemporary Perspectives on the Philosophy of Language*, Minneapolis: University of Minnesota Press.
Fine, K. (1994) "Essence and Modality," *Philosophical Perspectives* 8: 1–16.
Frege, G. (1974) *The Foundations of Arithmetic*, 2nd ed. revised, J. L. Austin (trans.), Evanston, IL: Northwestern University Press.
Kant, I. (1965) *Critique of Pure Reason*, N. K. Smith (trans.), New York: St. Martin's Press.
Kitcher, P. (1983) *The Nature of Mathematical Knowledge*, New York: Oxford University Press.
Kripke, S. (1971) "Identity and Necessity," in M. K. Munitz (ed.) *Identity and Individuation*, New York: New York University Press.
Kripke, S. (1980) *Naming and Necessity*, Cambridge, MA: Harvard University Press.
Plantinga, A. (1974) *The Nature of Necessity*, Oxford: Oxford University Press.
Putnam, H. (1983) "'Two Dogmas' Revisited," in Putnam, *Realism and Reason: Philosophical Papers*, Vol. 3, Cambridge: Cambridge University Press.
Quine, W. V. (1963) "Two Dogmas of Empiricism," in Quine, *From a Logical Point of View*, 2nd ed. revised, New York: Harper and Row.
Williamson, T. (2007) *The Philosophy of Philosophy*, Oxford: Blackwell.
Williamson, T. (2013) "How Deep is the Distinction between A Priori and A Posteriori Knowledge?" in A. Casullo and J. Thurow (eds.) *The A Priori in Philosophy*, Oxford: Oxford University Press.

Chapter 19

INTUITION AND MODALITY

a disjunctive-social account of intuition-based justification for the epistemology of modality

Anand Jayprakash Vaidya

19.1 The epistemology of modality

The philosophy of modality is the study of the metaphysics, semantics, epistemology, and logic of modal claims, such as 'it is necessary that 2 + 2 = 4', 'it is possible for France to win the 2016 UEFA CUP', 'it is possible for there to be more electrons than there actually are', 'it is essentially the case that whales are mammals', and 'it is essentially the case that 2 is the successor of 1'. The central question of the epistemology of modality is the following:

> How can we come to possess sufficient justification such that we can come to know (i) what is necessary, possible, contingent, essential, and accidental for the variety of kinds of entities there are, tables, numbers, soccer teams, etc., as well as (ii) what the modal properties are for particulars, such as Fido the dog or the table on the far side of the room?

One way to explore the central question involves reflection on the epistemology of possibility for concrete particulars. Consider claims (A) and (P), and the question (J):

(A) The cup, *c*, is actually located at *L* at time *t*.

(P) The cup, *c*, could have been located at $L^\star$ at *t*.
(J) Given (A), how can a subject S be justified in believing (P)?

There are at least seven different answers to (J).

> *Conceivability*: Even though *c* is not at $L^\star$ at *t*, S can conceive of a scenario G in which *c* is at $L^\star$ at *t*, where G is both consistent and coherent. S can derive justification for believing (P) by conceiving of G, and basing their belief in (P) on G.

Conceivability-based theories are the most developed and criticized theories in the epistemology of modality. They have a long historical lineage tracing back to medieval philosophers, such as St. Anslem, through modern philosophers, such as Descartes, and up through contemporary philosophers, such as Stephen Yablo (1993) and David Chalmers (2002).

> *Imaginability*: Even though *c* is not at $L^{\star}$ at *t*, when S imagines *c* being moved from its original position *O*, to $L^{\star}$ instead of *L*, S does not arrive at a contradiction under a robust search for one. If S bases their belief in (P) on their imaginative exercise, then S is justified in believing (P).

Imagination-based theories are the second most developed and criticized theories in the epistemology of modality. Typically, imagination-based theories are contrasted with conceivability-based theories along the dimension of sensory experience. Even though both Hume and Berkeley use the term 'conceivability', their accounts are better understood as imagination-based theories because of their empiricist accounts of the mind. Conceivability-based theories tend to be rationalist in nature. Counterfactual-based theories are classified within the realm of imagination-based theories because of their reliance on the counterfactual imagination when providing an account of the epistemology of modality. Williamson (2007) offers a counterfactual-based theory.

> *Deduction*: Even though *c* is not at $L^{\star}$ at *t*, S can deduce from justified beliefs about some of the essential properties of *c*, and the relevant details about the location $L^{\star}$, that *c* could have been at $L^{\star}$ at *t*. S can make this deduction, since the *essential* properties of *c* that S is justified in believing in are compatible with *c* being at $L^{\star}$ at *t*. If S makes this deduction, they will be justified in believing (P).

Deduction-based theories aim to identify specific modal principles from which one can derive modal knowledge through deductive inference. Hale (2013), working off of Kripke's (1971) account, has offered an engaging version of the deductive approach.

> *Theory*: Even though c is not at $L^{\star}$ at *t*, if S is justified in believing theory T, T implies that *c* could be at $L^{\star}$ at *t*, and S believes that *c* could be at $L^{\star}$ at *t* on this basis, then *S* can be justified in believing (P).

Theory-based theories focus on the role that theory selection plays in modal knowledge. The core idea is that one has a justified belief for a given modal claim only when they believe a general theory that justifies the specific modal claim. Fischer (2016) has defended an impressive and detailed theory-based theory.

> *Similarity*: Even though *c* is not at $L^{\star}$ at *t*, from S's prior observation of objects similar in relevant respects to *c*, and their actual locations and movement, S can come to be justified in believing (P).

Similarity-based theories hold that much of the much of the knowledge of possibility we have in ordinary cases derives from making an inference, relying on the uniformity of nature, from an observed entity that has a property to a relevantly similar entity possibly having the same property. Roca-Royes (2016) has developed a similarity-based theory that covers a wide range of ordinary possibility claims.

> *Perception*: even though *c* is not at $L^{\star}$ at *t*, S sees that *c* could have been at $L^{\star}$ at *t*, and on that basis is justified in believing (P).

Perception-based theories are controversial, since they push against the main rationale for investigating the epistemology of modality – the claim that perception is *categorically inappropriate*,

since perception is only a guide to the actual world, whereas modality is about necessity (what is true in all worlds) and non-actual possibility (what is true in some non-actual possible world). Perceptual theories aim to give us an account of how perception puts us in contact not only with present objects, but also possibilities for them. Perception-based theories typically do not attempt to account for all kinds of modal knowledge. Rather, they focus on specific kinds of ordinary possibility claims, such as 'I can see that the cup could have been at $L^\star$ at t, even though it is at L at t'. Margot Strohminger (2015) has developed an engaging version of a perception-based theory.

> *Intuition*: even though c is not at $L^\star$ at t, S has a non-sensory-based intuition that c could be at $L^\star$ at t when S entertains the question: could c have been at $L^\star$ at t? If S bases their belief in (P) on the non-sensory-based intuition, S will be justified in believing (P).

In the rest of this chapter I will focus on unpacking and developing an intuition-based theory that is distinct from Bealer's (2002) *modal reliabilism* (MR). Bealer's account aims to draw a connection between the epistemology of intuition, concept possession, thought experiments, and modality.

(MR) holds that there is a strong metaphysical connection between understanding a concept C and having truth-tracking intuitions about whether C applies in a given scenario. Bealer's aim is to establish that if a subject S determinately understands a concept C and cognitive conditions are ideal, then S must have truth-tracking intuitions about whether C applies in a given case D. The central question concerns how to explain a subject's failure to have truth-tracking intuitions. On (MR) there are three options: (i) the concepts are not of the right kind, (ii) cognitive conditions are not ideal, or (iii) the subject does not determinately understand the concepts. As a consequence, if the concepts are of the right kind, cognitive conditions are ideal, and the subject determinately understands her concepts, it appears difficult to explain how she could fail to have truth-tracking intuitions about C's application in a given case D. Consider the following illustrative example from Bealer (2002: 103).

> Suppose that in her journal a sincere, wholly normal, attentive woman introduces through use (not stipulation) a new term 'multigon'. She applies the term to various closed plane figures having several sides (pentagons, octagons, chiliagons, etc.). Suppose her term expresses some definite concept—the concept of being a multigon—and that she determinately understands this concept. By chance, she has neither applied her term 'multigon' to triangles and rectangles nor withheld it from them; the question has just not come up. Eventually, however, she considers it. Her cognitive conditions (intelligence, etc.) are good, and she determinately understands these concepts. Suppose that the property of being a multigon is either the property of being a closed, straight-sided plane figure, or being a closed, straight-sided plane figure with five or more sides. Then, intuitively, when the woman entertains the question, she would have an intuition that it is possible for a triangle or a rectangle to be a multigon if and only if being a multigon = being a closed, straight-sided plane figure. Alternatively, she would have an intuition that it is not possible for a triangle or a rectangle to be a multigon if and only if being a multigon = being a closed, straight-sided plane figure with five or more sides. That is, the woman would have truth-tracking intuitions. If she did not, the right thing to say would be that either the woman does not really understand one or more of the concepts involved, or her cognitive conditions are not really ideal.

Using the multigon example as a backdrop, (MR) would address our question about the cup as follows. When S is asked, 'Could *c* have been at $L^\star$ at *t*?' S will have a modal intuition about the modal property *possibly located at* $L^\star$ at *t* with respect to *c* on the condition that S is attentive, correctly possesses the appropriate concepts of a *cup*, *location*, *position*, *c itself*, and is engaged in determining the answer to the question. I now turn to some general issues in the epistemology of intuition leading up to a skeptical argument against intuition-based justification for beliefs about modality, I then sketch an alternative account of intuition-based justification for beliefs about modality.

19.2 The epistemology of intuition

In order to understand how intuition can be a source of justification for beliefs about modality, we need to look at some core questions about the epistemology of intuition, in general, and then apply those to the case of modality.

One of the best modern places to think about intuition is Chudnoff (2013). Here I offer my own pathway to an investigation over intuition. On my view, an investigation into intuition often starts with *the identification question*: what kind of mental state is *intuition*? Reductive views aim to reduce intuition talk to talk about some other kind of mental state. For example, doxastic views hold that intuition talk reduces to talk about beliefs or dispositions to believe, but there is nothing that is uniquely picked out as a mental state by *intuition*. Non-reductive views hold that intuition talk is not reducible to another kind of mental state, such as belief. Rather, intuition is a unique, *sui generis*, natural kind that has its own distinct phenomenology and cognitive role. Bealer offers a non-reductive view. Alongside the identification question is the *natural kind question*: does talk of intuition form a natural kind with a fixed set of criteria? In important recent work, Jennifer Nado (2014a) has argued that intuition talk does not form a natural kind.

After engaging these metaphysical questions, the epistemology of intuition often continues on to *the reliability question*: do intuitions derive from a reliable faculty of the mind? This question can be taken either in the sense of positing the existence of a faculty of intuition that is distinct from other faculties or in the sense of positing a set of faculties that generate intuitions, where no single faculty alone is *the faculty of intuition*.

Critically, Stacy Swain, Joshua Alexander, and Jonathan Weinberg (2008, hereafter SAW) have argued that there is evidence (deriving from studies) that shows that non-expert intuitions about concept application triggered by thought experiments are susceptible to order-embedding effects. An order-embedding effect occurs when an intuition about a thought experiment depends on the order in which the thought experiment is presented relative to other thought experiments. The basic critique is that *if intuitions are subject to order-embedding effects, then the faculty or faculties from which they arise is (are) unreliable*. Chudnoff voices a version of the worry that can be derived from SAW (2008):

> If philosophers' intuitions about thought experiments are influenced by factors that do not track the truth about their subject matter, then it is unreasonable to accord intuitive judgments expressing them an epistemically privileged role in philosophical methodology.
>
> (Chudnoff 2018: 196)

The skeptical argument that puts pressure on (MR) has three main parts. In part 1, the argument aims to establish that non-expert intuitions concerning concept application in thought experiments are not a source of justification. For example, SAW (2008) shows that non-expert

intuitions about whether a described case is a case of *knowledge* are subject to order-embedding effects, and thus intuitions about the application of the concept of knowledge are not reliable. In part 2, the argument aims to establish that non-expert intuitions concerning modality are not a source of justification because they are sufficiently similar to intuitions about the application of the concept of knowledge. The core idea is that every modal intuition, such as the intuition that it is possible for *c* to be located at $L^{\star}$ at *t*, can be generated through a question based on a description of a scenario in a way that is similar to how an intuition about the concept of knowledge can be generated from a thought experiment and a question about whether the concept of knowledge applies. In part 3, the argument aims to establish that neither expert nor non-expert intuitions about modality are a source of justification for beliefs about modality because the evidence pertaining to non-experts being susceptible to order-embedding effects also applies to experts, such as philosophers. While there is still debate over the status of the expertise defense (for example Nado 2014b, 2015; Horvath & Wiegmann 2016), it is clear that more theorizing and investigation are needed.

I now offer what I call the *Master Skeptical Argument Against Justification for Modal Beliefs Based on Intuition*. This argument is similar in kind to the one offered by Chudnoff (2018: 181–183). However, Chudnoff's skeptical argument is focused on the conclusion that philosophical methodology is unreasonable, while the present argument aims at the conclusion that intuitions about modality fail to provide sufficient justification.

(1) Non-expert intuitions about concept-application in thought experiments are susceptible to order-embedding effects, since, following SAW (2008), they have been shown to be susceptible to order-embedding effects in the case of intuitions about the application of the concept of *knowledge* across case descriptions, such as Lehrer's Truetemp case.
(2) If faculty F, which produces mental state M as an output, is susceptible to order-embedding effects, then faculty F is unreliable.
(3) If faculty F is unreliable, then M, which derives from F, cannot be taken to provide sufficient justification.

∴

(4) *Conclusion 1*: Non-expert intuitions produced through faculty F about concept application in thought experiments cannot be taken to provide sufficient justification.
(5) If non-expert intuitions produced through faculty F about concept application in thought experiments are sufficiently similar to non-expert intuitions produced through faculty G about modality, then non-expert intuitions produced through faculty G about modality are also susceptible to order-embedding effects.
(6) Non-expert intuitions produced through faculty F about concept application in thought experiments are sufficiently similar to non-expert intuitions produced through faculty G about modality.

∴

(7) *Conclusion 2*: Non-expert intuitions produced through faculty G about modality are susceptible to order-embedding effects, and, consequently, cannot be taken to provide sufficient justification.
(8) If non-expert intuitions (both about concept application in thought experiments and modality) are sufficiently similar to expert intuitions, in that there is no difference in skill or no real expertise, then expert intuitions are also susceptible to order-embedding effects.

(9) There is no preponderance of evidence that expert intuitions via some skill (both about concept application in thought experiments and modality) are superior to those of non-experts who lack that skill by definition.

∴

(10) *Conclusion 3*: Intuitions produced through faculty G about modality (both from non-experts and experts) cannot be taken to provide sufficient justification for beliefs about modality.

At (5), it is assumed that the faculty that produces intuitions about concept application is distinct from that of the faculty that produces intuitions about modality. And at (9), it is assumed that there is no significant difference between experts and non-experts in both the case of concept application and modality. The virtue of this setup is that it allows for a strategy of *dual insulation*.

On the one hand, one can insulate evidence about order-embedding effects in the case of concept application from those concerning modality. On the other hand, one can insulate the lack of skilled difference between non-experts and experts in the case of concept application from that of modality. This master argument allows one to evaluate the plausibility of the skeptical argument against the view that intuitions can provide sufficient justification for beliefs about modality.

19.3 Disjunctivism about intuition

One way to challenge the argument is to block the relevance of the evidence from order-embedding effects in the case of concept application. To do that I explore an analogy between, perception and intuition with respect to disjunctivism, and then move to a discussion of the relevance of social engagement with respect to being justified in believing a modal claim on the basis of intuition.

John McDowell (2008) articulates and defends what he calls a *disjunctive conception* of perceptual experience. He argues that the disjunctive conception of experience can provide resources for a transcendental argument against skepticism about the external world based on perceptual experience.

Disjunctivism about perceptual experience is best presented by contrasting it with the account it opposes: the highest common factor (HCF) account of experience. HCF maintains that veridical and non-veridical experiences share *a common kind of mental state*. HCF is motivated partly by the argument from illusion. McDowell's (2009) understanding of the reasoning involved in HCF can be captured as follows.

(1) If two states are *first-person-phenomenologically-indistinguishable*, FPPI, then they should be categorized as falling under a *common epistemic kind*.
(2) If two states fall under the same epistemic kind, then they provide the same warrant.
(3) Veridical and non-veridical perceptions are FPPI.

∴

(4) Veridical and non-veridical perceptions provide the same warrant.

Against HCF, McDowell presents the *disjunctive conception* of perceptual experience. His disjunctive conception involves three important theses.

(i) Perception is a capacity for knowledge
(ii) Non-veridical perceptions are metaphysically distinct from veridical perceptions.
(iii) Non-veridical perceptions do not have the same warrant as veridical perceptions.

With respect to intuition, we should note that a similar argument to that of HCF in the case of perceptual experience could be made with respect to intuitional experience. The phenomenology of intuition does not allow us to distinguish, from the inside, the difference between a veridical case and a non-veridical case. There is no difference between intuition and perception with respect to the property of being FPPI. Thus, were we to follow the other premises, we would be led to the conclusion that our intuitional experience has the same warrant in both veridical and non-veridical cases.

Given the outline of the disjunctive conception of perceptual experience offered by McDowell, the disjunctive conception of intuitional experience that I defend holds that there is a distinction between non-veridical intuitions and veridical intuitions. In order to further develop this view, I will offer responses to two critical questions.

Question 1: Disjunctivism about perception holds that in veridical cases perception puts us in contact with the external world through a relation, but that relation is not factorizable into a mental component and an external world component. How can this be true for intuitions about modality? What would the relevant objects be that we are in contact with?

Response 1: In the case of intuitions about modality, it is important to note that there are two kinds of objects that intuition can relate us to. On the one hand, if one is a strong realists about modality, and holds, like Lewis, that possible worlds are real concrete particulars just like the actual world, it follows that the objects of modal intuition will be possible worlds. On the other hand, if one is a moderate realist and holds that possible worlds are real, not like the actual world but like numbers construed as abstract objects, then the objects of modal intuition will be possible worlds construed as sets of propositions or sentences, perhaps world books. Either way, if one is a realist about modality and accepts the possible worlds framework, the objects of modal intuition are simply possible worlds.

Question 2: Disjunctivism about perception holds that in veridical cases, perception puts us in contact with the external world. At least one way in which perception does this is through a causal relation between the mind and the world. The causal relation is a necessary condition, but not a sufficient condition for perception putting us in contact with the world. How can this be true for intuitions about modality? Possible worlds are causally isolated from us when they are taken to be either concrete particulars on the model of Lewis's modal realism or abstract objects on the model of Platonism in mathematics. Thus, disjunctivism cannot make sense for intuitions about possible worlds.

Response 2: An answer to this objection requires that we point out why we are inclined to say that when A perceives x, it is in part because A bears a causal relation to x. Inquiry reveals that part of the motivation for articulating causation as a necessary condition on perception is to account for *intentionality*. When A perceives x, A's mind is directed at x. Causation is part of perception because perception is an intentional relation between the mind and particulars in the world. As a consequence, we can get a key to what is important in the case of intuition by looking at the idea that intuition is an intentional relation to abstract objects; but in the case of intuition, it is not causation that explains the intentional relation. Rather, it is some other relation that does so. At this point I admit that I face an incredulous stare similar to the one Lewis faced when he proposed his modal realism where possible worlds are concrete particulars just like our whole universe, but are causally isolated from us. I grant the force of the rhetorical question: *What else could explain intentionality other than causation?* However, the idea that intentionality can only be explained by

causation is to reduce our ontology to only those things we can bear a perception-like causal relation to. It is worth pointing out that there is a tradition, stretching back at least to Plato and Plotinus, through Descartes and Spinoza, and as recently as Edmund Husserl and Kurt Gödel, of philosophers who have not been worried about the idea that intuition can be directed at entities that we have no perception-like causal relation to.

In general, the claim that intuition cannot be directed intentionally at a possible world because there is no causal connection between a person's mind and a possible world rests on a prejudice that holds that *the only way to explain intentionality is through causal relations* between the mind and the truthmakers for the content the mind is intentionally directed at. Once we move beyond the idea that causation is the only way to explain intentionality, at least one pathway is open for a disjunctive account of intuition. Our minds can be directed at all kinds of things. Causation is only one way in which intentionality is realized. Importantly, causation is important for intentionality directed at concrete particulars, but not for non-concrete entities.

The disjunctive account of intuition, (DI), holds that *when one has an intuition experience, either they have a veridical intuitional state or they have a non-veridical appearance that is phenomenologically similar to an intuition, but because something has gone wrong it is not a genuine intuition*. This account provides for a response to the *Master Skeptical Argument Against Justification for Modal Beliefs Based on Intuition*.

With (DI) in place, one can argue that the studies in SAW (2008) *do not show* that we fail to have *genuine intuitions*, which do provide sufficient justification. Rather, those studies show that we might not be able to discriminate between pseudo-intuition and genuine intuition without the help of others. However, the fact that we cannot alone determine that an intuition experience is a genuine intuition experience from the inside, doesn't show that we cannot be confident, through further epistemically responsible behavior with others that our intuitional experience is genuine. Moreover, we could be justified in believing that we have a genuine intuition when our intuition experience is shared by a *robust* and *sufficiently diverse* set of people that allows us to inductively infer that our intuition experience is genuine.

Thus, in the case of the cup, one might argue as follows: because S, and many others that are sufficiently different from S, which S has discussed the matter with, have the intuition that the cup, c, located at L at t, could have been located at $L^{\star}$, S is justified in believing that their intuition that c could have been at $L^{\star}$ at t is genuine, and thus they are sufficiently justified in believing that c could have been at L at $t^{\star}$. That intuitions suffer generally from order-embedding effects does not preclude one from (a) having a genuine intuition, and (b) being inductively justified in believing, because of the robustness of the intuition across a diverse set of agents, that their intuition is genuine.

That is, we can have meta-evidence about our first-order evidential states, which, when conjoined with our first-order evidential state, provides us with justification for beliefs about modality. Even if the faculty of intuition for modality is unreliable, it does not follow that a given intuition about modality is not factive. Reliability is about a source; factivity is about a given instance from the source. Given that a modal intuition, such as that c could have been at $L^{\star}$ at t, could be genuine, one can be justified in believing the content of the intuition, if it is robustly shared across various parameters, even if they fail to have a reliable source.

19.4 Reliability, learnability, and ordinary vs. extraordinary modal intuition

In addition, when we look at the results of the experimental studies that suggest that intuitions are unreliable because they are subject to order-embedding effects, we need to look at an

important contrast in the way this worry is voiced concerning the notion of *unreliability*. Here I offer two distinct voicings one should consider.

On the *careless consideration voicing*, we look at unreliability through the lens of a careless person who is epistemically irresponsible when collecting evidence about matters. They appeal to intuition because they are epistemically lazy. On the *conscientious learner voicing*, we look at unreliability through the lens of a person conscientiously learning how to be a competent judge concerning matters in some domain, while not yet meeting the standard that is recognized to be relevant. On the latter conception, we are likely to say that a person can get it right for the right reasons, but that she is not yet there; hence the unreliability. We do this all the time as educators in a variety of domains. Moreover, we are inclined to hope that the person can learn how to get things right, be a reliable judge, and engage in epistemically responsible behavior. The results from SAW (2008) say nothing about these two alternative voicings. While it is important to take note of whether participants *can* be manipulated in the survey environment, it is also important to take note of what they would do after they have an intuition experience to determine whether or not it is genuine. We need to check further into what the subject does with their intuition experience in order to check whether we ought to take their intuition experience seriously. Do they seek corroboration? Of course, this might not be required for every single belief formed on the basis of intuition. But we ought also to check into a subject's epistemic behavior, when they have an evidential state.

The upshot of drawing attention to the contrast is that we should critically examine the view that holds that (i) an individual S can have a modal intuition that is genuine in the disjunctive sense, and (ii) neither S nor anyone else, S*, should take the intuition as providing more than *prima facie* justification, because there are skeptical reasons against taking the intuition more seriously, such as order-embedding effects or peer disagreement. That is, from the individual perspective, it could be true that a person has a modal intuition of the form, it is possible that *p*, because, as is the case in perception, they have the right kind of connection to an entity that registers the modal intuition as correct, even if it is not causal. However, the individual perspective can be contrasted with the social perspective where we take into further consideration the idea that the intuition is unreliable, not in the careless sense, but in the learning sense. Let me close with a further elaboration of this point. Consider the following pair of modal claims:

(Z) Zombies (physical duplicates of humans, which lack phenomenal consciousness) could exist.
(P) The cup, *c*, could have been located at $L^\star$ at *t*.

Arguably, (P) falls within the scope of what van Inwagen (1998), and others, have called ordinary modal claims, while (Z) falls within the scope of extraordinary modal claims. Upon considering both (Z) and (P), an individual, expert or not, can have a modal intuition. However, we should have different attitudes about the epistemic standing of the intuitions generated from considering these claims. While we could have genuine intuitions with respect to both (Z) and (P), (Z) is not shared across cultures and disciplines as much as (P) is. That is, (Z) fails the robustness test, while (P) passes it. In addition, whether one is an expert or not, we should treat intuitions about modality by taking into consideration their learnability.

While individuals from many different cultures and disciplinary backgrounds can have an intuition about (Z) and (P), only the subject matter of (P) is learnable in a recognizable way that enables us to understand how one can improve with respect to having an intuition about (P).

In the case of (P), one can observe relevantly similar particulars and how they are moveable in space, and they could even come to have a folk-theory of ordinary objects that underlies their intuitive reactions to modal questions about objects falling under the folk-theory.

By contrast, in the case of (Z) it is difficult to explain how one could learn to make a reliable judgment in the domain to which (Z) falls. For while we do have experience that puts us into contact with individuals who are culturally different from us. And we do have experience that puts us in contact with subjects who lack partial phenomenal consciousness, such as blindsight subjects, we appear to have no contact with individuals that lack complete phenomenal character across all sensory modalities in a way where they could assert their lack. In order to assert the lack of phenomenal character, one would have to already possess the concept of phenomenal character in a way that applies to their own experience, which by definition zombies fail to have. Of course, one can counter this asymmetry argument by holding that through the practice of philosophy, one can learn to have better and better intuitions about (Z). On the social model I advocate, this would open up the question of whether or not the intuition is widely shared across a diverse set of individuals, such that one can take their, say, positive intuition about (Z) as being genuine. In addition, it appears that (Z) faces wide peer disagreement and is even unavailable to members of certain cultures and disciplines.

Finally, the learnability of the subject matter to which an intuition experience belongs is relevant to the attitude we should take toward the intuition having epistemic standing. In both cases, one could fail to have a genuine intuition, but in the case of (P), unlike (Z), it appears that the intuition is shared, and that we can appeal to methods outside of intuition in order to corroborate or help explain the correctness of the intuition, such as actually moving c to $L^{\star}$ at a distinct time, and showing that some things are contingent. In contrast to (C), nothing outside of intuition is available for corroborating (Z).

While I have not offered a complete account of intuition-based justification for beliefs about modality, I have sketched two main components of that account. The first is the acceptance of a disjunctive account of intuition experiences as being either genuine intuitions or intuition-like experiences that have gone wrong in some way. The second is to join the disjunctive account of intuition to a social dimension characterized by epistemic responsibility. For example, when a subject S has an experience, which can either be an intuition or merely intuition-like, that p is possible, does S seek to discover whether the intuition is shared? The conjunction of the two parts yields an alternative to Bealer's *modal reliabilism*, as well as a response to skepticism about intuition-based justification in the epistemology of modality.

References

Bealer, G. (2002). Modal Epistemology and the Rationalist Renaissance. In Gendler, T. & Hawthorne, J. (eds.), *Conceivability and Possibility*, Oxford: Oxford University Press, pp. 71–125.

Chalmers, D. (2002). Does Conceivability Entail Possibility? In Gendler, T. & Hawthorne, J. (eds.), *Conceivability and Possibility*, Oxford: Oxford University Press, pp. 145–200.

Chudnoff, E. (2013). *Intuition*. Oxford: Oxford University Press.

Chudnoff, E. (2018). Intuition in the Gettier Problem. In Hetherington, S. (ed.), *The Gettier Problem*, Cambridge: Cambridge University Press, pp. 177–198.

Fischer, B. (2016). A Theory-Based Epistemology of Modality. *Canadian Journal of Philosophy* 46.2: 228–247.

Hale, B. (2013). *Necessary Beings: An Essay on Ontology, Modality, and the Relations between Them*. Oxford: Oxford University Press.

Horvath, J. & Wiegmann, A. (2016) Intuitive Expertise and Intuitions about Knowledge. *Philosophical Studies* 173.10: 2701–2726.

van Inwagen, P. (1998). Modal Epistemology. *Philosophical Studies* 92.1: 67–84.

Kripke, S. (1971). Identity and Necessity. In M.K. Munitz (ed.), *Identity and Individuation*, New York: New York University Press, pp. 135–164.
McDowell, J. (2008). The Disjunctive Conception of Experience as Material for a Transcendental Argument. In Haddock, A. & Macpherson, F. (eds.), *Disjunctivism: Perception, Action, Knowledge*, Oxford: Oxford University Press, pp. 376–389.
McDowell, J. (2009). Criteria, Defeasibility, and Knowledge. In Byrne, A. & Logue, H. (eds.), *Disjunctivism: Contemporary Readings*, Cambridge, MA: MIT Press, pp. 75–91.
Nado, J. (2014a). Why Intuition? *Philosophy and Phenomenological Research* 86.1: 15–41.
Nado, J. (2014b). Philosophical Expertise. *Philosophy Compass* 9.9: 631–641.
Nado, J. (2015). Philosophical Expertise and Scientific Expertise. *Philosophical Psychology* 28.7: 1026–1044.
Roca-Royes, S. (2016). Similarity and Possibility: An Epistemology of *De Re* Modality for Concrete Entities. In Fischer, B. & Leon, F. (eds.), *Modal Epistemology after Rationalism*, Synthese Library, New York: Springer Publishing.
Strohminger, M. (2015). Perceptual Knowledge of Non-actual Possibilities. *Philosophical Perspectives* 29: 363–375.
Swain, S., Alexander, J., & Weinberg, J. (2008). The Instability of Philosophical Intuitions: Running Hot and Cold on Truetemp. *Philosophy and Phenomenological Research* 76.1: 138–155.
Williamson, T. (2007). *Philosophy of Philosophy*. Oxford: Blackwell Publishing.
Yablo, S. (1993). Is Conceivability a Guide to Possibility? *Philosophy and Phenomenological Research* 53: 1–42.

Further Reading

Chapters 1 and 4 of C. Peacocke, *Being Known* (Oxford: Oxford University Press, 1998), are where one finds a classic articulation of the integration challenge for the metaphysics and epistemology of modality as well as an account of *the implicit knowledge of principles of possibility* as a model for modal knowledge. T. Gendler and J. Hawthorne, *Conceivability and Possibility* (Oxford: Oxford University Press, 2002) is a classical source for issues in the epistemology of modality, especially concerning the work of David Chalmers and George Bealer. T. Williamson, *Philosophy of Philosophy* (Oxford: Blackwell Publishing, 2007), Chapter 5, is the most sustained articulation of the counterfactual account to the epistemology of modality. B. Hale, *Necessary Beings: An Essay on Ontology, Modality, and the Relations between Them* (Oxford: Oxford University Press, 2013), Chapter 11, is the most sustained account of the deduction model for modal knowledge.

PART 5

Modality and the metaphysics of science


Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Chapter 20

MODALITY AND SCIENTIFIC STRUCTURALISM

Steven French

Introduction

Prominent indicators of modality appear to be scattered across the face of science; think of laws, causality, probability and so on. Here I shall focus on the first and, furthermore, in the context of modern physics only; causality is notoriously problematic in this context, and probability requires a whole separate entry in itself. So, let us begin by considering the nature and role of laws. On the face of it, they seem to have an inherent modal aspect: take Boyle's Law, PV = RT, where P is the pressure of an ideal gas, V the volume, T the temperature and R a constant. Given V and T, say, P *has* to be such-and-such. Or consider Newton's Second Law, $F = ma$, where F is the force, m the mass and a the acceleration; given a certain force applied to an object of a certain mass, that object *must* accelerate at a certain rate. Both seem to exhibit a kind of necessity, but whether this counts as 'metaphysical' or 'nomic' (Kment 2017) is not something I'll go into here. Instead I shall focus on what for many is the more fundamental and more pressing question of whether these laws exhibit any kind of necessity at all.

Thus, Earman, for example, writes,

> The topic of laws of nature provides a kind of Rorschach test for philosophy. Some philosophers see in laws only Humean regularities; others see a kind of physical necessity; others see a necessity closer to logical necessity; others see expressions of causal powers; others see inference tickets; still others see relations between universals … and some see only a messy inkblot.
>
> (Earman 1993, p. 413)

Here, at least initially, I shall narrow my scope even further to the debate between those who reduce laws to Humean regularities and those who see in them a kind of physical or 'objective' modality; to put it another way, between those who take modality to be 'in the models' and those who take it to be 'in the world' (in some sense). I shall begin with the former view and outline some of the concerns it must face. Then I'll consider perhaps the most well-known contrasting view, namely dispositionalism, which situates modality in the objects that populate the world. Here I'll also run through certain critical issues before indicating how we might move beyond these deeply entrenched positions by adopting a form of structuralism. My overall thesis will be

that it is the latter that best accommodates certain fundamental features of modern science and which represents a new development, insofar as it takes modality to be 'in' the world but not situated in objects.

Modality and the mosaic

Different characterizations have been given of the Humean picture, but for our purposes the following will suffice: the underlying ontology consists of a 'mosaic' of properties (the 'perfectly natural' ones, to use Lewis's phrase) which are instantiated at space-time points and between which there are no modal relationships (these are the 'categorical' properties). This mosaic exhibits certain regularities, and according to the most widely accepted version of this account (sometimes referred to as the Mill-Ramsey-Lewis view), those that are represented by means of the 'best' theoretical system are deemed to be laws. There has been, of course, considerable discussion as to how this notion of the 'best' might be articulated, but typically it is taken to involve achieving a balance between simplicity and 'strength', or informativeness (for a useful discussion see Cohen and Callender 2009). There are thus two components to this account: a specific ontology of properties that are not modally informed, plus a theoretical framework in which modality features only insofar that we deem certain patterns of these properties to be laws. It is in this sense that 'modality is in the models'.

As compelling as this picture might seem, it faces certain difficulties (see also Berenstain and Ladyman 2012). First of all, concerns have arisen about the extent to which it meshes with the practice of physics. In particular, when it comes to the regularities, that practice makes a clear distinction between, on the one hand, the *dynamics*, according to which the state of the system considered will evolve; and the *initial conditions*, representing the state from which the system has evolved (see Hall 2015). Now, as far as accommodating the initial conditions is concerned, physicists want the framework to cover as wide a range of these as possible. However, when it comes to the dynamics, they want that aspect to be as specific as possible, for obvious reasons.

In other words, a given theoretical framework will be deemed to be 'better' to the extent that the representation of the initial conditions is as *uninformative* as possible, in the sense that it should admit as broad a range of initial conditions as nomologically permitted, whereas the representation of the dynamics should be as *informative* as possible, in the sense that it should admit as narrow a range of putative relevant laws as possible, and preferably only one (for the kind of phenomenon being considered). This combination then yields the maximum explanatory power, and if 'strength' is cashed out in these terms, this will result in the strongest theoretical framework.

The problem is, this distinction is entirely unmotivated by the underlying metaphysical picture of the mosaic of categorical properties (Hall 2015). To see this, consider how the Humean must set about identifying those regularities that are candidates for being regarded as laws: a claim about the world can be counted as law-like in the first place only if it can be regarded as a 'distinctively appropriate target for scientific inquiry', independently, of course, of its nomological status. The question then is, how, from all the various distributions of properties spread across the mosaic, do we pick out those that count as 'distinctively appropriate'? Lewis, famously, argued that we should pick out those that meet certain standards of simplicity and informativeness, but the latter is precisely *not* what we want when it comes to accommodating the initial conditions.

Thus, the Humean faces a dilemma: on the one hand, she can ensure that her account meshes with the practice of physics by incorporating the aforementioned distinction, but then the choice of the resulting standards (for judging what is a law, etc.) is entirely unmotivated by the metaphysics of the mosaic; or she can choose standards that are so motivated, but then her

account will be out of kilter with scientific practice. As Hall puts it, 'This choice between a guilty intellectual conscience and insane revisionism is not a happy one' (ibid., p. 51).

And it gets worse. Further reflection on the practices of modern physics reveals that not just laws and claims about possible initial conditions feature in physicists' theories, but certain symmetry principles as well. Indeed, their significance is now broadly recognised (see the collection of essays in Brading and Castellani 2003), and they add a further set of constraints that must be accommodated. Thus, consider a fundamental symmetry in quantum theory that states, crudely, that a permutation of particles of the same kind, such as two electrons, for example, does not result in a new, physically significant arrangement, unlike the case in classical mechanics. This Permutation Symmetry effectively breaks up the space of possible states of quantum systems into disconnected sectors, such that once a particle is in one of these sectors, the laws cannot lead to it escaping. Particles in different sectors behave statistically in very different ways; so, particles in the sector corresponding to Bose-Einstein statistics tend to aggregate together (this includes photons, and this behaviour helps account for lasers, for example), whereas those in that corresponding to Fermi-Dirac statistics tend not to (this includes electrons, and their aggregate behaviour is encapsulated in the Pauli Exclusion Principle in terms of which atomic structure and chemical bonding can be understood). Or consider the Special Theory of Relativity which incorporates the symmetries represented by the Poincaré group. This imposes a certain 'light-cone' structure on space-time such that the laws cannot take a system from within the light-cone, where, to put it crudely again, effects travel at speeds slower than or equal to that of light, to beyond the light-cone where interactions would take place faster than the speed of light.

How may the Humean accommodate such constraints? Let us first consider the mosaic. Here symmetries will be manifested in the form of meta-regularities; that is, regularities among the regularities in the properties. So, consider the example of Permutation Symmetry and quantum statistics again: we begin by noting certain regularities in the behaviour of particles with certain properties, such as charge and mass and spin. We then note that these regularities themselves exhibit certain regularities, corresponding to the distinction between bosons and fermions – and we realise that the mosaic is even more complex than we first imagined!

Now let us consider our theoretical framework. Which of these meta-regularities will be given the status of 'symmetry principles'? Presumably we would apply the same approach as with laws, granting such status to those regularities the incorporation of which leads to the 'best' system in terms of balancing simplicity and strength. Again, this adds a further layer of complexity; still, we can see how this would go (for one of the few attempts to accommodate symmetry along Humean lines, see Callum 2017). But of course, that further complexity renders the aforementioned dilemma even more acute. In addition to those restrictions on the set of initial conditions that would allow us, for example, to construct a mathematically acceptable dynamics to begin with, we have the further restrictions imposed on the latter by the relevant symmetries. Again, we face the issue of how these further restrictions might be motivated by the metaphysics of the mosaic with its modally inert, categorical properties.

Furthermore, a claim about the world counts as a 'meta-regularity', in this sense, in the first place, only if it too can be regarded as a 'distinctively appropriate target for scientific inquiry', independently of the meta-law-like status that is subsequently granted to it. But again, what criteria do we use for determining which of the various meta-regularities spread across the mosaic are deemed to be 'distinctively appropriate' in this sense? The Humean might, for example, revert to the kind of traditional criteria already mentioned, such as 'informativeness' or explanatory power. But the extent to which symmetry principles might be taken to possess such power is only just being explored, and certainly further work needs to be undertaken here (see French and Saatsi 2018).

Finally, there is the connection between these symmetries and the properties that make up the mosaic. The Humean picture is a 'bottom-up' one, metaphysically speaking. We begin with the mosaic of properties that are categorical in the sense of being modally 'inert' and by representing certain regularities and meta-regularities within our 'best' theoretical framework, we obtain the laws and symmetries that currently play such a significant role in modern physics. However, the physics itself seems to present an entirely different, 'top-down' view. Recall Permutation Symmetry, which divides up the state space of quantum systems into non-connected sectors, such that bosons occupy one such sector and fermions another. Broadly speaking, bosons are the 'force-carrying' particles – so photons 'carry' the electro-magnetic force – whereas fermions are the 'material' particles – electrons, for example, are fundamental in constituting the material objects all around us. Hence one of the most fundamental divisions in kind-properties effectively 'drops out' of this symmetry. And we can go further: it turns out that the Poincaré group can be used to further classify elementary particles and properties such as spin likewise 'drop out'. This picture suggests that the properties of the mosaic are 'grounded' in these symmetries, which sits in tension with the view that the latter are meta-regularities among the former.

Now of course, the Humean could insist that this top-down picture can be reconstrued: we begin with the bottom-up view, starting with the mosaic, the regularities, the meta-regularities, and we take those that feature in our 'best' theoretical framework to be the laws and symmetries. We then understand the relationship between symmetries and properties as presented in physics practice as a misconstrual of the actual metaphysical situation – although it captures the close metaphysical relationship, it gets the 'directionality' of the grounding wrong. Thus, the Humean picture would not be *inconsistent* with what physics appears to be telling us, but it would require some careful thinking through of what's going on when it comes to the relationship between the mosaic and the elements of the best system.

All of this is in addition to the well-known problems presented by quantum entanglement, whereby systems are modally inter-related in ways that are not straightforwardly accommodated within the Humean approach (Maudlin 2007). Perhaps a neo-Humean account can be constructed that accommodates this inter-connectedness between elements of the mosaic and still draws the line at incorporating the kind of necessity typically associated with laws (and symmetries), but this remains as work to be done.

Dealing with dispositionalism

What of the alternative view, that takes modality to be 'in' the world, in some sense? One way of understanding this is by situating the modality with the fundamental objects of physics and understanding their properties not as categorical or modally inert, but as dispositional and modally powerful. Here, an explicit appeal to the practice of physics is often made; thus, Molnar writes,

> Physics tells us what result is apt to be produced by the having of gravitational pull or of electromagnetic charge. It does not tell us anything else about these properties. In the Standard Model the fundamental physical magnitudes are represented as ones whose whole nature is exhausted by their dispositionality: that is, only their dispositionality enters into their definition. Properties of elementary particles are not given to us in experience: they have no accessible qualitative aspect or feature. There is no 'impression corresponding to the idea' here. What these properties are is exhausted by what they have a potential for doing, both when they are doing it and when they are not.
>
> (Molnar 1999, p. 13; see also Mumford 2011, p. 267)

However, problems also arise when we try to accommodate physics within the standard dispositional framework, as captured by the Stimulus and Manifestation Characterisation (S&M):

> for a property to be dispositional is for it to relate a stimulus (S) and a manifestation (M) such that, if an object instantiates the dispositional property, it would yield the manifestation in response to the stimulus.

The principal positive feature of this framework, at least in this context, is that it yields the laws of nature, and thus, the modal nature of the latter can be grounded in the dispositions or powers of the fundamental objects (see Bird 2007). In particular, if we accept that the identity of properties is entirely cashed out in dispositional terms, so that, in effect, what makes a property the property that it is are the powers that it has (this is the Dispositional Identity Thesis), then, adopting the aforementioned expression of S&M, laws are grounded in the relevant set of dispositions. And if we then consider a different possible world, populated by the same set of objects, then we will have the same properties, and hence by the Dispositional Identity Thesis, the same dispositions and the same laws, thereby accounting for the (physical) necessity of the latter (for a critical analysis, see Vetter 2009).

An unfortunate side effect of this grounding of the modality associated with laws in the powers of objects is that metaphysically speaking, laws become redundant (Mumford 2004). At best they are epiphenomenal with all the modal work performed by the properties, understood as bundles of dispositions. Here we have an interesting point of comparison with the Humean approach. Note also that this is another bottom-up picture, albeit with added modal 'punch'.

And likewise, the dispositionalist faces problems when it comes to accommodating symmetries (see Psillos 2006; Lange 2013). Consider the S&M characterisation described earlier: it does not seem to be the case that it is in response to a certain kind of stimulus that a quantum particle manifests fermionic or bosonic behaviour, or cannot accelerate past the speed of light, say. At the very least, it seems that the relevant symmetry principles cannot be straightforwardly related to dispositions seated in the objects in the way that laws can.

One option would be to bite the bullet and to reject these symmetry principles as nothing more than 'pseudo-laws' to be written out of our scientific world view (Bird 2007, p. 214). This flies in the face of the practice of modern physics (Livanios 2010). Thus, consider again the eliminativism about laws touched on earlier. Note that this does not deny the apparent role of laws in physics practice; rather, the claim is that because we can reduce their modal force to the action of dispositions via the S&M characterisation, we can drop them as fundamental elements of our world view. It is hard to see how we could do something similar with symmetries since it is hard to see how such a reduction could even begin.

Another option might be to expand our notion of disposition and metaphysically characterise the symmetries as dispositional properties of the world as a whole (see Bigelow et al. 1992; also Bird op. cit., p. 213; Chakravartty 2019). This might appeal to monists of a certain stripe and the local dispositions we are more familiar with, associated with properties, such as charge, etc., could be understood as derivative of these 'global' dispositions associated with the world-as-object. This kind of dispositional monism would certainly accommodate the 'drop-down' relationship between symmetries and properties but it does have a somewhat ad hoc flavour (Bird op. cit.; Livanios op. cit.). Furthermore, it remains unclear – given the Dispositional Identity Thesis – what the relevant property would be that would be decomposed into the appropriate dispositions. If it is 'being the world', then we would have one property associated with myriad different dispositions and we lose the specificity behind the Identity Thesis.

Alternatively, the dispositionalist could argue that the relevant properties of the world should be understood as the laws themselves. However, the Dispositional Identity Thesis can't be invoked here in the way that it can for properties, since the identity of scientific laws is not cashed out in terms of the associated symmetries and, furthermore, that fails to capture the constraining role of symmetries. Finally, if it is the world as a whole that is taken to possess, in whatever sense, the relevant dispositions, with the symmetries as manifestations, then, within the S&M characterisation, the dispositionalist again has to say something about what the associated stimuli could be.

Nevertheless this suggestion of re-situating the relevant modality in the laws, rather than objects, leads nicely on to an alternative way of conceiving of modality as 'in the world' that is immediately compatible with modern physics.

Structuralism: modality from the top down

We begin with the suggestion that we should take these laws and symmetry principles to be ontologically primitive. As Maudlin notes,

> The notions of law, possibility, counterfactual and explanation are deeply interconnected. The directions of the connections ought to be open to all hypotheses. If we take laws as primitive, relatively clean analyses can be given of the rest, analyses that fit intuitions, predict degrees and sorts of context dependence, and describe actual scientific practice, at least in some notable instances.
>
> (Maudlin 2007, p. 49)

So, for example, we can analyse a counterfactual claim as follows: we take a set of physical magnitudes that make the antecedent true, then consider the future values of those magnitudes generated by the relevant laws. If the consequent holds across all the states so generated, then the counterfactual is true, and if it holds in none, then the counterfactual is false (and if it holds in only some, then the counterfactual has an indeterminate truth-value).

However, taking the laws (and symmetries) as unanalysable primitives raises the question: 'what are laws such that we must be so mindful of them?' (Lange 2009, p. 199). The thought here seems to be that by virtue of refusing to offer any analysis of laws, this view cannot account for why we take them, and the form of necessity they are associated with, as significant. In particular, how do we explain why a given fact acquires some form of necessity by virtue of following from the relevant laws (ibid., p. 198)?

Obviously we cannot answer the first question by giving some sort of reductive account, as the dispositionalist does, or by pursuing an eliminativist line, as with Humeanism. To offer such analyses would conflict with this primitivist stance. However, we can situate it within a broader metaphysical perspective in such a way as to address the concern.

One such perspective is provided by structuralism which can be broadly characterised in terms of shifting our metaphysical attention from objects to structures, however conceived (see Ladyman 2016). This of course covers a wide variety of positions, some of which sit closer to the Humean end of the spectrum when it comes to modality (see Lyre 2010) and some of which are more closely aligned with dispositionalism (Chakravartty 2007). Here I shall focus on what has come to be known as 'ontic' structural realism that is motivated primarily by developments in 20th-century physics (Ladyman 1998). Within this view, laws and symmetries are conceived of as features of the structure of the world (French 2014; see also Cassirer (1936), who also gives primacy to laws and symmetries, albeit in a neo-Kantian context). Then the answer to

the question 'what are laws such that we must be so mindful of them?' is straightforward: they are part of the fundamental structure of the world.

Note that this directly accommodates those features of the practice of modern physics that Humeanism and dispositionalism struggle with, in particular the symmetry principles, such as Permutation Symmetry and Poincaré symmetry. In one sense, then, the structuralist has an immediate answer to the question, 'What is this structure that you speak of?' as she can simply point to the relevant laws and symmetries. Of course, insofar as these are expressed mathematically, this runs the risk of accusations of Platonism. The structuralist can respond in two ways. First, she can insist that what she is concerned with is *physical* structure, not mathematical, although delineating the two can be difficult. Appealing to causation in the quantum context is obviously problematic, and perhaps all that the structuralist can do is insist that physical structure is that which can be related to 'the phenomena', in the usual way (see French 2014, ch. 8). Secondly, she can add some metaphysical 'flesh' to the bare theoretical bones of the laws and symmetries as they are presented in the relevant theories, such as the Standard Model mentioned by Molnar.

So, consider again the way in which fundamental kinds, such as bosons and fermions or properties, such as mass and spin, 'drop out' of the symmetries – another feature that Humeanism and dispositionalism had difficulties accommodating. The structuralist can avail herself of various devices from the metaphysicians' toolbox to capture this sense of 'dropping out' (see French and McKenzie 2015). One such is the notion of 'determinable' (see Wilson 2017) – Permutation Symmetry is the determinable, of which the bosonic and fermionic representations are the determinates.

This then helps us answer the second question: why do certain facts acquire some form of necessity by virtue of following from the relevant laws? Consider the statement 'the pencil had to drop once I let it go' (cf. Maudlin 2007). What are the truth-conditions for such a claim? Well, the laws of classical mechanics, approximately, or General Relativity, more precisely, determine a class of models that can be regarded as 'possible worlds' providing those truth-conditions. The laws, of course, are trivially nomically necessary since they are true in every such possible world, by virtue of underpinning their construction. And the sense of 'had to' is explicated by noting that in every such model in which I let go of the pencil, it falls (ibid.). The claim is thus physically necessary, acquired by virtue of following from the relevant laws because these are features of the fundamental ontology of the world. Or consider the statement 'the electron had to occupy a different state (from another electron)', made, perhaps, in the context of the composition of atoms. This follows from the assignment of electrons to the Fermi-Dirac representation as determined by Permutation Symmetry. Again, the physical necessity of this statement, insofar as electrons cannot do otherwise, is inherited from that of the symmetry, as a feature of the world.

Now, as noted earlier, it has been suggested that such symmetries count as *determinables*, so in what sense are these features inherently *modal*? Consider Permutation Symmetry yet again: it turns out that Bose-Einstein statistics and Fermi-Dirac statistics aren't the only possible ones – there are an infinity of others, known as para-statistics (indeed, for a brief time during the later 1960s and early 1970s, it was thought that quarks were a kind of paraparticle). It is by virtue of this feature that we can understand how symmetry principles explain phenomena, within the framework of the so-called 'counterfactual approach' to explanation (French and Saatsi 2018). Again, one might worry that this sense of possibility reduces to mere mathematical possibility, insofar as such different statistics correspond to different representations of the Permutation Group in terms of which the symmetry principle can be represented. Indeed, one could go further and argue that insofar as this is all part of the mathematics of group theory, which can

be related to still further mathematical structures (see Bueno and French 2018), we cannot robustly delineate the kind of 'physical' modality we are concerned with from the kind of 'formal' possibilities offered by mathematics. But any feature of our theories that is expressed mathematically may face similar concerns. In the end, as was suggested earlier in this chapter, all we can do is note that this modality counts as physical insofar as it is embodied in a feature of our theories that is related to the physical world (however one wants to characterise that relationship).

Furthermore, just as we can add metaphysical flesh onto the bones of our formally expressed structure, so we can metaphysically 'beef up' this sense of modality by again utilising some of the tools in the metaphysicians' toolbox. We noted that one way in which the dispositionalist could accommodate laws and symmetries would be to take them as the 'seat' of her dispositions, except she could not then adhere to the S&M characterisation, given the lack of an appropriate stimulus. Responding to an entirely different set of concerns, Vetter has dropped the antecedent stimulus from this characterisation and developed an account in which certain possibilities are identified as 'potentialities' via their manifestation alone (Vetter 2015). Thus, a vase is 'fragile' in the sense that it has the potential to break easily, and this is expressed by '*x* can M', where M is the manifestation. The structuralist could adapt this device and take '*x*' here to be the relevant symmetry and as the manifestation, the property of spin, say, in the case of Poincaré symmetry or that of being a boson or fermion in the case of Permutation Symmetry. Within this framework, the symmetries can be understood as certain potentialities that are manifested in the actual world in certain ways. These particular manifestations in effect pick out the actual world from the set of possibilities and hence may be regarded as 'existential witnesses' that appropriately leaven the modality encoded in the symmetries (Wilson 2012). We can also go a little further and appropriate Vetter's distinction between 'mere' possibility and potentiality in terms of the latter being the 'localized counterpart of a non-localized modality' (Vetter 2014, p. 23) to help with the concern about distinguishing mathematical from physical possibilities. The full range of mathematical structures associated with the Permutation Group, say, can then be understood as the 'non-localized' possibilities and those that we regard as 'physical', yielding Bose-Einstein, Fermi-Dirac and para-statistics, can be understood as the localized counterparts.

There is still work to be done, of course. But if our metaphysics of modality is to be appropriately naturalised, it needs to set the laws and symmetry principles of modern physics at its heart. Failing to do so will leave it out of kilter with that practice, and although we can always try to set things straight with metaphysical adjustments, taking these laws and symmetries as primitive and modally informed offers the possibility of a well-balanced and nicely structured account.

References

Berenstain, N. and Ladyman, J. (2012) "Ontic Structural Realism and Modality", in E. Landry and D. Rickles (eds.), *Structural Realism*. Dordrecht: Springer, pp. 149–168.

Bigelow, J., Ellis, B., and Lierse, C. (1992) "The World as One of Kind: Natural Necessity and Laws of Nature", *The British Journal for the Philosophy of Science*, 43: 371–388.

Bird, A. (2007) *Nature's Metaphysics: Laws and Properties*. Oxford: Oxford University Press.

Brading, K. and Castellani, E. (eds.) (2003) *Symmetries in Physics: Philosophical Reflections*. Cambridge: Cambridge University Press.

Bueno, O. and French, S. (2018) *Applying Mathematics: Immersion, Inference and Interpretation*. Oxford: Oxford University Press.

Callum, D. (2017) "Humean Metaphysics and the Philosophy of Science", PhD Thesis, University of Leeds.

Cassirer, E. (1936) *Determinismus und Indeterminismus in der Modernen Physik*. Göteborg: Göteborgs Högskolas Årsskrift 42. Translated as *Determinism and Indeterminism in Modern Physics*. New Haven: Yale University Press, 1956.

Chakravartty, A. (2007) *A Metaphysics for Scientific Realism*. Cambridge: Cambridge University Press.

Chakravartty, A. (2019) "Physics, Metaphysics, Dispositions, and Symmetries – à la French", *Studies in History and Philosophy of Science* 74: 10–15.

Cohen, J. and Callender, C. (2009) "A Better Best System Account of Lawhood", *Philosophical Studies* 145: 1–34.

Earman, J. (1993) "In Defence of Laws: Reflections on Bas van Fraassen's Laws and Symmetry", *Philosophy and Phenomenological Research* 53: 413–419.

French, S. (2014) *The Structure of the World*. Oxford: Oxford University Press.

French, S. and McKenzie, K. (2015) "Rethinking Outside the Toolbox: Reflecting Again on the Relationship between Philosophy of Science and Metaphysics", in T. Bigaj and C. Wuthrich (eds.), *Metaphysics in Contemporary Physics*. Poznan Studies in the Philosophy of the Sciences and the Humanities. Amsterdam, the Netherlands: Rodopi, pp. 145–174.

French, S. and Saatsi, J. (2018) "Symmetries and Explanatory Dependencies", in J. Saatsi and A. Reutlinger (eds.), *Explanation beyond Causation*. Oxford: Oxford University Press, pp. 185–205.

Hall, N. (2015) "Humean Reductionism about Laws of Nature", in B. Loewer and J. Schaffer (eds.) *A Companion to David Lewis*. Oxford: Wiley.

Kment, B. (2017) "Varieties of Modality", in E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2017 Edition), URL: https://plato.stanford.edu/archives/spr2017/entries/modality-varieties/.

Ladyman, J. (1998) "What Is Structural Realism?", *Studies in History and Philosophy of Science* 29: 409–424.

Ladyman, J. (2016) "Structural Realism", in E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition), URL: https://plato.stanford.edu/archives/win2016/entries/structural-realism/.

Lange, M. (2009) "Review of *The Metaphysics within Physics*", *Mind* 118: 197–200.

Lange, M. (2013) "How to Explain the Lorentz Transformations", in M. Tugby and S. Mumford (eds.), *Metaphysics and Science*. Oxford: Oxford University Press, pp. 73–100.

Livanios, V. (2010) "Symmetries, Dispositions and Essences", *Philosophical Studies* 148: 295–305.

Lyre, H. (2010) "Humean Perspectives on Structural Realism", in F. Stadler (ed.), *The Present Situation in the Philosophy of Science*. Dordrecht: Springer, pp. 381–397.

Maudlin, T. (2007) *The Metaphysics within Physics*. Oxford: Oxford University Press.

Molnar, G. (1999) "Are Dispositions Reducible?", *The Philosophical Quarterly* 49: 1–17.

Mumford, S. (2004) *Laws in Nature*. Oxford: Oxford University Press.

Mumford, S. (2011) "Causal Powers and Capacities", in H. Beebee, P. Menzies and C. Hitchcock (eds.), *The Oxford Handbook of Causation*. Oxford: Oxford University Press, pp. 265–278.

Psillos, S. (2006) "What Do Powers Do When They Are Not Manifested?", *Philosophy and Phenomenological Research* 72: 135–156.

Vetter, B. (2009) "Review of *Nature's Metaphysics: Laws and Properties," Logical Analysis and History of Philosophy* 8: 320–328.

Vetter, B. (2014) "Dispositions without Conditionals", *Mind* 123: 129–156.

Vetter, B. (2015) *Potentiality: From Dispositions to Modality*. Oxford: Oxford University Press.

Wilson, J. (2012) "Fundamental Determinables", *Philosophers Imprint* 12, URL: http://www.philosopher-simprint.org/012004/.

Wilson, J. (2017) "Determinables and Determinates", in E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2017 Edition), URL: https://plato.stanford.edu/archives/spr2017/entries/determinate-determinables/.

Chapter 21

LAWS OF NATURE, NATURAL NECESSITY, AND COUNTERFACTUAL CONDITIONALS

Marc Lange

21.1 The topic

In 1772, a distinguished committee of the Royal Society concluded that pointed lightning rods work better than blunt ones. However, in 1777, King George III decided to reject the committee's recommendation; he had sharp conductors replaced by blunt ones on his palace, and he lobbied the Royal Society to rescind its resolution – at least partly because the committee had been chaired by Benjamin Franklin, and after the Battle of Bunker Hill, the King was reluctant to demonstrate respect for any of Franklin's views, even about lightning rods. But when the King appealed to the President of the Royal Society, Sir John Pringle, to retract the committee's decision, Pringle is said to have replied, "Sire, I cannot reverse the laws of nature" (Weld 1848: 101).

Pringle's remark gestures towards not just the objectivity but also the necessity of the natural laws. If the laws of nature prohibit some kind of event (such as increasing the total quantity of energy in the universe, or accelerating a body from rest to beyond the speed of light – or improving a lightning rod by making it blunter), then it is not merely the case that such an event *does* not happen. Such an event *cannot* happen; it is impossible. The laws are not merely true; they could not have been false. Their necessity is often called "natural," "physical," "nomological," or "nomic" necessity in order to distinguish it from logical, conceptual, metaphysical, mathematical, and other species of necessity. Natural necessity is not possessed by contingent facts that are not laws (that is, by "accidents"), such as the fact (presuming it to be a fact) that all gold cubes in the entire history of the universe are smaller than one cubic meter. There could have been a gold cube exceeding two cubic meters, but in fact, none ever exists.

That the laws are (naturally) necessary, whereas the accidents are not, is associated with another difference between laws and accidents – namely, in their relation to the facts expressed by "counterfactual conditionals": if-then statements that concern what would have happened under circumstances that never actually come to pass. (For example, here is a counterfactual conditional that is true: "If I had gone to the market today, then I would have purchased a quart of milk." That I went to the market today – a falsehood – is the "counterfactual antecedent.") The laws, being necessary, would still have been true even if other things had been different, whereas an accident has less perseverance under counterfactual antecedents. For instance, since

it is a law that no body is accelerated from rest to beyond the speed of light, this cosmic speed limit would not have been broken even if the Stanford Linear Accelerator had now been cranked up to full power. (That was a counterfactual conditional.) On the other hand, since it is an accident that every gold cube is smaller than a cubic meter, this pattern would (presumably) have been broken if Bill Gates had wanted to have a gold cube constructed exceeding a cubic meter. (That was another counterfactual conditional.)

Thus, natural lawhood, natural necessity, and counterfactual conditionals are closely interrelated. This essay aims to sketch briefly an account of their relations.

21.2 Sub-nomic stability

Of course, laws would not still have remained true under counterfactual suppositions with which they are logically inconsistent (such as "Had a body been accelerated from rest to beyond the speed of light"). But presumably, the laws would still have held under any counterfactual antecedent that is logically consistent with all of the laws. On the other hand, trivially no fact that is an accident is preserved under all of those antecedents (since one such antecedent posits the accident's negation, and the accident is obviously not preserved under that antecedent). Making a few details more explicit, we arrive at the following proposal:

> It is a law that *m* if and only if in any conversational context, for any circumstance *p* that is logically consistent with all of the facts *n* (taken together) where it is a law that *n*, it is true that if *p* had been the case, then *m* would still have been the case (that is, *p* $\Box\!\!\rightarrow$ *m*).

In this proposal (and until further notice), I reserve lowercase letters (such as "*m,*" "*p,*" and "*n*") for "sub-nomic" claims – that is, for claims such as "The emerald at spatiotemporal location … is 5 grams" and "All emeralds are green," as contrasted with "nomic" claims such as "It is a law that all emeralds are green" and "It is an accident that the emerald at spatiotemporal location … is 5 grams." (On my view, a claim is "sub-nomic" exactly when in any possible world, what makes the claim hold, or fail to hold, is not – even partly – that a given fact in that world is a law or that a given fact in that world is an accident.) Let me also note that the account I am sketching presupposes that every logical consequence of laws qualifies as a law and that every broadly logical truth (that is, every truth holding with some species of necessity stronger than natural necessity, such as logical, metaphysical, mathematical, and conceptual necessity) is by courtesy a natural law since it has all of the necessity of a law and then some. The proposal had to refer to the conditional's being true *in all conversational contexts* because the truth-values of counterfactual conditionals are notoriously context-sensitive.[1]

This proposal captures an important difference between laws and accidents in their behavior toward counterfactuals. However, even if this proposal is correct, there is an obvious limitation on how enlightening it can be. That is because the laws appear in this proposal on *both* sides of the "if and only if." That is, the proposal picks out the laws by their invariance under a certain range of counterfactual antecedents *p* (namely, under any *p* that is logically consistent with all of the facts *n* where it is a law that *n*) – but this range of antecedents, in turn, is picked out by the laws. Therefore, although this proposal's truth is not thereby made trivial, this proposal fails to specify what it is in virtue of which *m* is a law. It also fails to account for what makes the laws *important*. The laws' invariance over the particular range of counterfactual antecedents that the proposal mentions makes the laws special only if there is already something special about having this particular range of invariance. But the laws are precisely what pick out this range. So there

is something special about being invariant over this range only if there is already something special about the laws. If there is no prior, independent reason why this particular range of counterfactual antecedents is special, then the laws' invariance under these antecedents fails to make the laws special. They merely have a certain range of invariance (just as a given accident has some range of invariance).

However, these deficiencies can be overcome by tweaking the proposal. It says roughly that the laws form a set of truths that would still have held under every antecedent with which the set is logically consistent. In contrast, take the set containing exactly the logical consequences of the accident that all gold cubes are smaller than a cubic meter. This set's members are *not* all preserved under every antecedent that is logically consistent with this set's members. For instance, had Bill Gates wanted to have a gold cube constructed exceeding a cubic meter, then such a cube might well have existed, and so it might not have been the case that all gold cubes are smaller than a cubic meter. Yet the antecedent p that Bill Gates wants such a cube constructed is logically consistent with all gold cubes being smaller than a cubic meter.

The idea we have just seen motivates the definition of "sub-nomic stability":

> Consider a non-empty set Γ of sub-nomic truths containing every sub-nomic logical consequence of its members. Γ possesses *sub-nomic stability* if and only if for each member m of Γ and for any p where $\Gamma \cup \{p\}$ is logically consistent (and in every conversational context), it is not the case that if p had held, then m might not have held (i.e., then m's negation might have held) – that is, $\sim (p \Diamond\!\!\rightarrow \sim m)$.[2]

Notice that $\sim (p \Diamond\!\!\rightarrow \sim m)$ logically entails $p \Box\!\!\rightarrow m$. (In other words, that it is not the case that $\sim m$ might have held, if p had held, logically entails that m would have held, if p had held.) Therefore, a set of truths is sub-nomically stable exactly when its members would all still have held – indeed, not one of their negations even might have held – under any counterfactual antecedent with which they are all logically consistent.

In contrast to our initial proposal, stability does not use the laws to pick out the relevant range of counterfactual suppositions; stability avoids privileging the range of counterfactual antecedents that is logically consistent with the laws. Rather, each set picks out for itself the range under which it must be invariant in order for it to qualify as stable.

21.3 Stability linked to lawhood

According to our initial proposal, which was correct but relatively unilluminating, the set Λ of all sub-nomic truths m where it is a law that m is sub-nomically stable. In contrast, the set spanned by the fact that all gold cubes are smaller than a cubic meter is unstable because this set's members are all logically consistent with Bill Gates wanting a gold cube larger than a cubic meter, yet the set's members are not all invariant under this counterfactual supposition.

Let us look at another example. Take the accident g that whenever a certain car is on a dry, flat road, its acceleration is given by a certain function of how far its gas pedal is being depressed. Had the gas pedal on a certain occasion been depressed a bit farther, then g would still have held. Can a stable set include g? Such a set must also include the fact that the car has a four-cylinder engine, since had the engine used six cylinders, g might not still have held. (Once the set includes the fact that the car has a four-cylinder engine, the antecedent that the engine has six cylinders is logically *inconsistent* with the set, so to be stable, the set does not have to be preserved under that antecedent.) But since the set includes a description of the car's engine, its stability also requires that it include a description of the engine factory, since had that factory been

different, the engine might have been different. Had the price of steel been different, the engine might have been different. And so on. Let's see that this ripple effect propagates endlessly. Take the following antecedent: had either *g* been false or there been a gold cube larger than a cubic meter. Under this antecedent, is *g* preserved? In every context? Certainly not. This counterfactual antecedent pits *g*'s invariance against the invariance of the gold-cube generalization. It is not the case that in every context, *g* proves more resilient. Therefore, to possess sub-nomic stability, a set that includes *g* must also include the fact that all gold cubes are smaller than a cubic meter (making the set logically inconsistent with the antecedent, and so to be stable, the set does not have to be preserved under that antecedent). Since a stable set that includes *g* must include even the fact about gold cubes, I conclude that the only set containing *g* that might be stable is the set of *all* sub-nomic truths.

I conclude that *no* nonmaximal set of sub-nomic truths that contains an accident possesses sub-nomic stability. This suggests my proposal for the laws' special relation to counterfactuals: the set Λ of all sub-nomic truths *m* where it is a law that *m* is sub-nomically stable, whereas no set containing an accident is sub-nomically stable (except perhaps for the set of all sub-nomic truths, considering that the range of antecedents under which this "maximal" set must be preserved in order to be stable does not include any false antecedents since no falsehood is logically consistent with all of this set's members).

It is a law that *m*, then, exactly when *m* belongs to a nonmaximal sub-nomically stable set.

21.4 A hierarchy of sub-nomically stable sets

Are there any other nonmaximal sub-nomically stable sets besides Λ? The sub-nomic broadly logical truths (such as mathematical truths) form a stable set since they would still have held under any broadly logical possibility. I will now show that for any two sub-nomically stable sets, one must be a proper subset of the other. The strategy is to consider an antecedent of the sort that we just saw in connection with the example involving *g* and the fact about gold cubes – namely, an antecedent pitting the invariance of the two sets against each other:

1. Suppose (for *reductio*) that Γ and Σ are sub-nomically stable, *t* is a member of Γ but not of Σ, and *s* is a member of Σ but not of Γ.
2. Then (~*s* or ~*t*) is logically consistent with Γ.
3. Since Γ is sub-nomically stable, every member of Γ would still have been true, had (~*s* or ~*t*) been the case.
4. In particular, *t* would still have been true, had (~*s* or ~*t*) been the case. That is, (~*s* or ~*t*) $\boxrightarrow$ *t*.
5. So *t* & (~*s* or ~*t*) would have held, had (~*s* or ~*t*). Hence, (~*s* or ~*t*) $\boxrightarrow$ ~*s*.
6. Since (~*s* or ~*t*) is logically consistent with Σ, and Σ is sub-nomically stable, no member of Σ would have been false had (~*s* or ~*t*) been the case.
7. In particular, *s* would not have been false, had (~*s* or ~*t*) been the case. That is, ~((~*s* or ~*t*) $\boxrightarrow$ ~*s*).
8. Contradiction from 5 and 7.

Thus, the sub-nomically stable sets must form a nested hierarchy.

We were asking about nonmaximal sub-nomically stable sets besides Λ. Since no nonmaximal *superset* of Λ is sub-nomically stable (since it would have to contain accidents), we must look for sub-nomically stable sets among Λ's proper subsets. Many of them are clearly unstable. For instance, what would have happened had Coulomb's law (the law governing the strength and direction of electrostatic forces) been violated sometime in the past? In that case, it would not

have been in a law, and so it might also have been violated sometime in the future. Accordingly, consider a restriction of Coulomb's law to times *after* today, and take the set containing exactly this restricted generalization and its sub-nomic, broadly logical consequences. Our counterfactual antecedent (positing that Coulomb's law was violated sometime in the past) is logically consistent with this set, but as we just saw, the set is not invariant under that antecedent. Therefore, this proper subset of Λ is unstable.

However, some of Λ's proper subsets may be stable. For instance, consider the great conservation laws (of energy, momentum, etc.), together with the law by which forces combine (the "parallelogram of forces") and the fundamental dynamical law relating force to motion (which, in classical physics, may be Newton's second law together with its analog relating applied torque to angular acceleration, or Hamilton's principle, or the Euler-Lagrange equations, etc.). Plausibly, these laws would all still have held, even if the force laws had been different. For instance, had there been additional kinds of force in the world, then whatever they were, they would have had to obey the same conservation laws as the actual kinds of forces do. Intuitively, the conservation laws and their colleagues not only describe the kinds of forces there are, but also constrain the kinds of forces there could have been. In that case, the set containing exactly these constraints – despite omitting many laws of nature, such as the force laws – would be a sub-nomically stable proper subset of Λ.

As constraints on the force laws, the conservation laws would explain various features of the force laws. For instance, the reason why all of the various kinds of forces are alike in conserving energy, despite their diversity in other respects, would not be that the electromagnetic force conserves energy, the nuclear force conserves energy, and so on for each of the various actual species of forces – as if it were a coincidence that all of the various kinds of force are alike in conserving energy. Rather, there would be a common reason why they all conserve energy – namely, because the law of energy conservation limits the kinds of forces there could have been.

21.5 Stability and natural necessity

By the definition of sub-nomic stability, the members of a sub-nomically stable set would all still have held under any sub-nomic counterfactual antecedent with which they are all logically consistent. That is, a sub-nomically stable set's members *would* all still have held under any sub-nomic counterfactual antecedent under which they *could* (i.e., without contradiction) all still have held. In other words, a stable set's members are collectively as resilient under sub-nomic counterfactual antecedents as they could collectively be. They are *maximally* resilient. That is, I suggest, they are *necessary*. In other words, I propose that a sub-nomic truth has a species of necessity exactly when it belongs to a (nonmaximal) sub-nomically stable set, and that for each of these sets, there is a distinct species of necessity that is possessed by exactly its members.[3]

There could be many species of natural necessity – many strata of natural laws (see the right-hand side of Figure 21.1). A stable proper subset of Λ is associated with a stronger variety of necessity than Λ. That is, the range of antecedents under which the proper subset's members are all preserved, in connection with its stability, is wider than the range of antecedents under which Λ's members are all preserved, in connection with Λ's stability.[4] Stability associated with greater invariance corresponds to a stronger variety of necessity.

The actual inventory of forces and the individual force laws are matters of natural necessity, but they lack the stronger necessity possessed by a constraint. Therefore, that inventory and those force laws are modally too weak to be responsible for the modal strength of the law of energy conservation, if it is a constraint. They cannot help to explain why it is a law that energy is conserved.

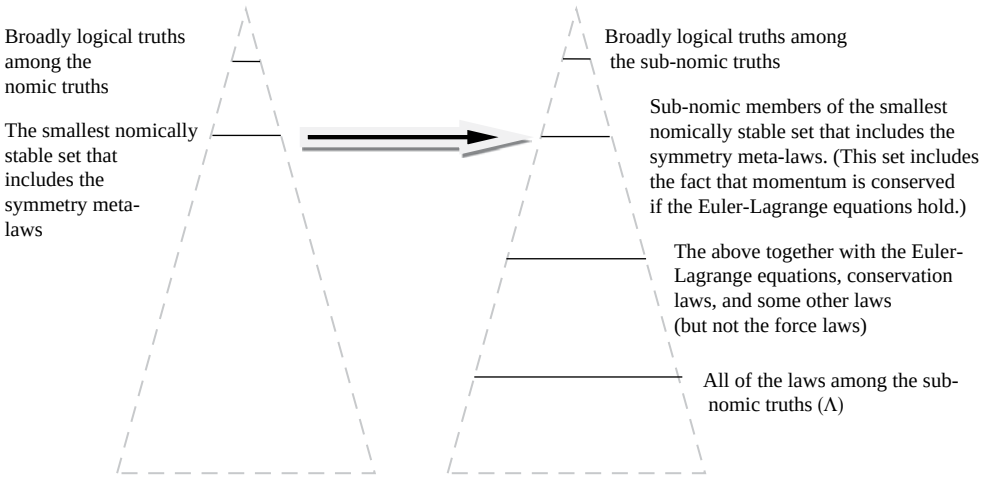


Figure 21.1 At right is the hierarchy of sub-nomically stable sets, with four of its members. (To the right of each line representing a rung on the hierarchy is specified the members of the stable set lying on that rung of the hierarchy.) Smaller sets, having stronger grades of necessity and larger ranges of invariance associated with their stability, appear higher on the hierarchy. At left is the hierarchy of nomically stable sets, with two of its members. (The members of a nomically stable set are specified to the left of the set's rung.) The rightward arrow depicts the way in which a nomically stable set's sub-nomic members are "projected" onto the hierarchy of sub-nomically stable sets.

Mathematical truths possess an even stronger variety of necessity than a conservation law that is a constraint. Accordingly, mathematical truths would still have held even if the conservation laws and force laws had been violated. This picture of necessity as associated with stability identifies what is common to broadly logical necessity and to the various grades of natural necessity in virtue of which they are all species of the same genus. This account also explains why there is a natural ordering among these species: because for any two sub-nomically stable sets, one must be a proper subset of the other.

21.6 Meta-laws, necessity, and nomic stability

Symmetry principles are closely associated with laws of nature. For instance, it is a symmetry principle that the laws privilege no particular moment in the universe's history – i.e., that they are invariant under arbitrary temporal displacement. Since a symmetry principle is made true by which facts are laws, it is not stated by a "sub-nomic" claim. Rather, it is stated by a "nomic" claim – that is, a claim that purports to describe which truths expressed by sub-nomic claims are (or are not) matters of law. (For instance, that all emeralds are green is a sub-nomic fact, whereas that it is a law that all emeralds are green is a nomic fact.)

Therefore, a symmetry principle is ineligible to belong to a sub-nomically stable set. Rather, a symmetry principle is a "meta-law": a law that governs the laws that are expressed by sub-nomic claims (the "first-order" laws). For instance, the symmetry principle I just mentioned explains why the first-order laws are invariant under arbitrary temporal displacement. To characterize the invariance under counterfactual antecedents that sets meta-laws apart from nomic and sub-nomic truths lacking the meta-laws' necessity, we need an analog of sub-nomic stability that applies to sets of claims that may contain both sub-nomic and nomic truths. Here is that analogue (now allowing lower-case letters such as "*p*" to stand for claims that are either sub-nomic or nomic):

> Consider a non-empty set Γ of truths that are nomic or sub-nomic containing every nomic or sub-nomic logical consequence of its members. Γ possesses *nomic stability* if and only if for each member m of Γ (and in every conversational context), $\sim (p \diamond\!\!\rightarrow \sim m)$ for any p where $\Gamma \cup \{p\}$ is logically consistent.

For instance, Λ lacks nomic stability because energy might not have been conserved if there had been no *law* requiring its conservation. (The antecedent of that counterfactual is a nomic claim that is logically consistent with Λ.) Perhaps, though, the set spanned by all of the truths about which sub-nomic claims are laws and which are not (such as the fact that it is a law that all emeralds are green but not a law that all gold cubes are smaller than a cubic meter – a fact not included in Λ) possesses nomic stability, just as Λ possesses sub-nomic stability. There may, in addition, be more exclusive sets that possess nomic stability. (By the same kind of argument I gave regarding the sub-nomically stable sets, the nomically stable sets form a nested hierarchy.) I suggest that the meta-laws are exactly the members of any nomically stable set more exclusive than the set spanned by all of the truths about which sub-nomic claims are laws and which are not.

The members of a nomically stable set *would* all still have held under any nomic or sub-nomic counterfactual antecedent with which they are all logically consistent – that is, under which they *could* (i.e., without contradiction) all still have held. Therefore, the members of a nomically stable set possess a kind of *maximal* resilience in that they are collectively as resilient under nomic or sub-nomic counterfactual suppositions as they could collectively be. Accordingly, they possess a species of necessity.

Famously, various conservation laws are logically entailed by various symmetry meta-laws within a Hamiltonian dynamical framework. It is widely believed that time-displacement invariance helps to explain why energy is conserved, space-displacement invariance helps to explain why linear momentum is conserved, and so forth. These explanations depend upon the connection between nomically stable sets and sub-nomically stable sets: for any nomically stable set Γ, its sub-nomic members must form a sub-nomically stable set Σ. Here is the proof:

1. If p (a sub-nomic claim) is logically inconsistent with a nomically stable set Γ, then Γ entails $\sim p$ (also sub-nomic), and so p is also logically inconsistent with the set Σ containing exactly Γ's sub-nomic logical consequences.
2. Conversely, if p is logically inconsistent with Σ, then obviously p is logically inconsistent with Γ.
3. Therefore (from 1 and 2), the sub-nomic claims that are logically consistent with Σ are exactly the sub-nomic claims that are logically consistent with Γ.
4. By Γ's nomic stability, Γ and hence its subset Σ are preserved under every sub-nomic antecedent p that is logically consistent with Γ – which (by 3) are exactly the sub-nomic antecedents that are logically consistent with Σ.
5. Hence, Σ is sub-nomically stable.

Thus, each nomically stable set Γ is associated with a sub-nomically stable "projection" Σ.

Suppose, then, that the symmetry meta-laws (forming a nomically stable set) entail various sub-nomic facts, including that a given conservation law holds under the Hamiltonian dynamical framework. The fact that the conservation law is true if the Hamiltonian dynamical law is true then belongs to the nomically stable set's sub-nomically stable "projection" (see Figure 21.1). Presumably, that projection is more exclusive than Λ since not all of Λ's members belong to the symmetry meta-laws' nomically stable set. Hence, there is a sub-nomically stable set possessing a strong variety of natural necessity that contains the fact that various conservation laws

hold if the Hamiltonian dynamical law holds, but does not contain the force laws or the Hamiltonian dynamical law (the Euler-Lagrange equations).

I suggested earlier that the fundamental dynamical law is also a constraint. Suppose that it does not belong to the sub-nomically stable projection of the symmetry meta-laws' nomically stable set, but rather to some sub-nomically stable set that is lower on the hierarchy (though above Λ), as depicted in the figure. Then, since the sub-nomically stable sets form a nested hierarchy, the fact that the conservation law holds under the Hamiltonian framework must also belong to that set, since it belongs to a set that is higher in the hierarchy (namely, the sub-nomic projection of the symmetry meta-laws). Therefore, since a sub-nomically stable set contains every sub-nomic logical consequence of its members, various conservation laws must also be members of any sub-nomically stable set that includes the Hamiltonian dynamical law along with the fact that those conservation laws hold if the Hamiltonian dynamical law holds. Thus, those conservation laws are constraints on the force laws; they have greater necessity than the force laws possess. In this way, the fact that a given conservation law is a constraint is explained by the fact that the fundamental dynamical law is a constraint and that the corresponding symmetry principle is a meta-law.

21.7 Conclusion

I have not argued that various facts are indeed laws of nature or that various laws are in fact constraints or meta-laws. These are matters for science to confirm empirically. My aim has instead been to make some progress toward understanding what difference it would make (in terms of counterfactuals, necessity, and scientific explanation) that various facts are laws and that various laws are constraints or meta-laws.

Facts about necessity, natural law, explanation, and counterfactuals are closely interrelated. Philosophy aims not only to identify these interrelations, but also to account for them. For instance, philosophy aims not only to identify the precise relation to counterfactuals that sets laws apart from accidents, but also to explain why laws (but not accidents) stand in that relation. Perhaps (as has traditionally been maintained) facts about what is a law and what is accidental are responsible (at least partly) for the facts expressed by various counterfactual conditionals. Or perhaps (as I proposed in Lange 2009) the facts expressed by various counterfactual conditionals are responsible for the facts about what is a law and what is accidental. Or perhaps facts about necessity or about explanation (see Kment 2014) are responsible (at least partly) not only for the facts expressed by various counterfactual conditionals, but also for facts about what is a law and what is accidental. Any account of these relations should do justice to the details of scientific practice.

Notes

1 A more sophisticated account would replace this appeal to "all conversational contexts" with a narrower range of contexts, thereby giving an account of the laws of some particular scientific field, such as aerodynamics, or island biogeography, or traffic science, or human medicine.

2 For the sake of simplicity, this definition of "sub-nomic stability" omits some details from Lange (2009). Also found in Lange (2009) is discussion of some alternatives and objections to this proposal.

3 The laws' natural necessity is therefore not a "cheap and trivial" matter of self-entailment (Fine 2005: 246). The laws earn their status as necessary by virtue of forming a sub-nomically stable set. Natural necessity is not on a par with "feline necessity" (where something is "felinely necessary" if and only if it is true in all possible worlds containing cats). As Bird (2007: 48) remarks, feline necessity does not name a genuine variety of necessity. Natural necessity does.

4 This does not mean that if some conservation law constrains the force laws, there is no antecedent (and context) under which a force law is preserved but the conservation law is not. Obviously, "had the force law held but the conservation law not held" is such an antecedent. Nevertheless, the range of antecedents under which the constraints are all preserved *in connection with their stability* (that is, the range logically consistent with the constraints) is wider than the range under which Λ's members are all preserved *in connection with Λ's stability* (that is, the range logically consistent with Λ). The aforementioned antecedent falls outside of both of these ranges since it is logically inconsistent with Λ.

References

Bird, A. (2007) *Nature's Metaphysics: Laws and Properties*, New York: Oxford University Press.

Fine, K. (2005) "The Varieties of Necessity," in K. Fine (ed.), *Modality and Tense*, Oxford: Oxford University Press.

Kment, B. (2014) *Modality and Explanatory Reasoning*, Oxford: Oxford University Press.

Lange, M. (2009) *Laws and Lawmakers*, New York: Oxford University Press.

Weld, C. (1848) *A History of the Royal Society, with Memoirs of its Presidents*, vol. 2, London: John W. Parker.

Chapter 22

NATURAL KINDS AND MODALITY

Alexander Bird

22.1 Introduction

Natural kinds are a particular locus of debate concerning modality. Modern essentialists claim that natural kinds are the subjects of necessary but aposteriori truths and have real essences. Here I look at these claims and the founding arguments for them, as presented in the work of Kripke and Putnam. From where do these arguments ultimately derive their support? From intuition? Or do they have an independent foundation in semantic theory? And if intuition is the foundation, is this intuition reliable?

22.2 Kripke on natural kinds, necessity, and essence

In *Naming and Necessity* and elsewhere, Saul Kripke makes a number of claims:

(I) For some natural kinds, A and B, it is a necessary aposteriori truth that A is B.
(II) By investigating the basic explanatory properties of a kind, science attempts to find its nature and so essence.
(III) Natural kind terms are rigid designators, designating natural kinds.
(IV) Where 'A' and 'B' are natural kind terms, 'A is B' expresses an identity.

Note that claims (I) and (II) and are in the material mode—they concern truths about kinds. Whereas claims (III) and (IV) are in the formal mode—they are about the properties of linguistic items such as names.

Let 'A' and 'B' be natural kind terms. Then according to (III) they rigidly designate natural kinds. So 'A is B' must express an identity between natural kinds, the proposition that natural kind A is natural kind B (which Kripke calls a 'theoretical identification'). This is not trivial, for if 'A' and 'B' were non-rigid definite descriptions, to be analysed as Russell proposes, then 'A is B' would not express an identity. So (IV) is a consequence of (III). Now consider the general claim that identities are necessarily true and are sometimes known aposteriori. This supports (I), if we think that what goes for identities in general goes for identities concerning natural kinds in particular. That is, we should accept (I) if we think that identities among

natural kinds are not special—they are not a subset of identities that are always knowable apriori (if true).

Examples of (I) are as follows (page references in brackets are to Kripke 1980):

> Gold is the element with the atomic number 79 [116]
> Water is H_2O [116, 128]
> Heat is the motion of molecules [99, 129]
> Temperature is mean molecular kinetic energy [129]
> Light is a stream of photons [116, 129]
> Flashes of lightning are flashes of electricity [132].

Kripke also includes in his discussion some propositions that do not strictly conform to (I), such as:

> Cats are animals [123, 125]
> Whales are mammals [138]
> Lightning is an electrical discharge [116]
> Light is a form of electromagnetic radiation [129].

These look to be examples of a claim weaker than (I), viz.:

V. For some natural kinds, A and B, it is a necessary aposteriori truth that all members of A are members of B.

We ought also to take note of other claims that Kripke makes that are weaker than (I) and (II):

VI. For some natural kinds, e.g. A, and properties, e.g. P, it is a necessary aposteriori truth that all members of the kind A have property P.
VII. By investigating the basic explanatory properties of a kind, science attempts to find its essential properties.

Claim (VII) is weaker than (II) because (II) claims that we can find something that is *the* essence of a natural kind, implying that possession of the essence is sufficient as well as necessary for kind membership—a full or complete essence. Whereas (VII) identifies only essential properties that are necessary but which may not be sufficient for kind membership—a partial essence. So, for example, the essence of gold might be the property of being comprised solely of atoms with 79 protons in the nucleus—which is both necessary and sufficient for something being gold. Whereas there might not be a property that is *the* essence of being a cat; even so, it is essential to being a cat that an entity is an animal.

22.3 Kripke's argument

Although Kripke starts his discussion of natural kinds by mentioning the theoretical identifications we find in (I), his arguments do not focus on those identifications directly. Rather he starts by arguing that:

> (i.a) Certain properties we might use to characterise a natural kind and which might be thought to be components of the definition of the kind term, are not necessary properties of the kind.

These properties are ones that in the subsequent literature have often been called 'superficial' properties—these being properties that are readily observable and which are apparent to those with little or no scientific training. He also argues that:

> (i.b) The superficial properties in the alleged definitions of the kind terms are not jointly sufficient for kind membership.

Kripke then argues that:

> (ii) Some other (non-superficial) properties that we find aposteriori to characterise the kind are necessary to it.

These properties are typically ones discovered by sophisticated scientific means. They have often been called 'structural' properties, that is inner, structural properties of member of the kind. While Kripke does often talk in terms of structure, it is not clear that Kripke thinks that all such properties are structural (subsequent authors—see McGinn 1976—have extended Kripke's essentialism about origin to biological kinds).

Regarding (i.a), Kripke considers a supposed definition of 'gold' as 'yellow metal' (which he finds in Kant). If correct, that would make being yellow a necessary condition on a piece of matter being gold. He then asks how we would respond if it were discovered that the yellow appearance of the stuff being dug out of gold mines is an illusion, that the stuff is in fact blue. If the definition of gold as yellow metal is correct, then it would have to be said that the stuff being dug in the mines or of which many rings are made and so on is not gold—that what we had been calling 'gold' is not gold after all. Kripke denies that this would be our response. Instead, we would say that we have discovered that gold is not yellow after all, but is blue.

Kripke presents an analogous argument regarding the proposal that tigers are, by definition, large carnivorous quadrupedal felines, tawny yellow in colour with blackish transverse stripes and white belly. If so there could not be three-legged tigers. But there is no contradiction in the idea of a three-legged tiger. Furthermore, it could be that all those who had ever seen the relevant animals had been subject to an illusion making it appear that they have four legs whereas in fact they all have three legs. Such a discovery would not lead us to say that there are no tigers. Rather, we would say that tigers are three-legged animals, not quadrupedal.

Kripke also considers whether the properties in the proposed definitions of 'gold' and 'tiger' are sufficient for membership of the kinds. We should pause to note that when Kripke talks of the proposed definitions he uses the indefinite article (as one naturally does): 'gold is *a* yellow metal', 'a tiger is *a* large carnivorous quadrupedal feline'. These could be read as implying that the properties listed are *not* to be taken as jointly sufficient for kind membership. If gold is *a* yellow metal, then perhaps there are other yellow metals that are not gold. Still, I think there is a reading on which the definiens properties are intended as jointly sufficient. Kripke assumes as much and then argues that something could satisfy the definiens without being of the kind. So something could have all the identifying features of gold without being gold—iron pyrites (fool's gold), he suggests. Iron pyrites is not a very good example, since the definition we are working with is 'yellow metal' and iron pyrites is not metal, nor does it have the readily observable properties of a metal (being ductile, malleable, etc.). Nonetheless we can understand the reference to fool's gold as intending some substance, real or imaginary, that has all the relevant allegedly defining 'superficial' properties of gold. Indeed, there is such as substance if the definition is 'yellow metal', since brass is a yellow metal; indeed, a kind of brass, Muntz metal (famous for sheathing the hull of the *Cutty Sark*), is also known as 'yellow metal'.

Clearly Muntz metal, an alloy of copper, zinc, and iron, is not gold. Likewise, if we discovered some hitherto unknown animals that resemble tigers in all the allegedly defining respects, but which were not mammals but reptiles, we would not say that some tigers are reptiles, but that these animals are not tigers.[1]

Arguments (i.a) and (i.b) establish that for some kinds at least, the 'superficial' properties by which we very frequently identify members of a kind are not properties that that are individually necessary or jointly sufficient for kind membership. So which properties are necessary and/or sufficient for kind membership? Kripke says that given that it is true that gold has atomic number 79, nothing that fails to have atomic number 79 would be gold. However similar something is to gold, if we found that it lacks this property, we would say that it is not gold. So having atomic number 79 is necessary to being gold. It is clear also that scientists had to discover that gold has atomic number 79, so this is a necessary aposteriori truth about the kind. This establishes claim (VI).

Kripke takes similar arguments to establish (V). Consider a non-actual world where there are cat-like entities that are in fact demons (1980: 126). It would not be that case that these are cats that are not animals. Rather, they would not be cats at all. So, Kripke concludes, it is necessary that cats are animals.

Now consider again the case of gold and the property of having atomic number 79. The arguments so far considered show that being made entirely of stuff with atomic number 79 is necessary for being gold. That is consistent with there being other stuff that is not gold but which also has atomic number 79, that is, with there being multiple types of stuff with atomic number 79, only one of which is gold. While Kripke does not directly argue that having atomic number 79 is sufficient for gold, the many arguments for (i.a) support this conclusion. Consider any property distinct from having atomic number 79 that one might consider necessary for being gold (e.g. being yellow in colour). By considering stuff with atomic number 79 but without these properties (e.g. that the stuff we dig out of gold mines is a blue metal), we conclude that none of these is necessary but that having atomic number 79 is necessary. But if no other property is necessary for being gold, then it is reasonable to conclude that having atomic number 79 is sufficient for being gold. So having atomic number 79 is necessarily both necessary and sufficient for being gold. This establishes (I).

22.4 Essentialism

Kripke calls identities that are instances of (I), e.g. that gold is the element with atomic number 79, 'theoretical identities' because they were once unknown truths that scientists had to hypothesise and then confirm with appropriate experiments and observations. Propositions of the forms in (V) and (VI) are likewise theoretical claims established by aposteriori scientific investigation. That much is clearly right. For (VII) to be true, it also needs to be the case that these aposteriori necessary properties, sub-kind relationships, and identities characterise features of the kind that are (a) explanatory in a basic way, and (b) its (partial) nature and essence. For (II) to be true, instead of (b) we need (b′), its (full) nature or essence.

It looks correct to say that the properties, etc., in question are explanatory in a fundamental way. The atomic number of an element determines its electron structure, which in turn determines is chemical behaviour, its colour, electrical conductivity, and so on. So atomic number is explanatory in a basic way that these other properties are not. That lightning is a form of electricity explains why it can burn what it strikes, why it can disrupt electrical equipment, why it can be attracted to a metal lightning rod, why it will charge a Leyden jar, and so on. So (a) is true.

What about (b)? In his brief argument for (II) (and (VII)) Kripke does talk of scientists discovering the 'very nature' of the substance (Kripke 1980: 124). But he does not at that point talk

about the essence of the kind. Later he does so, saying that, 'In general, science attempts, by investigating basic structural traits, to find the nature, and thus the essence (in the philosophical sense) of the kind' (1980: 138). But he does not really argue for these claims about nature and essence. It is as if he thinks that once one is persuaded of the relevant aposteriori necessities, it will be obvious that they concern the nature and essence of a kind.

Very plausibly our intuitive notion of 'nature' supports that. What is the nature of gold? It is the element with atomic number 79. What is the nature of water? H_2O. What is the nature of lightning? A form of electric discharge. What about essence? In *Naming and Necessity* Kripke does not discuss essentialism in detail. Essential properties are necessary properties. But are all necessary properties essential properties? Kripke does not say. Even so, the quoted sentence suggests that in his view, essential properties are those necessary properties that reflect the nature of the kind in question. So a plausible hypothesis to attribute to Kripke says that the concept of essence is much like the concept of nature, except that it also includes a modal commitment: whatever is an essential property of a kind X is necessarily necessary for being an X; whatever is the essence of a kind X is necessarily sufficient for being an X. This would fit with Fine's (1994) distinction between necessity and essence. Not all necessary properties are essential: it is a necessary but not essential property of *Socrates* that he is a member of the singleton set containing Socrates, whereas it is an essential property of that *set* that Socrates is a member of it. For Fine the difference between necessity and essence is that essence concerns the nature or identity of the entity in question. And so we can establish the essentialist elements of (b) and (b′), since the properties, etc., in question not only concern the nature of the kinds, they do so necessarily.

22.5 Putnam on natural kinds and modality

In 'The meaning of "meaning"', Putnam puts forward arguments that have conclusions very similar to Kripke's. Putnam (1975: 232) is far more explicit than Kripke that he holds these modal conclusions to be consequences of a semantic theory regarding natural kind terms:

> What Kripke was the first to observe is that this theory of the meaning (or 'use' or whatever) of the word 'water' (and other natural-kind terms as well) has startling consequences for the theory of necessary truth.

The theory of meaning and the related claims about modality are developed primarily in Putnam's famous 'Twin Earth' thought experiment. We imagine that there is another planet in our galaxy, Twin Earth, that is with respect to its superficial properties identical to our planet Earth. Some of its inhabitants speak a language almost indistinguishable from English, and they refer to the colourless, odourless, thirst-quenching liquid that fills the seas, lakes, and rivers of Twin Earth as 'water'. It turns out that this liquid, superficially just like water, has a complicated chemical formula quite unlike 'H_2O' which we may abbreviate by 'XYZ'. Putnam asserts:

a. On Twin Earth the word 'water' means XYZ.

whereas

b. On Earth the word 'water' means H_2O.

This has consequences for what the inhabitants of the two planets say: on Earth some people (those with an education in chemistry) will say 'water is H_2O' whereas their Twin Earth

counterparts will say 'water is XYZ'. However, if we turn the clock back to 1750, before the chemical revolution, such differences will disappear: the utterances of Earth-dwellers and their Twin Earth counterparts will be indistinguishable, and, says Putnam (1975: 224), they will be exact duplicates in 'appearance, feelings, thoughts, interior monologue, etc.'. The members of such pairs are 'in the same psychological state'.

Assuming that the extension of 'water' as used on Earth has not changed between 1750 and the present day (and so likewise on Twin Earth), this gives Putnam his principal conclusion: it is not the case *both* that for every term there is a meaning/intension that determines its extension, *and* that knowing this meaning/intension 'is just a matter of being in a certain psychological state'. This conclusion is semantic externalism. The argument for that conclusion depends on there being a reasonable sense of 'psychological state' such that the Earth-dwellers and their Twin Earth counterparts are in the same psychological state. Some might assert that any reasonable sense of 'psychological state' is such that these psychological states are themselves to be understood externalistically. Still, it will remain the case that two individuals can use words that have differing extensions despite being in identical brain states: 'Cut the pie any way you like, meaning just ain't in the head' (Putnam 1975: 227).

Putnam maintains that we would assent to (a) and (b) and that this is explained by his theory of the meaning of 'water' and other natural kind terms, which is as follows. He supposes that we give a meaning explanation for such terms by use of a particular sample of the kind in question—for example, a glass of water—and of a cross-world 'same kind relation'.[2] Such a meaning explanation may proceed by pointing to the glass of water and declaring 'this is water' while intending the following (where w ranges over possible worlds, and x ranges over particulars):

> (S) $\forall w \, \forall x$ in w (x is water $\leftrightarrow$ x bears the same kind relation to the entity referred to by 'this' in the actual world).

It is intending this that makes 'water' a rigid designator.

Putnam assumes (but does not explicitly state) something like:

> (MC) For many substance kinds such as water, something bears the same kind relation to a sample of the substance if and only if it has the same microstructural constitution (e.g. chemical composition) as the sample.

If the stuff in the glass on Earth is H_2O, then (S) and (MC) imply:

> (N) In all possible worlds, something is water if and only if it is H_2O.

And because (S) is part of a meaning explanation for 'water', we can say not only that water is H_2O, we can also say that 'water' means H_2O. Hence (b) is true (and, *mutatis mutandis*, so is (a)).

It is an aposteriori truth that water is H_2O. So (N) and its derivation imply something related to (I):

> (I′) For some natural kinds, A and B, it is a necessary aposteriori truth that something belongs to A iff it belongs to B.

This is an example of what Mackie (2006) calls 'predicate essentialism'.

22.6 The semantics of natural kind terms and modality

Putnam's argument appears to derive the essentialist conclusion (I′) from his theory of meaning, and the quotation at the beginning of Section 22.5 implies that he thinks that Kripke makes a similar derivation. Salmon (1979) argues that no such derivation can be made—attempts to do so rely on hidden essentialist assumptions.

Consider the argument for (N). This depends crucially on (MC), which Putnam does not discuss but just takes as given. Without (MC) there is no argument for (N). Consider an alternative to (MC):

> (SC) For many substance kinds such as water, something bears the same kind relation to a sample of the substance if and only if it shares the same superficially observable characteristics (e.g. appearance, taste, aroma, etc.) as the sample.

If (SC) were true, then we could not derive (N). Samples of XYZ would also be water. For that matter, there might even be possible worlds with slightly different laws of nature where H_2O does not have the superficial properties it has in the actual world. If so, some possible samples of H_2O are not water. So (N) would be false as would be (I′).

Where does this leave the argument for essentialism? Salmon (1979) thinks that Kripke's essentialist claims are independently plausible, whereas others, such as Mellor (1977), who take a similarly negative view of the derivability of essentialism from semantics, regard this as a reason to reject essentialism—what its supporters take to be the key argument for essentialism is unsound.

It is far from clear that Kripke—unlike Putnam and despite what Putnam says—does think that the argument for essentialism must proceed via a semantic route. As I have described it, Kripke's discussion starts with (i.a) and (i.b) before concluding with the essentialist (ii). Since (i.a) and (i.b) are the propositions that one would need (and which Kripke uses) to defeat a Fregean account of meaning, it is perhaps natural to suppose that what he is doing in *Naming and Necessity* is, first, arguing against Fregeanism about meaning, then, secondly, arguing in favour of an alternative semantics in which the idea of rigid designation plays a major role, and, thirdly, extracting from the latter an argument for essentialism. A different view of Kripke's strategy is that in establishing (i.a) and (i.b) he is providing premises from which, along with further plausible premises, the rejection of Fregeanism will follow. That in turn motivates the alternative semantics. But the argument for (ii) is independent of that alternative semantics. That proceeds naturally in the wake of the arguments for (i.a) and (i.b). For if the superficial properties of a natural kind do not characterise what that natural kind is, which, if any, properties do characterise it?

In favour of this interpretation is the fact that in the pages discussing (i.a), (i.b), and (ii) there is no in-depth discussion of semantics, and none of his own alternative account of reference and meaning. Kripke does not present his essentialism about natural kinds as deriving from his semantic theory. Just as I have outlined it earlier in this chapter, Kripke's argument is simple and straightforward. He asks, 'Given that gold *does* have the atomic number 79, could something be gold without having the atomic number 79?' He then considers counterfactual circumstances in which a substance, however similar to gold in superficial respects, was not an element with atomic number 79. And he declares that we would not say that such a substance is gold. Far from using semantic theory, Kripke appeals to our philosophical 'intuitions' about counterfactual situations in the manner that is familiar throughout philosophy (Williamson 2007).

So what of Putnam's argument, which Putnam does seem to think depends on semantic theory? That argument's principal aim is also to reject Fregeanism, so inevitably it draws on

semantic theory. The essentialist conclusion is presented as a corollary. So the key question is, which parts of the overall argument does the essentialist corollary require? Note first that if essentialism was all that Putnam was interested in, he could have proceeded just like Kripke, asking 'Given that water *does* have the chemical composition H_2O, could something be water without being H_2O?' The Twin Earth thought experiment provides us with the counterfactual example, in which we judge that the substance on Twin Earth, for all its superficial similarity to water, is not water because its chemical composition is XYZ, not H_2O. That argument, exactly parallel to Kripke's, does not use anything not found in Putnam's argument. It may not appear this way because the consequences of the thought experiment on which Putnam concentrates are (a) and (b), which are metalinguistic rather than object-language claims—these are supposed to be linguistic discoveries that the Earthian and Twin-Earthian travellers make. Putnam's discussion makes it clear that (a) and (b) are much the same as:

(a′) On Twin Earth the extension of the word 'water' is XYZ (and so not H_2O).

whereas

(b′) On Earth the extension of the word 'water' is H_2O (and so not XYZ).

However, (b′) could not be asserted if one thought that a sample of XYZ were water. I suggest then that although (b′) is a metalinguistic claim, it also embodies the judgment that the counterfactual XYZ would not be water, and it is this judgment that delivers the essentialist conclusion. Putnam is not explicit that this judgment is one that we would make. But he himself makes that judgment when he sets up the thought experiment in saying, 'I shall suppose that the oceans and lakes and seas of Twin Earth contain XYZ and not water, that it rains XYZ on Twin Earth and not water, etc.' (1975: 223). Putnam *should* have said 'I shall suppose that the oceans and lakes and seas of Twin Earth contain XYZ and not H_2O'. Assuming that he had said the latter, it would then be a matter of an additional non-linguistic judgment that the oceans, etc., of Twin Earth do not contain water. From that judgment (along with trivial linguistic propositions) claim (a′) follows. Independently, the essentialist claim also follows, that necessarily XYZ is not water, and we are at least close to the essentialist target, that necessarily water is H_2O.

Nonetheless, Putnam does think that (a)/(a′) and (b)/(b′) are explained by his semantic theory. So cannot essentialism *also* be derived from the semantic theory? As we saw earlier, these judgments are explained by three components of the semantic theory: (i) pointing to a sample that in fact is H_2O (on Earth; but XYZ on Twin Earth) and declaring 'this is water'; (ii) intending (S); (iii) the proposition (MC). But for the essentialist conclusion, we do not need (i) and (ii). (N) also follows from (MC) plus:

(S′) $\forall w \forall x$ in w (x is water $\leftrightarrow$ x bears the same kind relation to canonical water-samples in actual world.

and the fact that canonical water samples in the actual world are H_2O. These propositions are all in the object language. So again, it is nothing specific to the semantic theory that is doing the work in generating essentialism. Clearly it is (MC) that is doing the heavy lifting. As mentioned, Putnam does not explicitly assert (MC) but it is clearly implied by what he says.

The conclusion we should draw is that essentialism cannot be derived from the core of the semantic theory. But that need not undermine the case for essentialism. For Kripke's argument (as discussed in Section 22.3) does not use the semantic theory at all, depending instead on

intuitions concerning the kinds in question. And although Putnam's argument looks as if it uses the semantic theory, in fact it does not, depending instead on a general claim about kinds that must seem to him so obvious and unremarkable that it does not need stating explicitly at all. The defender of essentialism should therefore say that the argument for essentialism is simple—it just is intuitively highly plausible, as revealed either by counterfactual thinking or by our willingness to make tacit essentialist assumptions.

At the outset I noted that (I) is supported by (III), and that too looks like a claim about essentialism derived from a semantic claim. The response is the same: an argument for essentialism from rigid designation cannot provide independent grounds for asserting essentialism. For the very considerations that support the view that 'water', 'gold', and 'tiger' are rigid designators are those that directly support essentialism without a detour through semantics. It is because our counterfactual thinking does not track the superficial characteristics of these kinds that essentialism is plausible. And it is because our counterfactual thinking does not the track superficial characteristics of these kinds that it is plausible that 'water', 'gold', and 'tiger' each designate one kind across all possible worlds, where sameness of kind is a matter of the sameness of basic explanatory (usually structural) characteristics.

Although Salmon is right that essentialism cannot be derived from the semantics of natural kind terms, the discussion in the preceding paragraph shows that the semantics and modal metaphysics of natural kinds are not entirely independent. The point can be extended to suggest that rather than essentialism being derivable from semantics, it is semantics that is sensitive to modal metaphysics. Let K be a natural kind of the sort covered by (MC). So whether a particular belongs to the kind K depends on whether it has the right microstructural constitution. But the microstructures of things and kinds are properties of which we can be entirely ignorant and which we may not even be able to conceptualise—in 1750 people were not able to have the concept 'possessing the microstructure H_2O'. The Fregean theory requires that the extension of any term is determined by properties that the user of the term is able to conceptualise. So Fregeanism implies that we can have a term that refers to K only if we know what the microstructure of K is. So, in particular, (some) speakers today can use 'water' to refer to the kind whose microstructure is H_2O—because we can use 'having microstructure H_2O' as an intension that will determine the correct extension. But no speakers in 1750 could do that.[3] More generally, if essentialism is true about natural kinds, Fregeanism implies that scientifically ignorant speakers cannot refer to many natural kinds.[4] Consequently, if one thinks that essentialism is true and that the scientifically ignorant can refer to these natural kinds, then one must reject Fregeanism.

22.7 Objections to natural kind essentialism

One line of attack on Kripke–Putnam natural kind essentialism takes as its target the semantics of natural kinds terms that Kripke and Putnam develop—Mellor (1977) is a prime example of this approach (although Mellor has other arguments besides). Since, as I have argued, the support for natural kind essentialism comes from elsewhere, not from the semantic theory, I shall not outline such objections here. The three principal further kinds of response are:

(i) To reject the allegedly simplistic theoretical identifications upon which Kripke and Putnam rely. In particular, science tells us that it is not strictly correct to say that water is H_2O.
(ii) To argue that we do not in fact classify samples/items in a manner consistent with essentialism.
(iii) To question the reliability of the intuitions on which the arguments for essentialism are based.

22.7.1 We should not be essentialists, because water is not H_2O

Kripke and Putnam argue that the essential properties of many natural kinds are properties detailing some aspect of the microstructure of examples of the kind. The theoretical identities mentioned in (I) are often an identity formed by using a name of a kind on the one hand and a microstructural description on the other hand: water is H_2O, gold is the element with atomic number 79, and so on. A number of philosophers of science have argued that the science on which these identities is based is too simplistic. The microstructure of water just is not H_2O. Needham (2011) and van Brakel (1986), for example, argue for this on several grounds. First, 'H_2O' is not a microdescription at all, since what it tells us is that water is 'a compound of hydrogen and oxygen in the proportions 2 moles to 1', which is a macroscopic description (Needham 2011: 9). Secondly, the microstructure of liquid water is very different from that of water vapour and that of ice. So there is no such thing as *the* microstructure of water. Furthermore, they argue, certain clearly macroscopic descriptions of water *do* suffice to identify water, for example thermodynamic properties such as the triple point (the temperature and pressure at which the gaseous, liquid, and solid phases of a substance co-exist). Thirdly, not all H_2O is water, because D_2O, deuterium oxide, is a species of H_2O, but we would not and should not regard it as water (LaPorte 2004). This is because D_2O has different rates of reaction from the rates of reaction of H_2O, and this has significant consequences—D_2O can be toxic for some animals.

While these observations may require essentialists to be clearer about how their view should be articulated with respect to the particular case of water, they do not seem to undermine the case for microessentialism in general. Nor do they unequivocally refute microessentialism about water in particular (Hendry 2006). Because it is chemically surprisingly complex, water presents particular problems in getting the details right that other chemical substances do not. Indeed, Needham (2011: 4) accepts that gold's atomic number 79 is a sufficient microstructural condition for identifying gold. First, to be told that water is a compound of hydrogen and oxygen is to be told something about its microstructure, that it contains atoms of oxygen and hydrogen and not any of kind of atom. Secondly, although the different phases of water do have different microstructures, that is not to say that there is nothing that those microstructures have in common: for any sample of pure water, whether solid, liquid, or gas, is composed (almost) entirely of H_2O molecules.[5] Thirdly, there are good reasons for thinking that D_2O is a species of water (after all, it is called heavy water). While reaction rates are different for D_2O from 'normal' water, this is the same for any compound where there are isotopic variants for the constituent elements (which is to say: almost every compound).

22.7.2 We are not essentialists when we classify things as water

A distinct line of criticism of essentialism argues that what we regard as water is not determined by its microstructure. Some items that are largely H_2O are regarded as water, such as sea water and tap water, but others that are no less composed of H_2O are not regarded as water, such as tea and Sprite (Malt 1994; Chomsky 1995). Abbott (1997) responds first by remarking that we do say that tea is composed mostly of water precisely on the basis that it is mostly composed of H_2O. She proposes that our different classifications of impure substances are just another example of vagueness, whereby we tolerate a degree of imprecision that is appropriate to the context. And it does seem correct that we regard impure substances as water if their differences from pure water are irrelevant or negligible in the context—mineral water and herbal tea may be similar in their proportion of non-water constituents, but in the context of request for water with a meal in a restaurant, the impurities in the mineral water that make it differ from pure water are irrelevant, but corresponding differences between tea and pure water are relevant in that context.

22.7.3 We are essentialists, but this is just a cognitive bias

Philosophers' use of intuition as evidence for or against philosophical theories of all kinds has been subject to criticism (e.g. Swain et al. 2008). The intuitions to which Kripke and Putnam appeal have been equated with *psychological essentialism* (Leslie 2013). If psychological essentialism is an unreliable source of intuition, then we should not readily rely on Kripke–Putnam type intuitions, despite a strong human tendency to do so. Psychological essentialism is the thesis that humans have a cognitive disposition to believe that the underlying intrinsic structures and mechanisms of things are better predictors of behaviour than more superficially observable characteristics, and it is such underlying mechanisms that determine the nature and identity of a thing. Numerous studies show that this cognitive disposition is found in young children. Whether such findings are sufficient to undermine Kripke–Putnam intuitions depends on whether those cognitive dispositions are reliable. Leslie points to the rejection of microessentialism about water (as discussed earlier) as well as similar arguments against microessentialism about species as evidence for the unreliability of these cognitive dispositions. Note that the intuitions that Kripke and Putnam appeal to are not simply 'water has an essence that is its internal structure'. Rather they are, strictly, intuitions of the form, 'given that scientists tell us that all and only water around here is H_2O, then something would not be water if it is XYZ'. Now that intuition is not shown to be wrong by the finding (if correct) that it is false that scientists tell us that all and only water around here is H_2O. More generally, it has not been shown that the cognitive dispositions underlying such intuitions are unreliable when supplemented by philosophical reflection and correct scientific information.

22.8 Conclusion

I have outlined in this chapter some of the founding arguments for essentialism about natural kinds. Like so many philosophical arguments, these depend on 'intuitions'—quasi-intuitive judgments about counterfactual cases. Although it looks as if Putnam and (to a much lesser extent) Kripke think that the essentialist conclusions can be derived from semantic premises, this is an illusion since those semantic premises themselves contain substantive essentialist commitments. One could take this, as I do, to show the deep-seated nature of those intuitions. One could instead argue that this is wrong—we do not have such deep-seated intuitions; indeed, our common-sense classificatory practices are incompatible with essentialism. Even if one accepts that such intuitions are there, those do not show that they are reliable. They may be just the product of cognitive biases or of scientific ignorance (or of both working together).

Notes

1 For this purpose, one has to take 'feline' not to name a biological category, but to mean 'cat-like in appearance'.
2 Putnam talks of a 'same liquid relation', but that introduces avoidable difficulties concerning kinds, including water itself, whose samples may exist in gaseous or solid states as well as the liquid state.
3 Conceivably they could do so if they used an intension which is necessarily co-extensive with 'having microstructure H_2O'. It is an interesting question whether such an intension could have been available to speakers of English before the chemical revolution.
4 This in turn implies that if, being scientifically informed, we do now use terms such as 'water' to refer to natural kinds, then there has been meaning change over time. In that sense although essentialism is true, knowledge that what we refer to as 'water' has the essence 'H_2O' is apriori knowledge resulting from a semantic decision to change the meaning of 'water'. Something like this is Joseph LaPorte's (2004) view.

5 The 'almost' here refers to the fact that in liquid water a small proportion—fewer than 6 per billion at 37°C—of the molecules dissociate into OH^- and H_3O^+ ions. Hendry (2006) regards this dissociation as the biggest obstacle to simple microessentialism. Hendry regards water as what you get when you bring H_2O molecules together—what you get includes the ions produced by the dissociation caused by bringing the molecules together in a liquid.

References

Abbott, B. (1997) "A note on the nature of 'water'" *Mind* **106**, 311–19.

van Brakel, J. (1986) "The chemistry of substances and the philosophy of mass terms" *Synthese* **69**, 291–324.

Chomsky, N. (1995) "Language and nature" *Mind* **104**, 1–61.

Fine, K. (1994) "Essence and modality" in J. Tomberlin (Ed.), *Philosophical Perspectives 8: Logic and Language*, pp. 1–16. Atascadero, CA: Ridgeview.

Hendry, R. (2006) "Elements, compounds and other chemical kinds" *Philosophy of Science* **73**, 864–75.

Kripke, S. (1980) *Naming and Necessity*. Oxford: Blackwell.

LaPorte, J. (2004) *Natural Kinds and Conceptual Change*. Cambridge: Cambridge University Press.

Leslie, S.-J. (2013) "Essence and natural kinds: When science meets preschooler intuition" in T. S. Gendler and J. Hawthorne (Eds.), *Oxford Studies in Epistemology*, Vol. 4, pp. 108–65. Oxford: Oxford University Press.

Mackie, P. (2006) *How Things Might Have Been*. Oxford: Oxford University Press.

Malt, B. C. (1994) "Water is not H_2O" *Cognitive Psychology* **27**, 41–70.

McGinn, C. (1976). "On the necessity of origin" *Journal of Philosophy* **73** 127–35.

Mellor, D. H. (1977) "Natural kinds" *British Journal for the Philosophy of Science* **28**, 299–312.

Needham, P. (2011) "Microessentialism: What is the argument?" *Noûs* **45**, 1–21. (This paper lays out the problems for the microessentialism of Kripke and Putnam.)

Putnam, H. (1975) "The meaning of 'meaning'" in H. Putnam (Ed.), *Mind, Language and Reality: Philosophical Papers*, Vol. 2, pp. 215–71. Cambridge: Cambridge University Press.

Salmon, N. U. (1979) "How not to derive essentialism from the theory of reference" *Journal of Philosophy* **76**, 703–25.

Swain, S., J. Alexander, and J. Weinberg (2008) "The instability of philosophical intuitions: Running hot and cold on Truetemp" *Philosophy and Phenomenological Research* **76**, 138–55.

Williamson, T. (2007) *The Philosophy of Philosophy*. London: Wiley-Blackwell.

Chapter 23
MODALITY IN PHYSICS

Samuel C. Fletcher

23.1 Introduction

This review concerns the notions of physical possibility and necessity as they are informed by contemporary physical theories and the reconstructive explications of past physical theories according to present standards. Its primary goal is twofold: first, to motivate and introduce a range of accessible issues of philosophical relevance around these notions; and second, to provide extensive references to the research literature on them. Although I will have occasion to comment on the direction and shape of this literature, pointing out certain lacunae in argument or scholarly attention, I intend to advance no overriding thesis or point of view, aside from the selection of issues I deem most interesting.

I have grouped these issues into four categories by the domain of physical phenomena to which they pertain, with two or three issues in each, as follows.

1. Those that arise across physical theories of various sorts (Sections 23.2–23.5):
   (a) the modal commitments of accepting a physical theory, especially through the lens of structural realism and anti-realism (Section 23.3);
   (b) the notion of physical possibility seemingly invoked in variational principles throughout physical theory (Section 23.4); and
   (c) the role of physical possibility and necessity in accounts of symmetry in physics (Section 23.5).
2. Those in thermodynamics and its relation to statistical physics (Sections 23.6–23.8):
   (a) the possibilities allowed or forbidden by attributions of (ir)reversibility to thermodynamic processes, and how these possibilities interface with those allowed or forbidden by statistical mechanics (Section 23.7); and
   (b) the sense in which the adiabatic accessibility relation in the Lieb-Yngvason axiomatic formulation of thermodynamics is a modal accessibility relation (Section 23.8).
3. Those in spacetime theory, especially in general relativity (Sections 23.9–23.12):
   (a) the interpretation and ontology of spacetime events (Section 23.10);
   (b) the delineation of the physically possible worlds according to general relativity, viz., the *physically reasonable* spacetimes (Section 23.11); and
   (c) definitions and representations of (in)determinism (Section 23.12).

4. Notions of possibility and actuality in the interpretation of quantum theory (Sections 23.13–23.15), in particular:
   (a) in the many-worlds interpretation (Section 23.14) and
   (b) in the modal interpretation (Section 23.15).

I have not included any focused discussion of the role of modality in physical law,[1] or of modal reasoning with physical theories, for these are covered elsewhere in this *Handbook*. (But, see the discussion in Section 23.3 for some allusions.)

I have also not included much discussion of modality in quantum gravity (Muntean 2015) or in experimental physics. That's not to say that there are no interesting issues there. For instance, assumptions about what kinds of events are possible in a particle collider such as the Large Hadron Collider guides which data is discarded through hardware filtering before it even reaches analysis—a necessity when too much data is produced to even record (Perovic 2011). Thus, such experiments are laden by the modal assumptions of theories that physicists decide to probe (Karaca 2017). But these modal commitments could be understood through those of theories and models of experiment, data, and data acquisition (Suppes 1966; Karaca 2018), and so they don't obviously constitute a separate category. Indeed, the next Sections 23.2–23.5 discuss issues in the modal commitments of physical theory.

23.2 General issues: introduction

Each modern physical theory posits a collection of mathematical models that represent a range of physical phenomena or states of affairs within that theory's scope. Usually, each model of a theory represents some way or ways for that phenomena or state of affairs to be—its representational capacities (Fletcher 2020). This observation alone raises several general issues for modality in physical theory.

23.3 General issues: physical theory and modal commitment

One quite general question that arises in the interpretation of a physical theory is to what modal facts one commits when one commits to such a theory (Smeenk and Hoefer 2015: §4). For example, what does it mean to be a realist about physical possibilia that are not actual? Ought the realist thus commit themselves? Ought one commit thus to realism?

Constructive empiricists, while recognizing the need for modal reasoning in science (and physics in particular), abjure any realist commitment (van Fraassen 1980: ch. 5.4). Instead, they seek only to use modal talk as a means for commitment about observable phenomena. Some progress on an account of modal reasoning in science (and, again, physics in particular) without presupposing realist commitment has been made (Muller 2005; Fletcher 2019), but most prefer to accept some measure of realism.

Whether this belief beyond observables forms a mere modicum or a mound depends on who you ask. Some accept realist commitments to particular models, but not to the relations between models that represent modal structure (Dieks 2010; Brading 2011). Others are realists about physical laws, letting their preferred account of how laws ground physical necessity—of which there are a great many—discharge their modal commitments (Carroll 2016). Still others take modal structure to be real and irreducible, understood as arising from causal powers (Esfeld 2009; Esfeld and Lam 2011) or on their own (Ladyman 1998; French and Ladyman 2003; Berenstain and Ladyman 2012; French and McKenzie 2012; Ladyman 2018). Intellectual skirmishes between these different camps have focused on whether careful attention to the features

of physical theories themselves and how they are successfully applied rules in favor or against particular metaphysical theses—see, for example, Ladyman (2000), Monton and van Fraassen (2003), Ladyman (2004), French and McKenzie (2012), and Esfeld (2018).

23.4 General issues: variational principles

Most discussions of the modal commitments of a physical theory assume a binary classification of putative models—possible or not—and that explanations involving that theory use only a single model of the theory representing the phenomena explained. The widespread use of *variational principles* across physics seems to challenge both of these convictions. Roughly speaking, a variational principle is a rule for finding solutions to an equation—that is, models of a theory—that requires some property of the solution to be locally extremized—minimized, or maximized, compared with similar putative solutions. To do so, one considers putative solutions some of whose properties are variations—that is, perturbations—of the target solution and demonstrates that this target is in fact a true solution in virtue of its minimizing or maximizing a certain quantitative property. For example, Fermat's principle, one of the earliest used in physics, states that a light ray traveling from one point in space to another traverses the path that minimizes its duration of travel. Modern Hamiltonian and Lagrangian mechanics, and Hamilton-Jacobi theory, also use such principles, focusing instead on extremizing a quantity called the *action* of a solution (Yourgrau and Mandelstam 1960; Rojo and Bloch 2018). They are used throughout classical and quantum physics, including their statistical versions, spacetime theory, and thermodynamics.

The first challenge these principles present to the straightforward view of modal commitment is that when a particular model obtained using them represents some actual physical phenomena, it does so in virtue of the relations that model has to *other* models. In other words, it violates what Butterfield (2004b) calls the (weak) truthmaker principle, that what is actually true should supervene on the actual facts. Although there is extensive debate within metaphysics about the nature of such a principle, the challenge presented here is different in kind from those usually considered, which focus on the logical and metaphysical dimensions of propositions and facts.

The second challenge is that some of these principles consider variations which yield mathematical models that are not *themselves* models of the theory considered. This is in fact a double challenge, and corresponds to what Butterfield (2003, 2004a, 2004b) calls the third grade of modal involvement. Its first aspect is a magnification of the first challenge: not only does a model representing phenomena seem to do so in virtue of other models, but also in virtue of ones not even possible according to the theory. How can what's physically true supervene on what's physically impossible? The second aspect of this challenge is that because the variational principles quantify over models representing physically impossible phenomena, the principles seem both to commit to their possibility in some sub-physical or metaphysical sense, but not in the physical sense that the theory's models delineate.

Butterfield (2004a) argues that one can meet the first challenge first by showing that variational principles of the sort described can be understood as an infinite conjunction of counterfactual conditionals, so that the usual analyses of such conditionals compatible with the truthmaker principle can be applied. Alternatively, one can argue that the principles, though practically useful, are also equivalent to standard differential equations of motion, which make no appeal to counterfactual or counterlegal possibilities (cf. Section 23.11's discussion of modal properties of spacetimes used to define which spacetimes are physically possible). They are thus dispensable in principle, so that one may take an instrumentalist attitude towards them.

This insight, which also applies to the second challenge (Butterfield 2004b), seems to be a more satisfactory resolution of the challenge because it does not hinge on perhaps controversial assumptions about the semantics of counterfactual and counterlegal conditionals and better reflects the practice of mathematical physics, in which it is quite common to solve a problem by the output of an evinced algorithm that takes as input the solution to an entirely *different* problem. But, one might instead *embrace* the different levels of modal commitment in these principles as providing an impetus for new metaphysics (Terekhovich 2018).

23.5 General issues: symmetries

A bijective transformation on a space may preserve certain properties of its elements, and when it does, it is said to be a symmetry on that space of those properties. For example, consider the property of being a model of a particular theory. A bijective transformation that preserves this property is said to be a symmetry of the theory. Similarly, such a transformation that acts as an isomorphism on a model is said to be a symmetry of that model. Symmetries can be discrete, such as time-reversal, or continuous, such as spatial and temporal translation.

There is a voluminous literature on the role of symmetry in physics (Brading et al. 2017); here I confine the discussion to two related remarks. The first is that having a symmetry is a modal property of a model of a theory: it is defined essentially in terms of its relations with other models. For instance, whether a particular mode of a thermodynamic process is time-reversal invariant depends not just on the process itself, but what other processes are considered possible (Uffink 2001: §3)—cf. Section 23.7. So, debates about the symmetries of models of physical theories, such as that over time-reversal invariance, are debates about the ineliminably modal content of the theories.

The second remark is that the modal content of a physical theory can play a role in its interpretation, including its ontology, by way of its symmetries. In a word, what is the metaphysical nature of the modal features of symmetries? Should they be understood dispositionally, or in other terms (French 2017; Livanios 2018)? Some have argued as well that the modal structure of symmetries makes them more fundamental than, hence explanatorily prior to, any conservation laws that hold just in case such a symmetry does (Lange 2007).

23.6 Thermodynamics: introduction

Thermodynamics is the branch of physics concerned with heat, the relation of heat's interchange with mechanical work and other forms of energy, and auxiliary concepts pertaining to these relations, such as temperature and entropy. From its inception, its concerns were as much with practical engineering problems about the efficiency of engines, the generation of electricity, and the construction, design, and maintenance of machines as with scientific laws governing the phenomena in its domain of application (Uffink 2007). Far from an historical curiosity, this mixed genealogy is both apparent and important in the interpretation of the theory today, including certain of its modal features.

23.7 Thermodynamics: (ir)reversibility

Thermodynamics puts a set of constraints on the possible processes, or sequences of state change, of a thermodynamic system. Perhaps most famous of these is its Second Law. Although, as Uffink (2001) remarks, there is surprising difficulty stating exactly what the law demands and prohibits, what is relevant here is that most accounts of the law understand part of its content as a

statement of, or at least logically entailing, the impossibility of certain processes. For instance, one tentative statement of the Second Law is that the entropy of any isolated thermodynamic system cannot decrease. Thus, processes for which the entropy remains the same are also possible in reverse, while processes for which the entropy increases are not also possible in reverse. The former are accordingly called *reversible*, and the latter *irreversible*.

Because thermodynamics is a physical theory, one might just construe this notion of possibility as physical possibility *simpliciter*. The situation is not so straightforward, however, because of the theory's mixed genealogy adumbrated earlier. If thermodynamics is a theory born from engineering concerns, to what extent are its claims about what is (im)possible conditioned on seemingly contingent facts about human agents and their abilities to manipulate thermodynamic systems? Answering this question depends on a careful analysis of the scope of classical thermodynamics and its relation to other physical theories with overlapping scope, in particular statistical mechanics.

Classic thought experiments, such as Maxwell's demon (Earman and Norton 1998, 1999), may seem to suggest a positive answer. Maxwell envisaged a box of gas partitioned into two, with a small frictionless trap door in the partition operated by an intelligent demon. The demon lets only fast gas molecules through the door from one side, and only slow molecules from the other. Operating in this way, the demon increases the pressure of the gas on the first side, and decreases the pressure of the gas on the second. It is easy to show that, although the system involved is seemingly isolated, this process decreases its total entropy, in violation of the Second Law. Moreover, the classical mechanics on which kinetic theory is based is time-reversal invariant (cf. Section 23.5).

If this analysis is right, then the Second Law marks as impossible processes which are possible in statistical mechanics. Now, from a certain perspective this may not be so surprising: one expects that newer, more fundamental theories supplanting older, less fundamental theories will contradict them on what's possible. As discussed in Section 23.11, for example, general relativity seemingly allows for spacetime to have "holes" and causal irregularities of various sorts that are impossible according to pre-relativistic spacetime theory. But it does raise the question of how exactly one should understand the type of modal claim made by the Second Law. One answer is that thermodynamics does not concern physical possibility per se, but some more limited notion, such as physical possibility conditioned on the physical circumstances within the scope of thermodynamics, or perhaps some notion of *practical* possibility more closely aligned with human (engineering) capabilities.

But if the analysis is wrong, then the physical possibilities that statistical mechanics seemingly entails must be curtailed instead. There has been a huge range of responses along these lines—see Earman and Norton (1998, 1999)—but some of the most remarkable demand an accounting of the workings of the demon in information-theoretic terms (Norton 2005; Maroney 2009). Such accounting would connect physical possibility more tightly with what might have been understood as engineering concerns, namely, constraints on the processing of information. If this is right, it would provide a surprising way in which some aspects of physical possibility have precise expression in terms of practical possibility.[2]

23.8 Thermodynamics: adiabatic accessibility

One way to gain traction on the content and scope of thermodynamics is to axiomatize the theory. This is not just a pipe dream of philosophers quaintly enamored with logical empiricism, but a project that mathematical physicists take to be worthwhile. To focus on one example most relevant to the topic of modality in physics, Lieb and Yngvason (1999) have provided such an

axiomatization, for which they received the Levi Conant Prize of the American Mathematical Society in 2002, based on the binary relation of *adiabatic accessibility* between states. The interpretation of "adiabatic" is somewhat non-standard—see Uffink (2001: §11) for a discussion—but one can more or less understand it as a relation that holds between two states when the second can be reached from the first without adding any energy (except perhaps that which can be obtained from a weight falling in a uniform gravitational field). Their work is too rich to discuss in any detail here, but one feature worth emphasizing is that, among other properties, the relation is reflexive and transitive, but in general not symmetric. Importantly, the non-symmetry of this relation entails the existence of irreversible processes.[3]

Beyond questions of the nature of the modality thus invoked (and described in Section 23.7), this theory also suggests an unexplored connection with modal semantics. Recall that a "possible world" model of a modal language consists of a set W (of "worlds"), a binary accessibility relation R on that set, and a Boolean valuation v on worlds and sentences φ of the language such that v assigns "true" to $\Diamond\varphi$ at $w \in W$ if and only if for some $w' \in W$ accessible from w (i.e., wRw'), v assigns "true" to φ at w'.

The states of a thermodynamic system in the theory of Lieb and Yngvason (1999) are like W and their adiabatic accessibility like R, so the corresponding modal operator $\Diamond$ could be interpreted as "it is possible to adiabatically transform the system so that …"[4] In particular, adiabatic accessibility satisfies reflexivity and transitivity, yielding the modal logic **S4** (Burgess 2009).[5] It would be interesting to investigate whether this analogy is strong enough to support a semantics for modal reasoning within the theory along the lines suggested by Muller (2005).

23.9 Spacetime theory: introduction

Modern approaches to spacetime theory (e.g., by Friedman 1983) represent spacetime and matter using the mathematics of differential geometry, in particular a four-dimensional smooth manifold M to which one assigns fields $\Phi_1, \ldots, \Phi_m$. A four-dimensional smooth manifold is a structured space of points such that the space around each point is a chunk of R^4, where R is the real line. Each point represents an event, an occurrence or configuration of matter that has no temporal or spatial extension. Different spacetime theories—e.g., Newtonian, special relativistic, and general relativistic—employ somewhat different types of fields, but they all use fields to give extended collections of points temporal duration and spatial volume, and to represent the presence of matter at events.

23.10 Spacetime theory: ontology

What is the ontological status of events, and their relation to matter? Is one somehow prior to, or more fundamental than, the other? The opposed positions of substantivalism and relationalism take, respectively, spacetime and matter to be the one prior and more fundamental. (The names arise from an historical tradition of understanding this question in terms of the appropriate category of existence for space: substance or relation?) Whichever position one takes, one must account for the modal properties of spacetime, and events in particular. Traditional arguments about this topic focusing on whether temporally, spatially, or velocity shifted universes—cf. Section 23.5—constitute distinct possibilities can be couched in modal terms (Dasgupta 2015). (Discussion of the other modern classic argument bearing on this question, the hole argument, I will postpone for Section 23.12.) But this is so even confining attention just to how a spacetime model represents a physically possible spacetime. For instance, in the models of general relativity, certain smooth curves (the timelike ones) are picked out as the possible histories, or

worldlines, of massive particles. (These are the ones that every observer would agree are traveling slower than the speed of light.) Note that

> the modal character of the assertions (i.e., the reference to possibility) is essential. It is simply not true … that all images of smooth, timelike curves are, in fact, the worldlines of massive particles. The claim is that, as least so far as the laws of relativity theory are concerned, they could be.
>
> (Malament 2012: 121)

Generally, each model can equally well represent many different states of affairs, where particles are present or not; but one can add distinguished curves to the models to make them less idealized (Fletcher 2020).

Both substantivalists and relationalists must account for these representational properties, or else modify the theory. Subantivalists generally have an easier time meeting this demand, for they hang the modal properties on the events themselves. Relationalists must decide how these modal properties accrue to matter. Sklar (1974) raises the possibility, in the context of articulating a defense against Kant's incongruent counterparts argument against relationalism, of what has come to be known as modal relation(al)ism, the view that the relationalist should understand the events represented by the spacetime manifold as encoding the possible relations among the possible material bodies. Brighouse (1999) and Belot (2011) explicate and defend this position, especially in the light of technical results by Manders (1982). Yet objections come from substantivalists and relationalists alike (Field 1984; Earman 1989: ch. 6.12; Huggett 2006).

23.11 Spacetime theory: physically reasonable spacetimes

One of the ways that general relativity differs from its predecessor theories of space, time, and gravitation is that the properties of spacetime can differ much more significantly from model to model—that is, from one physically possible universe, or portion of a universe, to another. Some of these differences are extreme enough that physicists declare spacetimes with them "physically unreasonable"—they deny that they are genuinely physically possible after all. If one takes the models of a physical theory to represent the physical possibilities it allows, then this debate concerns the very *content* of the theory. That is, the question of which spacetime properties are physically (un)reasonable *is* the question of what's physically possible in general relativity.[6]

Most attention has focused on two broad classes of suspect properties: spacetime holes and causal irregularities (Manchak 2011, 2013: §4). One type of such hole involves seemingly "missing" events from spacetime. For a simple example, take a spacetime manifold M and remove a point from it. All the possible objects and matter that could have coincided at that event no longer can: it represents an "end" or "edge" to spacetime itself. But in this case it seems as though the spacetime "ends" prematurely. Such spacetimes are called *extendable*, and one proposal is that extendability is not a physically reasonable property for a spacetime to have (Earman 1995: 32). However, general relativity is not just a cosmological theory, but also a theory of more isolated spatiotemporal phenomena or states of affairs that have no pretension to representing an entire world. It's unclear why extendable models, when considered as more limited states of affairs, are physically unreasonable, although one must be careful about the inferences one draws from such models. Moreover, Manchak (2016b, 2017) points out that some extendable spacetimes only have extensions with other physically unreasonable properties. This puts different proposals for physically unreasonable properties in tension with one another.

Other types of holes cannot simply be filled in, because there is no consistent way to add events while extending the spacetime model's fields Φ_i to those added events. Because those fields involve the very structures that make events spatiotemporal, there is no sense in which any added points could represent the same sort of events as the others. Such spacetimes are *singular*, meaning (roughly) that there are curves representing the possible worldines of particles that end abruptly, without any means of continuing (Earman 1995: ch. 2). Yet some singularities, such as the past singularities of the Big Bang, are deemed by many not to be physically unreasonable. Although debate, proposals, and counterexamples abound, there seems to be no consensus on which sorts of singularities are physically reasonable and which are not (Manchak 2011, 2013: §4.1).

The other class of suspect properties centers on the existence of *closed causal curves*, which are widely accepted as representing a sort of time travel in general relativity (Earman 1995: ch. 6). The worldine of events representing a particle's history for such a curve is such that the time *line* for the particle is in fact a time *loop*. The claim that all physically reasonable spacetimes contain no such time travel is sometimes known as the chronology protection conjecture (Hawking 1992). Although physicists have given some evidence and arguments for why time travel is not physically reasonable, as with holes, there are also counterarguments and troves of purported counterexamples (Manchak 2013: §4.2).

These complications have led some to advocate for a simpler but much farther-reaching condition, such as a version of the so-called Cosmic Censorship Conjecture: "All physically reasonable spacetimes are globally hyperbolic" (Wald 1984: 304). Globally hyperbolic spacetimes are, in a sense, those spacetimes of general relativity whose properties are most like those of special relativity (see Wald 1984: 201 for a precise definition). In particular, while they rule out many seemingly physically unreasonable spacetimes, they seem to rule out much more besides (Manchak 2013).

Others have recently observed that some of the candidate physically (un)reasonable properties are themselves modal, since whether they obtain for a spacetime depends on what other spacetimes are considered physically (un)reasonable (Manchak 2016a). Allowing for such candidate properties seems to imply that one may not be able to give an explicit definition of the physically reasonable spacetimes, as that term appears in the definition of these properties. But some modal spacetime properties are extensionally equivalent to non-modal properties, so an explicit definition could be recovered if one replaced any of the former by the latter. For example, Hawking (1969) proved that the existence of a global time function in a spacetime is equivalent to stable causality. The former property is the existence of a consistent assignment of times that respects the causal order of events. Stable causality is a modal property of spacetime that holds when the spacetime does not allow for time travel, and the same for small perturbations thereof. But what count as legitimate perturbations themselves depend on what spacetimes are physically reasonable. However, it is unclear why an explicit definition is really necessary. Using modal conditions to delineate the physically (un)reasonable spacetimes is still compatible with their playing a role in an implicit definition thereof.

23.12 Spacetime theory: (in)determinism

Among the important modal properties that a spacetime model might have is *determinism* (Earman 2007; Hoefer 2016: §4). Roughly, a spacetime model is said to satisfy *Laplacian determinism* in a spacetime theory if and only if whenever that model and another model agree about the facts represented at a particular time-slice, they agree on all other facts represented. In a word, the state of the world at a time determines, in the theory, the state of the world for all

other times. It is a modal property because whether it obtains at a model depends not just on the model itself, but what other models are in the theory considered.

There are many variants on the definition of determinism, which modify what it means for models to agree on the facts and which facts are involved (Earman 1986: ch. 2). In particular, these variants and the debate surrounding them have been escalated through determinism's invocation in the influential hole argument (Belot 1995a; Leeds 1995). According to that argument (Earman and Norton 1987; Norton 2018), one considers two isometric relativistic spacetimes with the same manifold M but different metric fields g and g', each of which assigns temporal duration and spatial volume to collections of events. However, those considered are not entirely different: they in fact assign the same temporal durations and spatial volumes outside of some extended collection (the "hole") of events $O \subset M$ —in fact, they are identical outside of that collection. Moreover, there is a smooth bijective map ψ: $M \rightarrow M$ that acts as the identity outside of O, and within O maps regions of events with certain durations and volumes according to g to ones with identical durations and volumes according to g'. Indeed, ψ shows that the only differences between g and g' are the identities of the points to which they assign magnitudes, not the magnitudes themselves. A substantivalisist—cf. Section 23.10—seemingly must maintain that the spacetime model with g is distinct from the spacetime model with g' because they assign different such magnitudes to the same collections of points in O. Yet the facts of both models are compatible with the facts on any time-slice outside of O. Thus, the argument concludes, the manifold substantivalist is committed to indeterminism for metaphysical, rather than physical, reasons.

One influential line of response to this argument has focused on the definition of determinism, insisting that, for spacetime models to "agree" on the facts, it is not necessary for them to be identical, but only necessary that they be isomorphic, i.e., that there exists a smooth bijection ψ: $M \rightarrow M$ of the sort described earlier (Butterfield 1988, 1989; Brighouse 1994). This led to an edifying exchange of ideas in the 1990s about the nature of determinism and whether this response was sufficient to defuse the tensions of the hole argument (Belot 1995b; Brighouse 1997; Melia 1999), with some later skeptical re-evaluations (Brighouse 2008). Thus, the modal properties of spacetime are more relevant for spacetime ontology than they might at first seem.

For some, this suggests that a different structure would be appropriate to capture indeterminism in a single model, such as branching space-times (Placek et al. 2014; Müller and Placek 2018; Placek 2019). In such models, one captures the idea of an "open" future—i.e., future-indeterminism with past-determinism—by having the manifold M literally branch into alternatives. Belnap (1992, 2012) gives a version of this for Newtonian spacetimes, motivated in part also from the seemingly indeterministic and non-local outcomes of quantum experiments—see also Kowalski and Placek (1999), Placek (2000), Belnap (2002), Müller (2002), and Placek (2002). Since then, versions have been constructed for special relativity (Wroński and Placek 2009; Placek 2010; Müller 2013) and general relativity (Placek 2014). Part of the significance of this program, revealed in the face of criticism (e.g., by Earman 2008), is formalization of the alternative "Aristotelian" concept of indeterminism involving an open future (Placek and Belnap 2012).

23.13 Interpretations of quantum theory: introduction

Quantum theory deserves a more extended development of its basic concepts than I can provide here, where I only recount the basic ideas relevant for the issues in modality discussed in Sections 23.14 and 23.15 (for a brief introduction to the formalism of quantum mechanics, with many suggestions for further study, see Ismael 2015). Recall, then, that in elementary quantum mechanics, the state space of a physical system is represented by a vector in a Hilbert space, and

an observable for that system is represented by a Hermitian operator. If the system is not being measured, its time evolution is described by a smooth, one-parameter family of unitary operators—rotations in the Hilbert space, essentially—acting on its initial state. But if an observable is measured, the state projects, generally discontinuously and stochastically, onto one of the eigenstates of the observable whose eigenvalue represents the outcome of the measurement, with probability given by the Born rule. The inclusion of this "projection postulate" provides both an explicit mechanism for how experiments with quantum systems seem to yield definite outcomes, but also a conceptual puzzle: how are measurements supposed to be different from any other physical process? An interpretation of quantum theory, besides giving a description of what the world would be like if the theory were true, also must solve this so-called "measurement problem" (Myrvold 2018: §4). Some of these interpretations in particular involve explicit invocations of modality.

Interpretations of quantum mechanics in which modality plays an important role include some versions of the many-worlds interpretation and, unsurprisingly, the modal interpretation. Both of these interpretations propose to *eliminate* the projection postulate, so that measurement can be understood not as discontinuous changes of the state of the measured system, but as ordinary dynamical evolution of that system interacting with the measurement device. However, the way they do and their consequent connections with modality are different.

23.14 Many-worlds interpretations of quantum theory

The many-worlds interpretation attempts to explain how, without adding any extra formal ingredients, one can recover the use of probability, the semblance of definiteness to the outcomes of experiments, and, indeed, human experience itself. This is needed because if a system were to interact with a measurement device according to the standard (non-measurement) dynamics, their states would become entangled, in general consisting of a *sum* of terms in each of which the system is in some eigenstate of the measurement operator and the measurement device is in a state indicating as such. Apparently, the state has "split" into many worlds, each of which contains a measurement device with a definite outcome for the experiment.

What, then, is the status of these splitting worlds? Wilson (2013, 2015) has recently suggested that *each* of them is in fact a metaphysically possible world, not just some "part" of one metaphysically possible world in which the system and measurement device are entangled. This in turn has implications for the understanding of objective probability in the theory. See also Conroy (2018) for an actualist alternative and Skyrms (1976) for skepticism about relating the many-worlds interpretation to modal metaphysics.

23.15 Modal interpretations of quantum theory

In contrast with many-worlds interpretations, modal interpretations *add* structure to quantum theory after removing the projection postulate. In particular, they interpret the usual quantum state as purely modal—it describes merely what properties are *possible* for the system—and add a second state that represents what properties are *actual* for the system. This is a list of the actual properties of the system at a time, and so it is in general distinct from the dynamical quantum state. The structure of quantum theory prevents this list from being as comprehensive as it would be for systems described classically (de Ronde et al. 2014). There are many different proposals for what these actual properties should be, how they should be determined, and whether one's attitude should be empiricist (Dieks 2010) or realist (de Ronde 2010) towards the purely modal quantum state, but they all share the aforementioned basic structure (Lombardi and Dieks 2017).

Like with the adiabatic accessibility relation in axiomatic thermodynamics discussed in Section 23.8, one can interpret the standard quantum dynamics as providing a kind of accessibility relation between possible states, albeit one of a much more complicated structure, since the accessibility relations (i.e., the quantum states) are not determined by the actual state. It would be interesting to develop a sort of dynamic modal logic to capture this.[7]

Notes

1 This includes the somewhat subtle question of to what extent the laws themselves are necessary (Darrigol 2014).
2 Uffink (2001: §3) considers a third notion of possibility according to which "the claim that such a process [proscribed by the Second Law] is impossible, becomes a statement that transcends theoretical boundaries." But it isn't clear to me why this isn't just the notion of physical possibility haughtily immunized against disconfirmation. It is hard to see any basis for a claim that the Second Law is an analytic or a priori truth.
3 Uffink (2001) also claims that, despite this, their formulation of the Second Law is compatible with time-reversal invariance, although Henderson (2014) has more recently disputed this.
4 As states are not worlds in general but small fragments or parts of worlds, this perhaps brings the analogy closer to truthmaker semantics (Fine 2017).
5 For "simple" systems, Lieb and Yngvason (1999) also demand that they satisfy the "Comparability Hypothesis," namely that the adiabatic accessibility relation is total. For such systems, this would yield the modal logic **S4.3**. Cf. the comments of Earman (1986: 99–100) regarding the modal strength of physical necessity.
6 This analysis distinguishes being "physically reasonable," as a physical-modal attribute, from being "physically significant," as an epistemic-modal attribute (Fletcher 2016). Some of the exegetical puzzles for the physics literature on these issues may arise from conflating the two—cf. Earman (1995: 80).
7 There is already a modal interpretation of quantum logic (Dalla Chiara 1977; Dalla Chiara and Giuntini 2002), but the proposal here is for a *dynamic* logic that fits with the modified structure of modal interpretations of quantum theory.

References

Belnap, N. (1992) "Branching space-time," *Synthese* 92(3), 11–30.
Belnap, N. (2002) "EPR-like 'funny business' in the theory of branching space-times," in T. Placek and J. Butterfield (eds.), *Non-locality and Modality*, Dordrecht: Kluwer, pp. 293–316.
Belnap, N. (2012) "Newtonian determinism to branching space-times indeterminism in two moves," *Synthese* 188(1), 5–21.
Belot, G. (1995a) "Determinism and ontology," *International Studies in the Philosophy of Science* 9(1), 85–101.
Belot, G. (1995b) "New work for counterpart theorists: Determinism," *British Journal for the Philosophy of Science* 46(2), 185–195.
Belot, G. (2011) *Geometric Possibility*, Oxford: Oxford University Press.
Berenstain, N. and Ladyman, J. (2012) "Ontic structural realism and modality," in E. Landry and D. Rickles (eds.), *Structural Realism: Structure, Object, and Causality*, Dordrecht: Springer, pp. 149–168.
Brading, K. (2011) "Structuralist approaches to physics: Objects, models and modality," in P. Bokulich and A. Bokulich (eds.), *Scientific Structuralism*, Dordrecht: Springer, pp. 43–65.
Brading, K., Castellani, E. and Teh, N. (2017) "Symmetry and symmetry breaking," in E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, Winter 2017 edn, Stanford, CA: Stanford University, Metaphysics Research Lab.
Brighouse, C. (1994) "Spacetime and holes," in D. Hull, M. Forbes, and R. M. Burian (eds.), *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, Vol. 1, Philosophy of Science Association, Chicago: University of Chicago Press, pp. 117–125.
Brighouse, C. (1997) "Determinism and modality," *British Journal for the Philosophy of Science* 48(4), 465–481.
Brighouse, C. (1999) "Incongruent counterparts and modal relationism," *International Studies in the Philosophy of Science* 13(1), 53–68.

Brighouse, C. (2008) "Understanding indeterminism," in D. Dieks (ed.), *The Ontology of Spacetime II*, Amsterdam: Elsevier, pp. 153–173.

Burgess, J. P. (2009) *Philosophical Logic*, Princeton, NJ: Princeton University Press.

Butterfield, J. (1988) "Albert Einstein meets David Lewis," in A. Fine and J. Leplin (eds.), *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, Vol. 2, Philosophy of Science Association, Chicago: University of Chicago Press, pp. 65–81.

Butterfield, J. (1989) "The hole truth," *British Journal for the Philosophy of Science* 40(1), 1–28.

Butterfield, J. (2003) "David Lewis meets Hamilton and Jacobi," *Philosophy of Science* 71(5), 1095–1106.

Butterfield, J. (2004a) "Between laws and models: Some philosophical morals of Lagrangian mechanics." Unpublished manuscript. URL: http://philsci-archive.pitt.edu/1937/

Butterfield, J. (2004b) "Some aspects of modality in analytical mechanics," in M. Stöltzner and P. Weingartner (eds.), *Formale Teleologie und Kausalität in der Physik*, Paderborn: Mentis, pp. 160–198.

Carroll, J. W. (2016) "Laws of nature," in E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, Fall 2016 edn, Stanford, CA: Stanford University, Metaphysics Research Lab.

Conroy, C. (2018) "Everettian actualism," *Studies in History and Philosophy of Modern Physics* 63, 24–33.

Dalla Chiara, M. L. (1977) "Quantum logic and physical modalities," *Journal of Philosophical Logic* 6(1), 391–404.

Dalla Chiara, M. L. and Giuntini, R. (2002) "Quantum logics," in D. M. Gabbay and F. Guenthner (eds.), *Handbook of Philosophical Logic*, Vol. 6, 2nd edn., Dordrecht: Springer, pp. 129–228.

Darrigol, O. (2014) *Physics and Necessity: Rationalist Pursuits from the Cartesian Past to the Quantum Present*, New York: Oxford University Press.

Dasgupta, S. (2015) "Substantivalism vs relationalism about space in classical physics," *Philosophy Compass* 10(9), 601–624.

Dieks, D. (2010) "Quantum mechanics, chance and modality," *Philosophica* 83(1), 117–137.

Earman, J. (1986) *A Primer on Determinism*, Dordrecht: D. Reidel.

Earman, J. (1989) *World Enough and Space-Time: Absolute versus Relational Theories of Space and Time*, Cambridge, MA: MIT Press.

Earman, J. (1995) *Bangs, Crunches, Whimpers, and Shrieks: Singularities and Acausalities in Relativistic Spacetimes*, Oxford: Oxford University Press.

Earman, J. (2007) "Aspects of determinism in modern physics," in J. Butterfield and J. Earman (eds.), *Philosophy of Physics*, Vol. B, Amsterdam: Elsevier, pp. 1369–1434.

Earman, J. (2008) "Pruning some branches from branching spacetimes," in D. Dieks (ed.), *The Ontology of Spacetime II*, Amsterdam: Elsevier, pp. 187–206.

Earman, J. and Norton, J. (1987) "What price spacetime substantivalism? The hole story," *British Journal for the Philosophy of Science* 38(4), 515–525.

Earman, J. and Norton, J. D. (1998) "Exorcist XIV: The wrath of Maxwell's demon. Part I. From Maxwell to Szilard," *Studies in History and Philosophy of Modern Physics* 29(4), 435–471.

Earman, J. and Norton, J. D. (1999) "Exorcist XIV: The wrath of Maxwell's demon. Part II. From Szilard to Landauer," *Studies in History and Philosophy of Modern Physics* 30(1), 1–40.

Esfeld, M. (2009) "The modal nature of structures in ontic structural realism," *International Studies in the Philosophy of Science* 23(2), 179–194.

Esfeld, M. (2018) "Metaphysics of science as naturalized metaphysics," in A. Barberousse, D. Bonnay and M. Cozic (eds.), *The Philosophy of Science: A Companion*, Oxford: Oxford University Press, pp. 142–170.

Esfeld, M. and Lam, V. (2011) "Ontic structural realism as a metaphysics of objects," in P. Bokulich and A. Bokulich (eds), *Scientific Structuralism*, Dordrecht: Springer, pp. 143–160.

Field, H. (1984) "Can we dispense with space-time?" in *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, Vol. 2, Philosophy of Science Association, East Lansing, MI: Philosophy of Science Association, pp. 33–90.

Fine, K. (2017) "Truthmaker semantics," in B. Hale, C. Wright and A. Miller (eds.), *A Companion to the Philosophy of Language*, 2nd edn, West Sussex: Wiley Blackwell, pp. 556–577.

Fletcher, S. C. (2016) "Similarity, topology, and physical significance in relativity theory," *British Journal for the Philosophy of Science* 67(2), 365–389.

Fletcher, S. C. (2019) "Counterfactual reasoning within physical theories," *Synthese*, forthcoming. doi:10.1007/s11229-019-02085-0.

Fletcher, S. C. (2020) "On representational capacities, with an application to general relativity," *Foundations of Physics* 50(1), 228–249.

van Fraassen, B. C. (1980) *The Scientific Image*, Oxford: Clarendon Press.

French, S. (2017) "Building bridges with the right tools: Modality and the standard model," in M. Massimi, J.-W. Romeijn and G. Schurz (eds), *EPSA15 Selected Papers*, Cham: Springer, pp. 37–47.
French, S. and Ladyman, J. (2003) "Remodelling structural realism: quantum physics and the metaphysics of structure," *Synthese* 136(1), 31–56.
French, S. and McKenzie, K. (2012) "Thinking outside the toolbox: Towards a more productive engagement between metaphysics and philosophy of physics," *European Journal of Analytic Philosophy* 8(1), 42–59.
Friedman, M. (1983) *Foundations of Space-Time Theories*, Princeton, NJ: Princeton University Press.
Hawking, S. (1969) "The existence of cosmic time functions," *Proceedings of the Royal Society A* 308, 433–435.
Hawking, S. (1992) "The chronology protection conjecture," *Physical Review D* 46(2), 603– 611.
Henderson, L. (2014) "Can the second law be compatible with time reversal invariant dynamics?" *Studies in History and Philosophy of Modern Physics* 47, 90–98.
Hoefer, C. (2016) "Causal determinism," in E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, Spring 2016 edn, Stanford, CA: Stanford University, Metaphysics Research Lab.
Huggett, N. (2006) "The regularity account of relational spacetime," *Mind* 115(457), 41–73.
Ismael, J. (2015) "Quantum mechanics," in E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, Spring 2015 edn, Stanford, CA: Stanford University, Metaphysics Research Lab.
Karaca, K. (2017) "A case study in experimental exploration: Exploratory data selection at the Large Hadron Collider," *Synthese* 194(2), 333–354.
Karaca, K. (2018) "Lessons from the Large Hadron Collider for model-based experimentation: The concept of a model of data acquisition and the scope of the hierarchy of models," *Synthese* 195(12), 5431–5452.
Kowalski, T. and Placek, T. (1999) "Outcomes in branching space-time and GHZ-Bell theorems," *British Journal for the Philosophy of Science* 50(3), 349–375.
Ladyman, J. (1998) "What is structural realism?" *Studies in History and Philosophy of Science* 29(3), 409–424.
Ladyman, J. (2000) "What's really wrong with constructive empiricism? van Fraassen and the metaphysics of modality," *British Journal for the Philosophy of Science* 51(4), 837–856.
Ladyman, J. (2004) "Constructive empiricism and modal metaphysics: A reply to Monton and van Fraassen," *British Journal for the Philosophy of Science* 55(4), 755–765.
Ladyman, J. (2018) "Scientific realism again," *Spontaneous Generations* 9(1), 99–107.
Lange, M. (2007) "Laws and meta-laws of nature: Conservation laws and symmetries," *Studies in History and Philosophy of Modern Physics* 38(3), 457–481.
Leeds, S. (1995) "Holes and determinism: Another look," *Philosophy of Science* 62(3), 425–437.
Lieb, E. H. and Yngvason, J. (1999) "The physics and mathematics of the second law of thermodynamics," *Physics Reports* 310, 1–96.
Livanios, V. (2018) "Dispositionality and symmetry structures," *Metaphysica* 19(2), 201–217.
Lombardi, O. and Dieks, D. (2017) "Modal interpretations of quantum mechanics," in E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, Spring 2017 edn, Stanford, CA: Stanford University, Metaphysics Research Lab.
Malament, D. B. (2012) *Topics in the Foundations of General Relativity and Newtonian Gravitation Theory*, Chicago: University of Chicago Press.
Manchak, J. B. (2011) "What is a physically reasonable space-time?" *Philosophy of Science* 78(3), 410–420.
Manchak, J. B. (2013) "Global spacetime structure," in R. Batterman (ed.), *The Oxford Handbook of Philosophy of Physics*, Oxford: Oxford University Press, pp. 587–606.
Manchak, J. B. (2016a) "Epistemic "holes" in space-time," *Philosophy of Science* 83(2), 265– 276.
Manchak, J. B. (2016b) "Is the universe as large as it can be?" *Erkenntnis* 81(6), 1341–1344.
Manchak, J. B. (2017) "On the inextendibility of space-time," *Philosophy of Science* 84(5), 1215–1225.
Manders, K. L. (1982) "On the space-time ontology of physical theories," *Philosophy of Science* 49(4), 575–590.
Maroney, O. (2009) "Information processing and thermodynamic entropy," in E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, Fall 2009 edn, Stanford, CA: Stanford University, Metaphysics Research Lab.
Melia, J. (1999) "Holes, haecceitism and two conceptions of determinism," *British Journal for the Philosophy of Science* 50(4), 639–664.
Monton, B. and van Fraassen, B. C. (2003) "Constructive empiricism and modal nominalism," *British Journal for the Philosophy of Science* 54, 405–422.
Muller, F. A. (2005) "The deep black sea: Observability and modality afloat," *British Journal for the Philosophy of Science* 56(1), 61–99.

Müller, T. (2002) "Branching space-time, modal logic and the counterfactual conditional," in T. Placek and J. Butterfield (eds), *Non-locality and Modality*, Dordrecht: Kluwer, pp. 273–292.
Müller, T. (2013) "A generalized manifold topology for branching space-times," *Philosophy of Science* 80(5), 1089–1100.
Müller, T. and Placek, T. (2018) "Defining determinism," *British Journal for the Philosophy of Science* 69(1), 215–252.
Muntean, I. (2015) "Metaphysics from string theory: S-dualities, fundamentality, modality and pluralism," in T. Bigaj and C. Wuthrich (eds.), *Metaphysics in Contemporary Physics*, Leiden: Brill/Rodopi, pp. 259–292.
Myrvold, W. (2018) "Philosophical issues in quantum theory," in E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, Fall 2018 edn, Stanford, CA: Stanford University, Metaphysics Research Lab.
Norton, J. D. (2005) "Eaters of the lotus: Landauer's principle and the return of Maxwell's demon," *Studies in History and Philosophy of Modern Physics* 36, 375–411.
Norton, J. D. (2018) "The hole argument," in E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, Summer 2018 edn, Stanford, CA: Stanford University, Metaphysics Research Lab.
Perovic, S. (2011) "Missing experimental challenges to the standard model of particle physics," *Studies in History and Philosophy of Modern Physics* 42(1), 32–42.
Placek, T. (2000) "Stochastic outcomes in branching space-time: Analysis of Bell's theorem," *British Journal for the Philosophy of Science* 51(3), 445–475.
Placek, T. (2002) "Partial indeterminism is enough: a branching analysis of Bell-type inequalities," in T. Placek and J. Butterfield (eds.), *Non-locality and Modality*, Dordrecht: Kluwer, pp. 317–342.
Placek, T. (2010) "A locus for 'now'," in D. Dieks, W. Gonzalo, T. Uebel, S. Hartmann and M. Weber (eds.), *Explanation, Prediction, and Confirmation*, Dordrecht: Springer, pp. 395–410.
Placek, T. (2014) "Branching for general relativists," in T. Müller (ed.), *Nuel Belnap on Indeterminism and Free Action*, Berlin: Springer, pp. 191–221.
Placek, T. (2019) "Laplace's demon tries on Aristotle's cloak: On two approaches to determinism," *Synthese* 196(1), 11–30.
Placek, T. and Belnap, N. (2012) "Indeterminism is a modal notion: Branching space-times and Earman's pruning," *Synthese* 187(2), 441–469.
Placek, T., Belnap, N. and Kishida, K. (2014) "On topological issues of indeterminism," *Erkenntnis* 79(S3), 1–34.
Rojo, A. and Bloch, A. (2018) *The Principle of Least Action: History and Physics*, Cambridge: Cambridge University Press.
de Ronde, C. (2010) "For and against metaphysics in the modal interpretation of quantum mechanics," *Philosophica* 83(1), 85–117.
de Ronde, C., Freytes, H. and Domenech, G. (2014) "Interpreting the modal Kochen-Specker theorem: Possibility and many worlds in quantum mechanics," *Studies in History and Philosophy of Modern Physics* 45, 11–18.
Sklar, L. (1974) "Incongruous counterparts, intrinsic features and the substantiviality of space," *The Journal of Philosophy* 71(9), 277–290.
Skyrms, B. (1976) "Possible worlds, physics and metaphysics," *Philosophical Studies* 30(2), 323–332.
Smeenk, C. and Hoefer, C. (2015) "Philosophy of the physical sciences," in P. Humphreys (ed.), *The Oxford Handbook of Philosophy of Science*, Oxford: Oxford University Press, pp. 115–136.
Suppes, P. (1966) "Models of data," in E. Nagel, P. Suppes and A. Tarski (eds), *Logic, Methodology and Philosophy of Science*, Amsterdam: Elsevier, pp. 252–261.
Terekhovich, V. (2018) "Metaphysics of the principle of least action," *Studies in History and Philosophy of Modern Physics* 62, 189–201.
Uffink, J. (2001) "Bluff your way in the second law of thermodynamics," *Studies in History and Philosophy of Modern Physics* 32(3), 305–394.
Uffink, J. (2007) "Compendium of the foundations of classical statistical physics," in J. Butterfield and J. Earman (eds.), *Philosophy of Physics*, Vol. B, Amsterdam: Elsevier, pp. 923–1074.
Wald, R. M. (1984) *General Relativity*, Chicago: University of Chicago Press.
Wilson, A. (2013) "Objective probability in Everettian quantum mechanics," *British Journal for the Philosophy of Science* 64(4), 709–737.
Wilson, A. (2015) "The quantum doomsday argument," *British Journal for the Philosophy of Science* 68(2), 597–615.
Wroński, L. and Placek, T. (2009) "On Minkowskian branching structures," *Studies in History and Philosophy of Modern Physics* 40(3), 251–258.
Yourgrau, W. and Mandelstam, S. (1960) *Variational Principles in Dynamics and Quantum Theory*, 2nd edn, New York: Pitman.

Chapter 24

PHYSICAL AND METAPHYSICAL MODALITY

Ned Hall

24.1 Introduction

Philosophers widely if not universally agree that physical and metaphysical modality are to be distinguished. Physical modality is the kind of modality we have in mind when, say, we discuss free will and determinism, or when we pick out those truths about the world that are, or are guaranteed by, laws of nature. By contrast, *metaphysical* modality is supposed to be strictly broader, as witness what seems to many to be the obvious truth that the laws of nature – and so, the merely *physical* necessities – *could have been otherwise*, in an appropriately "metaphysical" sense of "could have been".

Against this consensus, this essay sketches a heretical account of metaphysical possibility and necessity according to which the "and" of the title may be replaced by an "is". Or better, as we will see: by a "just is".

There are obvious caveats. I will not have space to discuss the modal status of truths of mathematics, nor how the conception of metaphysical modality I will sketch interacts with such currently fashionable metaphysical categories as "ground", "essence", "real definition", and so on. There will be some choice points where it is not clear, to me at least, which is the best path for further development of the sketch. And the argument in favor of the account is in a certain sense quite modest; at least, it is very far from a clever deduction of the equivalence of metaphysical and physical modality from premises no one would have suspected could yield such a striking conclusion.[1] If you end up agreeing with the account, it will be because, having seen what it comes to, you concur with me that there doesn't look to be anything else that "metaphysical modality" could reasonably name – more exactly, that the only *point* there could be to having a concept of *metaphysical* modality will be satisfied if we simply equate this with physical modality. And so in another sense, the argument is quite *im*modest: for if I am right, then the persistent habit of distinguishing physical from metaphysical modality simply reflects a confusion about the function of these modal notions.

I'll lead up to the central idea with a little autobiography.

24.1.1 Some personal history

In my youth, I was taught to take seriously a familiar form of argument (FFARG):

1. It is conceivable that P.
   Therefore,
2. It is metaphysically possible that P.
   But
3. It is not metaphysically contingent whether P.
   Therefore,
4. It is true (and indeed metaphysically necessary) that P.

Of course, it was understood that one must use the FFARG with good taste and discretion. Only for certain choices of P would premise 3 be plausible – though the hope was always that for many philosophically *interesting* choices of P, 3 would hold.[2] And the transition from 1 to 2 needed to be handled with care, lest we contradict ourselves about undecidable mathematical claims, or implausibly deny that Hesperus is Phosphorus, or contravene established Kripke-Putnamian Doctrine concerning natural kinds.[3] All the same, one could quite reasonably be confident that the FFARG could do serious philosophical work – e.g., in establishing or at least strongly supporting mind/body dualism,[4] or in refuting or at least seriously undermining Humeanism about laws,[5] or perdurantism about persistence.[6] So, at any rate, I was led to believe.

And then – long post-youth, alas – it all began to seem too good to be true. Okay, that's being overly polite: it began to seem silly, ridiculous, a mere exercise in philosophical wishful thinking to suppose that our human powers of conceiving could, if but properly deployed, yield knowledge of substantive, profound conclusions about the nature of reality. What's more, disputes about particular uses of the FFARG too often seemed to devolve into fruitless squabbles over conceivability: "You claim that you can conceive of a situation in which P. But why should I trust you? *I* can't conceive of such a situation". And then there was the occasional, and regrettable, failure of philosophical nerve – e.g., Lewis's (1994) concession that it is only *contingently* true that objects persist by perduring (because of those pesky spinning spheres), or the more widespread acceptance by materialists about the mind that their view is only contingently true (because of those even peskier zombies). It started to look more and more attractive simply to reject the FFARG, root and branch.

Now, maybe that's hasty; maybe, for example, Descartes was right, and we have been so endowed by our Creator with a capacity for clear and distinct ideas that just by thinking hard and carefully, we can figure out, say, physics. But I doubt it, and at any rate no one ever suggested that that is why the FFARG is to be trusted. Come to think of it, no one ever mentioned *any* particularly compelling reason why the FFARG is to be trusted. Reject away, I say!

Or: continue to trust in the FFARG, but take pains to fashion your concept of metaphysical modality so as not to make such trust look badly misplaced. The most obvious way to do so is to *epistemologize* your conception of modality, perhaps along the lines of Rosen's (2006) "non-standard" conception, according to which (roughly) a world counts as "metaphysically possible" just in case its full description contains no latent contradictions that would be detected by a suitably informed and intelligent agent. On an approach of this sort, we will end up meaning by "metaphysically possible" nothing more than "correctly conceivable" – where there will be plenty of room for disagreement about what "correctly" means, but where this qualifier will permit a close enough connection to what *we* can conceive that we can go on treating our powers of imagination as a reasonably reliable guide to what is possible.

On such an approach, it will turn out to be possible, say, that objects persist through time by enduring – and likewise possible that they persist by perduring. It will turn out to be possible that laws of nature are mere patterns in the phenomena – and likewise possible that they govern such patterns while being wholly distinct from them. It will turn out to be possible that only the present exists, that only the past and present exist, and that the past, present, and future all equally exist. And so on. So we could continue to use the FFARG. But what would be the point? The very conception of modality that tries to vindicate it also renders it toothless.

And there is a further complaint. Having so directly tied the concept of possibility to what is (ideally, perhaps) conceivable, why should we persist in adding the label "metaphysical"? That label suggests a species of possibility and necessity that has something to do with *what reality is fundamentally like*. But (as Rosen himself very forthrightly notes) that connection has been jettisoned, on this "non-standard" conception. Much preferring terminology that does not mislead, I would rather reserve the name "metaphysical modality" for a kind of modality that *preserves* this connection.

Suppose you agree. Still, that leaves some questions. Just what is metaphysical modality, then? Is the notion useless, as Quine and others have thought? Muddling over those questions led me (as I think it has led others; see for example Maudlin 2007) to the nascent view I'll try to lay out here, and whose central idea comes next.

24.1.2 The central idea

I think of metaphysical modality as intrinsically connected to what we seek in explanation.

That's the core thought. I'll expand on it a bit here, and then spend the rest of the chapter further unpacking, and trying to figure out what it might come to.

So let's start with a truism about inquiry: In investigating our world, and trying to figure out what it's like, one of the central things we're after is *understanding*. That is, we don't merely aim to collect facts about what our world is like; we want, additionally and crucially, *explanations* for those facts (where we can have them, anyway!).

Now, more controversially: Part of what we seek, in seeking explanations, is knowledge of that upon which the target of our inquiry *depends*. Often this dependence is *causal*: we want to know what *brought about* the thing we're trying to explain. Often, but not invariably (as we'll see); and so I'll speak more generically of *relations of explanatory dependence*. Essential to any explanation is the exhibiting of such relations. These relations are *objective*, and *out there in the world*, and not merely grounded in how we happen to *represent* the world to ourselves. And these relations are not, finally, some miscellaneous and disparate jumble. No, they all have a common *source*; they are all grounded in *constraints* that certain aspects of the world's fundamental structure impose on what is the case.[7]

And "metaphysical necessity" is just a label for these fundamental explanatory constraints.

That they are *constraints* makes the word "necessity" appropriate. That they are rooted in *reality itself* makes the word "metaphysical" (as opposed to, say, "logical", or "conceptual") appropriate.

Here is another way to put the idea. The simplest – and, for my purposes, most revealing or canonical (which is *not* to say best!) – form an explanation can take is the following:

> "Why is it the case that P?"
>
> Answer: "Because, thanks to such-and-such fundamental feature of reality, it *must* be the case that P".

So the central idea, simplified slightly, is that metaphysical necessity *just is* the kind of necessity that figures in such explanations.

Now let's come at this idea from a slightly different direction, after a brief pause to dispense, once and for all, with the FFARG.

24.1.3 Death to the FFARG

What goes wrong with the FFARG, on my view? This is easy: it involves an equivocation. If you mean what *I* mean by "metaphysically possible", then the move from conceivability to possibility is absurd. If, on the other hand, you mean something by "metaphysically possible" that secures the move from 1 to 2 – for example, "correctly conceivable", a la Rosen (2006) – then, as noted earlier, no interesting instance of 3 will be true. For example, it will turn out to be "metaphysically" possible that presentism is true. And likewise possible that endurantism is true. And possible that every set of objects has a fusion. And possible that no set of objects has a fusion. And possible that objects persist by being composed of time slices. And possible that they persist by being wholly present at each moment of their existence. And possible that the continuum hypothesis is true. And possible that it's false. You get the idea.

Now, the argument here is hardly airtight; in particular, the fourth sentence in the last paragraph looks plausible *to me* simply because I can't see what metaphysical possibility could be, such that conceivability is a good guide to it, *and* such that many interesting instances of 3 will be true. But maybe that's just me. At any rate, I'm going to proceed on the following optimistic assumption of genuine personal growth: namely, that in rejecting FFARG, I've succeeded in throwing off the shackles of my youthful indoctrination, and not, say, just become too addled with age to any longer appreciate the sacred mysteries into which my elders had initiated me.

24.2 A background conception of metaphysics

The conception of modality that I'm recommending flows naturally from a certain conception of metaphysics itself.

Start with our truism: We try to *understand* stuff (scientific inquiry being our most disciplined and effective means for doing so). Now add another truism: We *posit* other stuff, in order to arrive at the explanations we seek of the stuff we're trying to understand. In so doing, we add to the stuff we're trying to understand, whence the opportunity arises for further posits, etc. On one conception, metaphysics just *is* the study, at the most abstract and general level possible, of these posits. In short: metaphysics is the *general theory of explanatory structure*.

On this conception, metaphysics is, *of course*, smoothly continuous with science (though this is not to say that scientists are any good at it). And on this conception, it's no surprise to find these among the most central questions in metaphysics: What are laws of nature? What is causation? What is objective chance? What makes time different from space? And so on.

It's also not surprising, given this conception, that very many comprehensive metaphysical views follow a kind of common pattern, in that they involve the construction of what I call a "cosmic SCARF".

24.2.1 Cosmic SCARFs

Suppose you're a comprehensive metaphysician. If you follow the pattern I have in mind, your SCARF will end up with the following threads in it:

- An account of the *S*tuff that reality contains.
- An account of the *C*onstitution relations connecting its bits.

- An account of the *A*rena in which its behavior unfolds.
- An account of the *R*ules governing this behavior.
- An account of the *F*eatures of the stuff that give its nature.

A couple of examples will help clarify what cosmic SCARFs can look like. To set the stage, let's suppose that classical particle mechanics (CPM) had turned out to be exactly right. Still, metaphysicians have gainful employment: for the physics – as presented by physicists – underdetermines the nature of its basic explanatory posits. Here are two strikingly different conceptions of these posits.

24.2.2 A sample SCARF

Our first CPM-friendly SCARF is inspired in part by Newton and in part by Tim Maudlin (2007). It takes the stuff to be enduring (in the technical sense) point-particles. The constitution relations are in part mereological (at any time, and for any set of particles, there is at least one thing composed at that time of just those particles), and in part involve some complicated story about the constitution and identity across time of macro-objects. The arena is Newtonian absolute space and time. The rules are the basic Newtonian laws – understood, à la Maudlin, as "FLOTEs", or fundamental laws of temporal evolution – together with fundamental laws restricting the space of possible instantaneous states (e.g., it might be a law that there are only finitely many particles). These basic laws are constituents of reality *distinct from*, but *governing*, the behavior of the particles. And the features, finally, are just mass and charge.

24.2.3 Another sample SCARF

Here's a rival CPM-friendly SCARF, this time inspired by Lewis. This time, the stuff is just spacetime points – or really, for reasons we're about to see, just *mereological simples*. The constitution relations are purely mereological: for any simples, there is a unique thing that is their fusion. Talk of the "arena" is redundant, as it is nothing more than the fusion of *all* the simples. There are no rules. (There are *patterns* in the phenomena, to be sure; and some of these, for reasons that are, ultimately, epistemological, deserve special attention. But to say *that* is not to add to the SCARF in any way. See Hall (2017, 2020). And the fundamental features are mass, charge, and spatiotemporal relations (Lewis's "perfectly natural properties"); whence there's no need to characterize the stuff as *spacetime* points, since all it comes to for them to deserve that label is that they are simples that instantiate spatiotemporal relations.

24.2.4 A missing ingredient?

It might seem – it might even seem *obvious* – that a metaphysician who has constructed a cosmic SCARF has more work to do. For why isn't it incumbent on her to explain, in addition, the *modal status* of the different elements of her SCARF? (E.g., are the rules metaphysically contingent or necessary? Etc.)

Switching, briefly, back to autobiography: When the "SCARF" framework first occurred to me, I found myself almost *automatically* worrying about this further question; given the psychological residue of all that youthful indoctrination, it seemed almost unavoidable that the question of the SCARF's modal status would arise. (Which was sad, as adding an "M" promised to ruin the acronym.)

But then it started to look avoidable – especially if we take seriously the conception of metaphysics as the general theory of explanatory structure. For in that case, why *bother* with a conception of "metaphysical" possibility and necessity that goes beyond what's needed, to articulate the relevant standards of explanation? Explanations, so the thought goes, succeed by articulating relations of dependence between explananda and explanantia. These relations of explanatory dependence all have their ultimate source in constraints imposed on what is the case by the most basic elements of reality – the ingredients of the cosmic SCARF. To say that something is *necessary* – in a metaphysical sense – is to say that it is *constrained to be that way* by the SCARF. So while it's perfectly sensible to ask about the *logical* or *conceptual* modal status of the SCARF,[8] it does not make sense to ask about its *metaphysically* modal status. The SCARF is the *source* of explanatory constraint, and for that reason is not itself constrained to be any way at all.

24.3 How one might be led toward this view

By this point, the reasoning is straightforward:

First, to explain something is, in part, to provide information about *what it depends on*. Second, these relations of explanatory dependence are all grounded in *constraints imposed by the nature and structure of the fundamental ingredients of the SCARF*. Finally – and crucially – there is no good reason to use "metaphysical necessity" as a label for anything *other* than these constraints. So "it is metaphysically necessary that P" comes to mean "the nature and structure of the fundamental SCARF ingredients constrains reality to be such that P". That doesn't *quite* get us all the way to an equation of metaphysical with physical modality; that will depend on whether we think that the "physical" necessities are a proper subset of the SCARF-imposed constraints. But it gets us close enough.

That's it. Well, almost. There's one bit of bookkeeping fussiness to attend to. If it's really important to you to be able to make sense of iterated modalities, or really important to you that metaphysical possibility be understood as the dual of metaphysical necessity, then you won't like the foregoing formula. No matter. There's a simple fix[9]. Count a world as "metaphysically possible" exactly if (i) its fundamental SCARF-ingredients are exactly as they actually are; (ii) in all other respects it conforms to the constraints imposed by these ingredients. Define "it is metaphysically necessary/possible that P" in the usual way (so that each is the dual of the other). Then it will turn out that it is metaphysically necessary that the fundamental SCARF ingredients are the way they are. But don't be distracted by what is, after all, *just* a bookkeeping maneuver: this is 'metaphysical necessity' by courtesy only; it remains just as confused as before to ask what "constrains" the SCARF ingredients to be this way.

24.4 Explanatory dependence and its structure

It will be a good idea to have some examples of explanations, to help make vivid the variety of explanatory dependence relations they can appeal to, and to help make clear how these dependence relations might be grounded in fundamental SCARF ingredients. We'll look at six examples, and then make some comments about the structure of metaphysical necessity, using classical particle mechanics once again as a foil.

24.4.1 First example: particle quintumvirates

Call a "quintumvirate" a collection of five particles, each of which is at exactly the same non-zero distance from every other. Suppose ours is a classical particle world, with a three-dimensional,

Euclidean space. There are not, have never been, and never will be any quintumvirates. Why not? Because the structure of space forbids it. Particles are constrained to occupy locations that the structure of space makes available; and there simply are no locations that permit the formation of a quintumvirate. This is a cleanly illustrative example: We explain a pervasive regularity by direct appeal to the constraints imposed on what is the case by one of the fundamental ingredients of the SCARF.

It's helpful to compare an alternative scenario: This time, space is four-dimensional and Euclidean, but the particles all start out in the same three-dimensional subspace. As before, there are not, have never been, and never will be any quintumvirates. Why not? The answer is more complicated, but goes roughly as follows: No particle will ever experience a force that will pull it *out* of the three-dimensional subspace in which all the particles start out, for such a force could only come from another particle, and all forces between two particles act along the line connecting them. So, given the way the force laws work, and given the laws of motion, and given the structure of space, every particle will forever be constrained to move in such a way as to stay within this subspace. So this time, the explanation is mixed, adverting in part to constraints imposed by the fundamental ingredients of the SCARF (as, on the view we're considering, *all* explanations must, at least indirectly), but also in part to a brute contingency (as, on this view, explanations may but don't have to).

24.4.2 Second example: particles in a box

Suppose that, in our classical particle world, there are some particles in a box at time t1. The walls of the box are impermeable, but there is an opening in it. By t2, all the particles are outside of the box, having passed through this opening. Consider the cube-shaped region of space R that exactly bounds the box. We can ask: Given that the particles were inside the box at t1 and outside of it at t2, why did they cross region R in the interim? The best answer – the one pitched at just the right level of explanatory generality – is that they were *constrained* to do so, in part by the structure of space, and in part by the laws of motion: for given the laws of motion, particles cannot move discontinuously; and given the structure of space, the only continuous paths available to the particles cross R. Here we see something that I think is quite common: constraints imposed by distinct fundamental aspects of the SCARF working together to yield some result. Note that we don't achieve the right sort of explanatory "depth" (cf. Strevens 2008) if we advert in our explanation to the particular forces that led the particles on their particular trajectories through the opening (or indeed, if we mention the opening at all). Given the particular why-question asked, the right level of depth is achieved only by adverting directly to the SCARF-imposed constraints on motion.

24.4.3 Third example: collapsing stars

The third example comes from Railton by way of Lewis (1986). Lewis writes:

> A star has been collapsing, but the collapse stops. Why? Because it's gone as far as it can go. Any more collapsed state would violate the Pauli Exclusion Principle. It's not that anything caused it to stop – there was no countervailing pressure, or anything like that. There was nothing to keep it out of a more collapsed state. Rather, there just was no such state for it to get into. The state-space of physical possibilities gave out.

Lewis says exactly the right thing about this example.[10] What explains the given phenomenon is a constraint imposed by the fundamental laws of nature (the laws of coexistence, specifically). Lewis focuses on this example, by the way, because he sees it as a threat to his thesis that an explanation of an event consists in information about its causal history. And it *is* a threat – a fully successful one. But Lewis wants to avoid this concession, and so offers an alternative gloss on the example in the next paragraph:

> I reply that information about the causal history of the stopping has indeed been provided, but it was information of an unexpectedly negative sort. It was the information that the stopping had no causes at all, except for all the causes of the collapse which was a precondition of the stopping. Negative information is still information. If you request information about arctic penguins, the best information I can give you is that there aren't any.

But that *can't* be right: just compare a case where the space of physical possibilities does not give out – but where the FLOTEs are so radically non-deterministic that they allow that events (e.g., the stopping) can happen uncaused. No, the example is, as Railton intended, a clear case of non-causal explanation of an event.

As an aside, it was a mistake for Lewis to worry about *this* sort of counterexample. For a natural and perfectly friendly amendment to his thesis about event-explanation is that the explanation of any event consists in information about what the event's occurrence *depends* on. Granted, in the vast majority of cases, the dependence in question will be *causal*. But so what, if in some cases it's a different variety?

24.4.4 Fourth example: time-traveling billiard balls

Before getting to the science fiction, consider an ordinary case of probabilistic explanation: You have a lump of radium. After 1601 years, very close to half of it has decayed into radon gas. Why? Answer: Because of the probabilistic FLOTE that says that radium has a half-life of 1601 years.

The probabilistic FLOTEs are doing real explanatory work in examples like this, that cannot be recovered if one gives a frequentist reduction of them (however sophisticated). Here's an example, designed to show why.

The world of the example is mostly boring: it contains billiard balls, floating around in empty space, occasionally undergoing perfectly elastic collision (but otherwise not interacting at all). But there's one interesting detail. In a certain corner of space, there is a time-travel entrance portal, with the exit portal nearby, facing it.

At a certain time, a billiard ball is about to sail right between the entrance and exit portals. What will happen? Well, two things *can* happen, given the laws and the spacetime structure and the exact positioning of the ball. The ball could sail through in a straight line. Or, as it approaches the line connecting the two portals, it could happen that a ball sails out of the exit portal, and collides with our ball so as to send it into the entrance portal. Whence, of course, it is the very same ball, colliding with itself.

Notice that there is nothing whatsoever in the fundamental structure of this world that could generate an *objective chance* for either outcome. And so if the ball undergoes self-collision, and we ask "Why?", there is *no answer to give* beyond "Well, that's one of the two things it *could* do".

This sad verdict won't change if we add more billiard balls. So suppose a gazillion goes sailing through this region, and that exactly 30% of them – randomly distributed – undergo self-collision. What could explain this statistical regularity? Nothing whatsoever. The contrast with the

radium example is stark: for in that case, we *can* see the statistical behavior of the radium atoms as being constrained – not perfectly, but highly – by the probabilistic FLOTEs. Not so, with our time-traveling billiard balls.

Notice that we have here yet another reason to doubt frequentism about objective chance. For a frequentist reduction of probabilistic laws will say that there are objective probabilities here, just as much as in the case of our lump of radium. It will see *no difference* between the cases, when there palpably is one.

24.4.5 Fifth example: different masses

Different particles frequently exhibit different values for mass. But it's an unbroken regularity that no particle by itself ever exhibits, at the same time, two distinct values for mass (or any other basic physical magnitude). Why not?

Yes, it's an odd question. But perhaps only because the answer is so obvious: no particle *can* have two mass-values at one time. Given the nature of mass, it is *metaphysically impossible* for a particle to exhibit distinct mass values. This answer – which I think is perfectly correct – might appear to generate a puzzle: for what could *account* for this metaphysical impossibility? But I would prefer a bit of theft over honest toil, here, and say instead merely that mass is (we think) a fundamental element of the SCARF, and that it has a certain *structure*, and that this structure *imposes a constraint* on anything that possesses mass: namely, that it possess exactly *one value* for mass. *That*, and no more, is what this metaphysical impossibility comes to.

24.4.6 Sixth example: causation

A window has broken. Why? Because Suzy, ever accurate, has thrown a rock at it.

This is the most familiar (and common) case: we explain something, by citing information about its causes. So how do metaphysical necessities (understood in the way I've suggested) get into the picture? Via two steps (presented here in a highly compressed fashion; though see Hall 2011 for a fuller treatment).

First step: analyze causal structure in terms of localized counterfactual dependence structure.

Second step: analyze the relevant counterfactuals in terms of facts about what happens, together with facts about what is nomologically (i.e., metaphysically!) possible.

Notice that causal explanations will quite liberally mix together information about constraints imposed by the fundamental SCARF ingredients (information presented very obliquely) with other sorts of information. That's fine, but awkward for my purposes: for it can make it difficult to appreciate the distinctive explanatory force of these constraints. Better to focus, say, on the lack of particle quintumvirates and other examples of explanation-by-constraint, neat.

24.4.7 The heterogeneous character of metaphysical necessity

If all of this is right, then metaphysical necessity has an ungainly number of dimensions:

There are constraints imposed by the structure of space and time.
There are constraints imposed by the FLOTEs (probabilistic and otherwise).
There are constraints imposed by the laws of coexistence.
There are constraints imposed by the natures of the fundamental physical magnitudes.

One might wonder whether a more seamless account of metaphysical necessity can be had. I do not know – nor whether, if we are stuck with heterogeneity, that is a feature or a bug. But for what it's worth, here is one interesting speculative position that substantially reduces the heterogeneity.

It borrows from Jonathan Schaffer's (2010) priority monism, and adds a bold thesis about the "F" ingredient of the SCARF. On this view, there is, fundamentally, just one thing (The World). And it is characterized by a highly structurally complex fundamental magnitude. And – just as the nature of the magnitude *mass* issues in a space of possibilities for the things that instantiate it (namely, the possible values they can exhibit) – so too does the World-magnitude, by its nature, yield a range of possible values that the one thing that instantiates it can exhibit. Each such value *just is* a distinct and complete physical – and we might as well say *nomological* – possibility for reality as a whole. So on this view, the arena and the rules are not needed as elements of the SCARF; their work is done by the stuff and the features (well, feature). I find such a reduction of laws of nature to the one perfectly natural magnitude very cool, though I worry a little bit about the constitution relations (for the story about how less fundamental entities are grounded in the most fundamental entity certainly won't be the familiar mereological one), and worry much more about how to accommodate probabilistic FLOTEs.

24.5 How one might be led away from this view

There are many routes.

One might think that the notion of metaphysical necessity has uses *beyond* (or just *other* than) the explanation-oriented ones I've emphasized. (*Counterfactuals* get frequent mention here – though not, I think, for very good reasons.) One might think that explanation just isn't that important (van Fraassen and other "pragmatists" about explanation would presumably concur; see in particular van Fraassen 1980). One might think that explanation is important, and even that certain key metaphysical notions are needed to say what explanation is (e.g., the notion of a causal power or capacity), but revolt at the "fundamentalist" claim that all relations of explanatory dependence are grounded in constraints imposed by the most basic ingredients of the SCARF (see for example Cartwright 1999).

Or, one might think that explanation itself has just been *misunderstood*. That's the option I'm going to explore here. Go back to what I called a "canonical" form of explanation:

> "Why is it the case that P?"
> "Because, thanks to such-and-such fundamental feature of reality, it *must* be the case that P".

The objection I'm now considering begins thus:

> Phoney baloney. This "canonical" form of explanation is, in fact, the canonical form of pseudo-explanation. You have misunderstood what it is that really brings increased understanding. In particular, proper explanations certainly do not present a "metaphysically special" kind of information. Rather, proper explanations earn their status as such for a quite different reason.

That's a good idea, well worth exploring; the next section begins to do so.

24.6 The "unificationist" face of explanation

I think that everyone, regardless of their metaphysics, ought to agree that there is *more* to explaining than pointing out what the target of explanation depends on. A good way to see this is to focus not on the word "explain", but on the word "understand". If I want to understand, say, why a particular window broke, a jumbled list of causes of the breaking won't help me very much. What else is required? I'm not entirely sure, but I will adopt as a working hypothesis the following key idea: We can achieve a kind of understanding of a given body of material by coming to see how to *organize it by means of a suitably unifying framework*.

Now, what this notion of a "unifying framework" exactly comes to – what that extra ingredient is, that goes beyond merely possessing the right sort of information, and somehow consists in having that information *packaged* or *organized* in the right sort of way – is not, to me at least, very clear. As yet, all I can do is point to the notion by means of vivid examples. That's what I'll try to do next.

24.6.1 A simple math game

Consider the following simple two-player game. The players begin with a pile of coins. They take turns; on each player's turn, she may remove one or two coins from the pile. The player who takes the last coin wins. (Note: the players know how many coins are in the pile at the start.)

Think about this game for a minute. You will likely hit upon a simple strategy for the first player, one that will allow her to win, *provided* the number of coins at the beginning of the game is not a multiple of three: take enough coins so that the remainder *is* a multiple of three. A simple argument (exercise) shows that this strategy will win.

Now imagine being provided with a *different* method for winning the game: a simple lookup table, that tells you, for any number of coins in the pile (up to some suitably large number), whether to take 1 or 2. Consider the epistemic situation of a player who is quite confident that this lookup table will guide her play as well as possible – but who doesn't know *why* it is successful in this way. If you want to *explain* to her why the table works, you certainly can. But not by providing her with information about what *causes* the table to succeed, or more generally what its success *depends* on. No, you will just give her the rule from the last paragraph, together with the simple argument you came up with for why that rule works.

She now *understands* why the table worked, whereas she did not before. What changed? It wasn't, I think, so much that she *acquired new information* (and certainly not: new information about dependencies), but rather that, thanks to the rule plus the argument you gave her, she came to *organize her information in a more effective manner.* Notice, for example, that she can now play the game much more easily, with far less cognitive effort than was required to constantly reference the table. More interestingly, she can easily see how to generalize the strategy to variants – say, where the number of coins that can be taken is one, two, or three. She is, in short, in a much better position to reason about this game, and other games like it.

There's an important insight here, though I wish I knew better how to identify and exploit it. For the present, this will have to do: I suggest that one of our central aims as inquirers is that of developing more and more *cognitively effective ways of organizing our information* – ways, roughly, that will enhance our ability to engage in reasoning and deliberation about questions concerning the given subject matter. And I suggest that we judge goodness of explanations in part by how well they help us achieve this goal.

If these suggestions are on the right track, then the natural place to look, for examples that throw this epistemic aim into sharp relief, is to mathematics. For explanatory

dependency relations are thin on the ground there; and yet mathematicians routinely draw distinctions of explanatoriness, e.g. when comparing different proofs of the same result. The next example, which I learned from Tristam Needham's excellent *Visual Complex Analysis* (1997), is of this type.

24.6.2 A serious, math-geek example

Consider these two functions on the reals:

$$g(x) = \frac{1}{(1 - x^2)}$$

and

$$h(x) = \frac{1}{(1 + x^2)}$$

Each has a Taylor expansion, valid for the range $-1 < x < 1$:

$$g(x) = 1 + x^2 + x^4 + x^6 + \ldots$$
$$h(x) = 1 - x^2 + x^4 - x^6 + \ldots$$

But now a puzzle emerges. For both of these expansions become undefined at $x = \pm 1$. For $g(x)$, that's no surprise: after all, the function itself blows up at those values. But $h(x)$ doesn't. Is it just some sort of coincidence that its expansion becomes ill-behaved at exactly the same values?

Note clearly what's going on here: We have a result (that both expansions fail at $x = \pm 1$) whose *correctness* is not at all in question. But we don't understand *why it obtains* – and are even open to the possibility that there is *no explanation*, that this really *is* just a mathematical coincidence. It's not plausible that what we're hankering after, here, is some sort of *dependency* information.

Well, what *would* satisfy us? To see, let's take a look at what *does* (for happily, there *is* an explanation of the common behavior of the Taylor expansions).

The key move is to think of these functions as defined not just on the reals, but on the entire complex plane. Then we notice that $g(z) = h(iz)$, so that these functions are *essentially the same*: one is just the result of rotating the other 90° through the complex plane. Finally, it is easy to show that the expansion of this function is undefined for *all* complex numbers z such that $|z| = 1$. Mystery solved!

I think this is what just happened: we didn't merely learn a new way to derive the result we were wondering about. We learned a new and much more effective way of categorizing functions – a way that allowed us to see much more clearly that, and how, the difference between our two functions is mathematically superficial. We gained some mathematical knowledge, to be sure; but more importantly, we learned how to *organize* our mathematical knowledge in a more useful manner.

24.7 A unificationist alternative?

It would be beyond delightful to have a clear, sharp, illuminating account of this thing I've been calling "effective cognitive organization", an account that showed why achieving it is such an important epistemic goal. I don't have one. All I have are some examples that seem highly

suggestive, and the strong suspicion, in light of them, that philosophers such as Kitcher (1989) who have argued for the importance of explanatory unification were onto something. (And here I'm simply using "explanatory unification" as a name for what we achieve, when we manage to organize our information in the kind of cognitively effective manner illustrated in our examples.) But in the current dialectical context this lack doesn't matter. First, as noted, it seems to me overwhelmingly plausible that a complete theory of explanation must contain a unificationist component. And second, what I'm trying to do here is to make trouble for my proposal about metaphysical modality – or more exactly, the presuppositions about explanation that lie behind it. So I'm happy to spot the opponent a good theory of explanatory unification. With that theory in hand, she has a natural move to make, which is that this is *all we need* – that a good theory of explanatory unification is a good theory of explanation *simpliciter*.[11]

I have claimed that to explain something is, in part, to give information about that which it depends on – and that such dependency-relations are perfectly objective, and furthermore are grounded in metaphysical necessity (i.e., in constraints imposed on what is the case by the fundamental ingredients of the cosmic SCARF). Here is my opponent's attractive response:

> To explain something may, in many cases, involve providing information that is *couched in dependency-language*. But this language is *misleading* (philosophically speaking, anyway). Ultimately, what makes these explanations valuable as such – what makes them enhance our understanding – is *only* that they provide us with the means to achieve a better-organized view of the given target of inquiry. We find ourselves naturally working with concepts of *dependency* (and making a familiar sort of projective error, in thinking that these concepts answer to some metaphysically special feature of reality itself) simply because, for creatures like us, doing so provides an effective means for securing the kind of cognitive organization we crave.

That may well be right – though defending this unificationist alternative in detail is, I think, *much* more difficult than this quick sketch of it would suggest. The strongly felt intuition that, for example, the structure of spacetime *really does* constrain possible motions will have to be abandoned. Some novel account of why explanation is (typically?) asymmetric will need to be provided. And there are other costs. For example, taking this position will steer you straight towards a Humean position about laws of nature, which is a position not easy to square with actual scientific practice (as argued in Hall 2017). Still, you should take these worries in the proper spirit: In particular, I do not at all think they show that the balance of plausibility favors my proposal over this alternative. Rather, what they show is that we've got a real – and, to my mind, *very* interesting – philosophical debate here, concerning the fundamental nature of explanatory inquiry. If I'm right, the possibility of an interesting notion of *metaphysical* modality hinges on the outcome.

Notes

1 Compare, e.g., Shoemaker (1980, 1998).
2 For a very helpful survey of standard choices for P, see Rosen (2006).
3 Thanks to Jason Decker for the excellent label.
4 For surely *zombies* are conceivable; see for example Chalmers (1996).
5 For surely *empty* or *really boring worlds* with non-trivial laws of nature are conceivable. For more creative examples aimed at the same conclusion, see Carroll (1990).
6 For surely *homogeneous spinning spheres* (of various angular velocities) are conceivable; see Armstrong (1980).

7 Incidentally, you can see, right in this preamble, why a full and proper development of this sketch must investigate how metaphysical modality and grounding are connected.
8 Though typically the questions will have boring answers: "Is it logically or conceptually possible that the SCARF be different, in such-and-such a way?" "Um, yes."
9 And thanks to Tim Maudlin for pointing it out.
10 At least, he does if we assume that the underlying dynamics is stochastic, so the possible end states for the star are *not* constrained by the combination of the dynamical laws, together with the initial conditions.
11 That, I think, was Kitcher's view, though his own theory seems to me demonstrably inadequate.

References

Armstrong, David (1980). Identity through Time. In Peter van Inwagen (ed.), *Time and Cause: Essays Presented to Richard Taylor*. Dordecht: Reidel. pp. 67–78.
Carroll, John (1990). The Humean tradition. *Philosophical Review* 99 (2):185–219.
Cartwright, Nancy (1999). *The Dappled World: A Study of the Boundaries of Science*. Cambridge: Cambridge University Press.
Chalmers, David J. (1996). *The Conscious Mind: In Search of a Fundamental Theory*. Oxford: Oxford University Press.
Hall, Ned (2011). Causation and the Sciences. In Steven French & Juha Saatsi (eds.), *Continuum Companion to the Philosophy of Science*. New York: Continuum. pp. 96–119.
Hall, Ned (2017). Humean Reductionism about Laws of Nature. In B. Loewer & J. Schaffer (eds.), *A Companion to David Lewis*. Oxford: John Wiley and Sons.
Hall, Ned (2020). *Respectful Deflationism*. Unpublished manuscript.
Kitcher, Philip (1989). Explanatory unification and the causal structure of the world. In Philip Kitcher & Wesley Salmon (eds.), *Scientific Explanation*. Minneapolis: University of Minnesota Press. pp. 410–505.
Lewis, David K. (1986). Causal explanation. In David Lewis (ed.), *Philosophical Papers*, Vol. II. Oxford: Oxford University Press. pp. 214–240.
Lewis, David K. (1994). Humean Supervenience Debugged. *Mind* 103 (412):473–490.
Maudlin, Tim (2007). A modest proposal concerning laws, counterfactuals, and explanations. In Tim Maudlin (ed.), *The Metaphysics within Physics*. Oxford: Oxford University Press.
Needham, Tristan (1997). *Visual Complex Analysis*. Oxford: Oxford University Press.
Rosen, Gideon (2006). The Limits of Contingency. In Fraser MacBride (ed.), *Identity and Modality*. Oxford: Oxford University Press. pp. 13–39.
Schaffer, Jonathan (2010). Monism: The Priority of the Whole. *Philosophical Review* 119 (1):31–76.
Shoemaker, Sydney (1980). Causality and properties. In Peter van Inwagen (ed.), *Time and Cause*. Dordecht: Reidel. pp. 109–135.
Shoemaker, Sydney (1998). Causal and metaphysical necessity. *Pacific Philosophical Quarterly* 79 (1):59–77.
Strevens, Michael (2008). *Depth: An Account of Scientific Explanation*. Cambridge, MA: Harvard University Press.
van Fraassen, Bas, C. (1980). *The Scientific Image*. Oxford: Oxford University Press.

PART 6

Modality in logic and mathematics


Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Chapter 25
MODALITY IN MATHEMATICS

Øystein Linnebo and Stewart Shapiro

25.1 The modal status of pure mathematics

There is a long-standing and widely held view that mathematical objects—such as numbers, functions, and sets—exist of necessity, and that all truths of (pure) mathematics are themselves necessary. The necessity here is metaphysical, the broadest variety of modality (unless logical necessity counts as a kind of modality). This view can be traced to Plato, who in the *Republic* VII took the mathematicians of his day to task for using a dynamic language, talking of what one can *do* with the objects of mathematics:[1]

> [The] science [of geometry] is in direct contradiction with the language employed by its adepts … Their language is most ludicrous … for they speak as if they were doing something and as if all their words were directed toward action … [They talk] of squaring and applying and adding and the like … whereas in fact the real object of the entire subject is … knowledge … of what eternally exists, not of anything that comes to be this or that at some time and ceases to be.
>
> (Republic, VII)

Similar views remain popular among philosophers of mathematics. Of course, not all of them are Platonists. But even among anti-Platonists, it is widely assumed that *if* numbers or sets exist, then they "eternally exist", and facts about them hold of necessity.

Indeed, the modal status of numbers is sometimes taken as a reason to doubt that they exist, for the following reason. In most ordinary cases of knowledge, our belief is responsive to the relevant state of affairs, often by being caused by it. Consider someone's knowledge that there is a computer in front of her. It is the computer that causes her to form this belief. Moreover, had there not been a computer in front of her, she would not have formed that belief. Thus, the belief about the existence of a computer *covaries* with the presence or absence of a computer in front of the agent. This kind of covariation of a belief state with what the belief is about becomes impossible if the belief is about objects that exist and have their properties necessarily. On the view inherited from Plato, for example, it makes no sense to vary the arithmetical facts and examine how this would affect our arithmetical beliefs. Based on considerations of this sort, some philosophers contend that the Platonistic view renders us unable to make sense of

mathematical knowledge, and that we should therefore reject Plato's necessarily existing mathematical objects (see, for example, Goodman and Quine (1947), Benacerraf (1973), Field (1989)).

As is well known, Aristotle rejected Plato's view of mathematics. Aristotle is often understood as conceding that there are mathematical objects but maintaining that these depend for their existence on being exemplified by contingent physical objects. For example, the shape of triangularity exists, but only in virtue of being instantiated by physical triangles. This view appears to introduce an element of contingency into mathematics (though see Section 25.3 in this chapter). Destroy all triangles, and you destroy triangularity itself! In the past two centuries, this alternative view has lost favor among mathematicians and philosophers, who want a greater degree of autonomy for mathematics. Mathematics should not be entangled with empirical matters in this way. Triangularity is a perfectly good mathematical object, worthy of geometrical investigation, regardless of whether the physical world contains perfect triangles. Moreover, modern mathematics investigates mathematical structures of vast complexity and size (or "cardinality"), undeterred by the (very real) possibility that these abstract structures may lack instantiation in the physical world.

There remain some cases, however, where mathematicians study notions that are at least *prima facie* modal. Intuitively, a function is *computable* just in case it is possible to compute it; a formula is *deducible* just in case it is possible to deduce it; a property is *definable* just in case it is possible to define it; a function is *differentiable* if it is possible to differentiate it. It is common to define a "knot" to be an embedding of a circle in 3-dimensional Euclidean space. Intuitively, two knots are equivalent if it is possible to transform one of them into the other without "tying" or "untying". However, in each of these cases, the relevant property or relation is ultimately understood in terms of the *existence* of a particular mathematical object: a function is computable only if there is a Turing machine that computes it, or it has a recursive derivation, etc.; a formula is deducible just in case there is a derivation of it; a property is definable just in case it has an explicit or implicit definition; a function is differentiable just in case it has a derivative; and two knots are equivalent if there is an ambient isotopy that maps one into the other. In short, the dominant view appears to be that the prima facie modal notions are expressed as features of the eternal and unchanging universe.

Another possible exception to the dominant view that mathematical objects (if any) exist of necessity concerns what are sometimes called *impure* mathematical objects. An example would be the set whose single element is Plato. Since Plato (presumably) exists only contingently, so too, many philosophers contend, does his "singleton set". If Plato did not exist, there could be no such thing as the set whose single element is Plato (Fine (1981), Maddy (1990)). However, this is, at most, a small deviation from the orthodox view. For example, Fine accepts the orthodox view once restricted to *pure* mathematical objects (such as numbers or the empty set).

A small minority of philosophers take the more radical step of rejecting even the thesis that pure mathematical objects exist of necessity (if at all). According to Quine, the only reason we have to believe in abstract objects (or anything else, for that matter) is that they are necessary for natural science. Given that mathematics does appear to be indispensable to modern science, he therefore accepts the existence of mathematical objects. This is the so-called Quine-Putnam indispensability argument for the existence of mathematical objects. Now, as Mark Colyvan (2001) reminds us, the truth of any given scientific theory is only a contingent matter. Therefore, he contends, the existence of mathematical objects is itself contingent. Suppose, for example, that classical mechanics can be formulated adequately with only rational numbers, but that quantum mechanics or general relativity (or any theory that can reconcile those) requires real numbers. Then, for Colyvan, the existence of real numbers would turn on the truth of quantum mechanics or relativity over classical mechanics, a contingent matter.

25.2 Modal accounts of mathematics

Some programs in the philosophy of mathematics reject the existence of mathematical objects but insist that correct mathematical assertions are objective and, perhaps, necessarily true. The idea is to formulate (or reformulate) mathematics in modal terms. Prominent members of this school include Geoffrey Hellman (1989) and Charles Chihara (1990).

According to some *structuralists* about mathematics, arithmetic is not about particular objects—namely certain abstract natural numbers—but about any system of objects that are structured in the appropriate way. The relevant structure is characterized by the second-order Dedekind-Peano axioms; let PA be their conjunction. Consider a domain X, with a designated member a, and on which is defined a unary function f. What is it for this system of objects to be "structured in the appropriate way"? The answer is that PA must hold of the system—with X, a, and f playing the roles of N, 0, and the successor function s, respectively. This requirement is expressed by a formula $\mathrm{PA}(X, a, f)$ which is obtained from PA by substituting X, a, and f for N, 0, and s, respectively. Consider now an arithmetical statement φ. According to our structuralists, φ should be understood as stating that some structural relationship obtains in any appropriately structured system of objects, that is:

$$\forall X \forall a \forall f \left(\mathrm{PA}\left(X,a,f\right) \rightarrow \varphi\left(X,a,f\right)\right). \tag{25.1}$$

At least from a mathematical point of view, this analysis works perfectly—provided there exist systems with the appropriate structure. If this proviso fails, however, the antecedent of (25.1) is always false, thus making the entire generalization vacuously true, with the plainly unacceptable consequence that both φ and its negation come out true.

What to do? Hellman's *modal structuralism* makes an interesting proposal. It is immaterial whether there actually are systems with the appropriate structure. What matters is that it is *logically possible* for there to be such systems; that is:

$$\Diamond \exists X \exists a \exists f (\mathrm{PA}\left(X,a,f\right)$$

Suppose that we analyze φ as the statement that (25.1) holds by logical necessity. Hellman proves that this analysis yields all the right truth-values and inferential relationships between the formulas of arithmetic. We thus avoid a problematic ontological commitment to an appropriately structured system of objects. The price, however, is the adoption of some new "ideology", namely operators for logical possibility and logical necessity (and second-order quantifiers). These operators are taken as primitive, and thus cannot be understood as quantifiers over possible worlds or in terms of satisfaction over sets.

Chihara's (1990) main innovation is a modal primitive, a "constructibility quantifier". Syntactically, it behaves like an ordinary quantifier: if φ is a formula and x a certain type of variable, then $(Cx)\varphi$ is a formula, which is to be read "it is possible to construct an x such that φ". Chihara also invokes terminology for the satisfaction of sentences by objects. He then formulates a version of simple type theory in these terms, and shows how to reconstruct mathematics in this language (following the usual logicist reduction of arithmetic to type theory). For example, there is a theorem that it is possible to construct an open sentence (of level 2) which is satisfied by all and only the level 1 open sentences that are satisfied by exactly four objects. This open sentence plays the role of the number 4. But Chihara does not assert the outright existence of such open sentences. The theory only entails that it is *possible* to construct sentences like this.

With both of these programs, we see modality invoked to make sense of mathematics; in particular, to avoid problematic ontological commitments to vast structures of mathematical objects or concrete instantiations thereof. Much earlier, Hilary Putnam (1967) sketched a modal account of mathematics similar to Hellman's, but unlike Hellman and Chihara, Putnam argued that the modal account is in some sense "equivalent" to the ordinary non-modal one based on the existence of abstract objects. The modal claim about structural possibilities and the ontological claim about existence of abstract objects are "equivalent descriptions" of reality (see also Shapiro (1997) Chapter 7).

25.3 Potential infinity

Let us return to antiquity. Aristotle emphatically rejects the Platonic view that geometry produces "knowledge … of what eternally exists, not of anything that comes to be this or that at some time and ceases to be". For Aristotle, mathematics is about ordinary physical objects, except that they are studied in a particular manner. For example, the surfaces, lines, and points studied in geometry are literally contained in physical objects.

An interesting consequence is that mathematics becomes inherently modal. In particular, the scope of geometry is not limited to actual physical objects, but rather to *possible* physical objects. It is generally agreed that Euclid's *Elements* captures the mathematics of Plato's and Aristotle's day. Consider the first postulate, "To draw a straight line from any point to any point". Consider a corner of a cube, and a corner of a different cube. The postulate says that there is a line that connects those two points. But, clearly, there need not be a physical object with an edge that connects those two points. From Aristotle's perspective, one way to read this postulate is that if we start with two points, there *could be* a line that connects them.[2]

This sets the stage for consideration of another Aristotelian notion, that of potential infinity. Aristotle, along with ancient, medieval, and early modern mathematicians, recognized the existence of certain *procedures* that can be iterated indefinitely, without limit. Examples are the bisection and the extension of line segments. What was rejected are what would be the end results of applying these procedures infinitely many times, namely self-standing points, infinitely long regions, and infinite collections. Thus, in *Physics* 3.6 (206a27–29), Aristotle wrote, "For generally the infinite is as follows: there is always another and another to be taken. And the thing taken will always be finite, but always different" (206a27–29). This orientation towards the infinite was endorsed by mainstream mathematicians as late as Gauss (1831), who wrote: "I protest against the use of infinite magnitude as something completed, which is never permissible in mathematics".

Let us sketch an account of the Aristotelian notion of potential infinity in modal terms (see Linnebo, Shapiro, and Hellman (2016) and Linnebo and Shapiro (2019) for more detail). Aristotle's claim that matter is infinitely divisible provides a nice example. Consider a stick. However many times one has divided the stick, it is always possible to divide it again (or so it is assumed). It is fairly natural to explicate Aristotle's temporal vocabulary in a modal way. This yields the following analysis of the infinite divisibility of a stick s:

$$\Box\forall x\big(Pxs \to \Diamond\exists y\, Pyx\big), \tag{25.2}$$

where Pxy means that x is a *proper* part of y. If, by contrast, the divisions of the stick formed an actual infinity, the following would hold:

$$\forall x\big(Pxs \to \exists y\, Pyx\big). \tag{25.3}$$

According to Aristotle, it is not even possible to complete infinitely many divisions of the stick, that is:

$$\neg\Diamond\forall x\left(Pxs \rightarrow \exists y\, Pyx\right). \tag{25.4}$$

By endorsing both (25.2) and (25.4), Aristotle is asserting that the divisions of the stick are *merely potentially infinite*.

According to Aristotle, the sequence of natural numbers too is merely potentially infinite.[3] We can represent this view as the conjunction of the following theses:

$$\Diamond\forall m\Diamond\exists n\mathrm{Succ}\left(m,n\right) \tag{25.5}$$

$$\neg\Diamond\forall m\exists n\mathrm{Succ}\left(m,n\right) \tag{25.6}$$

Thus, provided we are willing to use the resources of modal logic, there is a simple and intuitive way to distinguish the merely potential infinite from the actual infinite.

There is, of course, a question about the nature of the indicated modality. Aristotle appears to have had in mind some form of metaphysical modality, and the resulting notion is also available to contemporary philosophers who believe there is such a thing as metaphysical necessity. There are other options as well. For example, constructivists take mathematics to be about actual and possible constructions—where possibility is understood in some suitably idealized form of possibility for epistemic agents such as ourselves. On this view, an existential generalization over mathematical objects is true in virtue of the possibility of constructing a "witness". With the mentioned modality in play, many constructivists proceed to endorse the potential infinite but reject even the possibility of the actual infinite. Yet another option is discussed in the next section.

Our modal analysis of potential infinity prompts a crucial question. Ordinary mathematical language—including that used by proponents of potential infinity, such as Gauss—is non-modal. How, then, should our modal analysis be connected up with the non-modal statements that we set out to analyze? Clearly, we need a translation to bridge the non-modal language of ordinary arithmetic and the modal language in which our analysis is stated. While several "bridges" are possible, we contend that one is particularly plausible.[4] When Gauss asserts that for every natural number there exists a successor, he should be understood as asserting (25.5). More generally, we should translate his arithmetical quantifiers as follows:

$$\forall \mapsto \Box\forall \qquad \exists \mapsto \Diamond\exists$$

Let us call this *the potentialist translation*, and let $\varphi^{\Diamond}$ be the translation of φ. Every statement φ of ordinary non-modal mathematics can thus be understood as implicitly making a modal statement $\varphi^{\Diamond}$ in a context that is friendly to potential infinity.

Suppose we use this bridge to connect the modal and non-modal languages. Let us say that an entailment "$\varphi_1, \ldots \varphi_n$, therefore ψ" obtains in the non-modal language just in case some corresponding entailment "$\varphi^{\Diamond}_1, \ldots, \varphi^{\Diamond}_n$, therefore $\psi^{\Diamond}$" obtains in the modal language. What logical principles does this definition validate in the non-modal language?

The question has a pleasing answer. For first-order languages, precisely the laws of ordinary classical logic are validated—no more, no less!—provided that certain plausible principles are assumed in the modal language. The most important of these principles captures the idea that the existential possibilities are cumulative: if two extensions of the domain are individually possible, then they are also jointly possible.

This mathematical result provides a justification of the use of classical first-order logic when reasoning about the potentially infinite. This would no doubt have pleased proponents of potential infinity, from Aristotle to Gauss, who relied on classical logic when reasoning about potentially infinite domains.[5]

25.4 Potentialism versus actualism

As mentioned, a broadly Aristotelian view of infinity was dominant in mathematics and philosophy for more than two millennia. The watershed in the eventual fall of this view was Georg Cantor's pioneering work of the late nineteenth century. For the sorts of cases discussed so far—such as geometrical regions and natural numbers—Cantor defended the exact opposite of Aristotle, namely that the potentially infinite is dubious, unless it is somehow backed by a corresponding actual infinity:

> I cannot ascribe any being to the indefinite, the variable, the improper infinite in whatever form they appear, because they are nothing but either relational concepts or merely subjective representations or intuitions (imaginationes), but never adequate ideas.
>
> (Cantor (1883), 205, note 3)

> … every potential infinite, if it is to be applicable in a rigorous mathematical way, presupposes an actual infinite.
>
> (Cantor (1887), 410–411)

As part of his defense of this view, Cantor developed an elegant and extremely fruitful mathematical theory of actually infinite sets and numbers, which was soon absorbed into mainstream mathematics.

We think it safe to say that this Cantorian orientation is now dominant in the relevant intellectual communities, especially concerning geometry and arithmetic. Indeed, the view is often known as "Cantor's paradise", in the famous words of the leading mathematician of the time, David Hilbert (1925). The main exceptions are constructivists of various forms, who still largely reject the actually infinite.

There is, however, some tension in Cantor's writings. On closer examination, we find that even Cantor leaves room for a limited form of potentiality or incompletability in mathematics, at least in places. While it is unproblematic to regard the sequence of *natural* numbers as a completed or actual infinity, the same does not hold of what Cantor called "the extended number sequence", namely the sequence of all *ordinal* numbers. We see this by reflecting on Cantor's view that for every "definite" sequence of numbers, there is a further number that specifies the order type of the given sequence (1883). Suppose that the extended number sequence is definite. It follows that there is a further number that is larger than each number in a sequence that has been supposed to comprise all numbers whatsoever. But this is a contradiction, as the mentioned number would have to be larger than itself. So, by *reductio ad absurdum*, we conclude that the extended number sequence is not definite after all. Instead, this sequence is inherently potential or incompletable. In short, while Cantor insists that Aristotle and his followers were wrong to deny that the sequence of natural numbers can be completed, he ascribes a form of incompletability to the sequence of ordinal numbers. While the former sequence is "transfinite", the latter is "absolutely infinite".

Cantor returns to this theme when the set theoretic paradoxes were discovered in the late 1890s. He insists that his theory is unaffected by the paradoxes as he was never committed to the sorts of set that would engender paradox, such as the Russell set or the set of all ordinal (or cardinal) numbers. He admits there is a "multiplicity" of all non-self-membered sets and of ordinal (or cardinal) numbers. But he denies that these multiplicities form sets. As justification for the latter claim, Cantor asserts that these are absolutely infinite multiplicities, or what he dubbed "inconsistent multiplicities", with features closely analogous to those of the potentially infinite. In a much-quoted letter to Dedekind, in 1899, he wrote:

> [I]t is necessary … to distinguish two kinds of multiplicities (by this I always mean definite multiplicities). For a multiplicity can be such that the assumption that all of its elements 'are together' leads to a contradiction, so that it is impossible to conceive of the multiplicity as a unity, as 'one finished thing'. Such multiplicities I call *absolutely infinite* or *inconsistent multiplicities* … If on the other hand the totality of the elements of a multiplicity can be thought of without contradiction as 'being together', so that they can be gathered together into 'one thing', I call it a *consistent multiplicity* or a 'set'.
>
> (Ewald (1996), 931–932)

An 1897 letter to Hilbert is even more suggestive:

> I say of a set that it can be thought of as finished … if it is possible without contradiction (as can be done with finite sets) to think of all its elements as existing together … or (in other words) if it is possible to imagine the set as actually existing with the totality of its elements.
>
> (Ewald (1996), 927)

Let *potentialism* be the view that there are (as Cantor says) types of object whose instances cannot "exist together" or "be thought of as finished". This view generalizes the ancient idea of potential infinity by allowing transfinite multiplicities, such as that of natural numbers, to be completed but insists on the incompletability of other, absolutely infinite multiplicities. Let *actualism* be the opposing view that all multiplicities are completable, or, relatedly, that we have no need for the concept of "incompletability", which should accordingly be abandoned (see Linnebo (2013) and Parsons (1977)).

Let us sketch one way to clarify the two views. As in the analysis of orthodox potential infinity, we use modal operators to represent the modal notions involved in the view. Additionally, we need a way to represent that some objects "exist together". The resources of plural logic allow us to do this. This logic adds to ordinary first-order logic plural variables, such as xx and yy, each of which can have one *or more* values, rather than just a single value, as in the case of ordinary (or singular) variables. We also add a logical predicate $\prec$, where $x \prec yy$ is to be understood as "x is one of yy". We require that the (one or more) values of each plural variable "exist together". Where P is some singular predicate, let P^* be the plural predicate defined by letting $P^*(xx)$ abbreviate $\forall y(y \prec xx \rightarrow Py)$. For example, Dog* is true of some objects just in case each of them satisfies the predicate Dog.

We can now express the claim the multiplicity of sets is incompletable or cannot "exist together" as follows:

$$\Box\forall xx\left(\mathrm{Set}^*\left(xx\right) \rightarrow \Diamond\exists y\left(\mathrm{Set}\left(y\right) \wedge y \not\prec xx\right)\right)$$

That is, necessarily, for any coexisting sets, possibly there is another set that is not among the mentioned sets. Indeed, if one thinks—as on the influential iterative conception of sets (Boolos (1983) and (1989))—that no set can be an element of itself, then the set y can be chosen to be the set whose elements are precisely xx.

Of course, to clarify potentialism, it is not sufficient just to formalize the relevant claims: we also need to explain the notation used in our formalizations. We have already commented on how the plural resources are to be understood. It remains to say something about the modal operators. As discussed, Aristotle and some of his followers understood these operators in terms of some form of metaphysical modality. But this is not an attractive option in the present context, since pure sets are widely assumed to exist by metaphysical necessity, if at all (cf. Section 25.1). Fortunately, there are various alternatives. One option is to follow constructivists and understand the modality in terms of suitably idealized abilities that we (perhaps aided by computers) have for constructing mathematical objects. Another option is to follow Putnam and Hellman (cf. Section 25.2) and invoke some form of logical modality. Yet another option has been proposed by Fine (2006) and Linnebo (2012, 2013) namely to let the modality vary, not the circumstances (as in the previous proposals), but the interpretation of the language.

We cannot here attempt to adjudicate between the resulting forms of potentialism and actualism. Instead, we wish to close by briefly mentioning two other views that may be illuminated by means of potentialism, in the sense we have just articulated.

First, there is the view that the cumulative hierarchy of sets is "open ended". According to the iterative conception of sets, the sets are formed in stages. At stage 0, we begin with some non-sets or "urelemente". At stage 1, we form all sets of objects available at stage 0. We continue in this way up through all the finite stages. At stage ω and other stages indexed by so-called limit ordinals, we simply take the union of all the sets formed at earlier stages. This process is now continued "as far as possible". But what does this mean? However large a cumulative hierarchy we have formed, it seems possible to continue and form even more sets. Based on such considerations, many theorists have thought the cumulative hierarchy is somehow "open ended". One expression of this view is found in Ernst Zermelo's famous 1930 paper:

> But [the set-theoretic paradoxes] are only apparent "contradictions", and depend solely on confusing set theory itself, which is not categorically determined by its axioms, with individual models representing it. What appears as an "ultrafinite non- or super-set" in one model is, in the succeeding model, a perfectly good, valid set with both a cardinal number and an ordinal type, and is itself a foundation stone for the construction of a new domain.

This can be understood as the claim that necessarily, given any coexisting objects that form a model of set theory, it is possible to extend the model such that every proper class (or "ultrafinite non- or super-set") of the initial model corresponds to "a perfectly good, valid set" of the extended model.

Next, there is Michael Dummett's influential, but notoriously obscure, notion of an indefinitely extensible concept (1963; 1981, pp. 428–41; 1991, esp. pp. 312–21). Dummett characterizes a concept as indefinitely extensible just in case, whenever

> we can form a definite conception of a totality all of whose members fall under that concept, we can, by reference to that totality characterize a larger totality all of whose members fall under it.
>
> (1991, p. 441)

Let X range over "totalities", and let $D(X)$ mean that X is definite. A concept is then indefinitely extensible, according to Dummett, just in case:

$$\forall X\big(D(X) \wedge \forall u\big(u \in X \rightarrow Fu\big) \rightarrow \exists y\big(Ey \wedge y \notin X\big)\big) \qquad \text{(IE)}$$

Frustratingly, Dummett fails to explain how he understands the crucial notion of definiteness. This has understandably hampered later discussions of indefinite extensibility (but see Shapiro and Wright (2006)).

Our analysis of potentialism suggests an interesting way to provide this missing definition. Perhaps a definite totality is simply what Cantor calls a "consistent multiplicity", that is, what we represent as a plurality. This yields the following analysis of indefinite extensibility:

$$\forall xx\big(F^*(xx) \rightarrow \Diamond\exists yEy \wedge y \prec\!\!\!\!\!- xx\big) \qquad \big(\text{IE}^+\big)$$

That is, a concept F is indefinitely extensible just in case: necessarily, given any coexisting instances of the concept, possibly there is another instance that is not among the mentioned instances. In short, a concept is indefinitely extensible just in case it is impossible to specify some objects that are its extension, because any proposed extension makes it possible to define yet further instances of the concept.

25.5 Concluding remarks

In Section 25.1, we explained how the dominant view today is the Platonic one that all truths of pure mathematics are metaphysically necessary. Since these truths could not have been otherwise, they can safely be relied on when we theorize about counterfactual scenarios, as we routinely do when reasoning about an uncertain future or try to learn from past mistakes.

It is controversial, however, whether we should follow Plato in other respects, as we discuss in Sections 25.2 and 25.4. In Section 25.2, we described some views that reject a Platonic ontology of abstract mathematical objects in favor of modal claims about constructions or suitably structured systems of objects (which may well be concrete).

Returning to antiquity, we explained in Section 25.3 how there is a coherent and interesting concept of potential infinity. Although most mathematicians and philosophers today follow Cantor in accepting the actual infinite, this was a substantive intellectual victory over a subtle and resourceful opponent.

Drawing inspiration from Cantor himself, Section 25.4 explored a successor concept to the ancient one of potential infinity, which we called potentialism. According to this view, there are mathematical domains that differ from concrete ones by being incompletable or incapable of having all of its members co-exist. This too is a controversial step away from the Platonic conception of mathematics—but one that does not drive us out of Cantor's paradise.

Notes

1 According to Proclus (1970, p. 125) the "problem" of dynamic language in geometry occupied those in Plato's Academy for some time.

2 We make no attempt to make sense of Aristotle's own account of modality, a matter that is particularly vexed. Here we invoke contemporary modal notions in order to sketch an Aristotelian account of mathematics.

3 For Aristotle, a number is always a number of some things. So, it is perhaps a bit misleading to speak of *the* sequence of natural numbers. We trust that no confusion results from this minor deviation.
4 One historically and technically important option is Gödel's (1969) translation of intuitionistic logic into the modal logic S4. This translation agrees with the one we favor concerning the universal quantifier, but it translates the existential quantifier as itself, thus failing to capture the potentialist idea that existence talk in non-modal mathematics is short for talk about *potential* existence.
5 Various extensions of this result are discussed in Linnebo and Shapiro (2019). In higher-order settings, not all comprehension axioms are necessarily validated. Moreover, if intuitionistic rather than classical logic is employed in the modal language, but other assumptions are kept in place, then intuitionistic rather than classical logic is validated in the non-modal language.

References

Aristotle, *The basic works of Aristotle*, edited by R. McKeon, Random House, 1941.
Benacerraf, Paul, "Mathematical truth", *Journal of Philosophy* 70 (1973), 661–679.
Boolos, George, "The iterative concept of set", in *Philosophy of mathematics*, edited by Paul Benacerraf and Hilary Putnam, 2nd ed., Cambridge, Cambridge University Press, 1983, 486–502.
Boolos, George, Iteration again, *Philosophical Topics* 17 (1989), 521.
Cantor, G., *Grundlagen einer allgemeinen Mannigfaltigkeitslehre. Ein mathematisch-philosophischer Versuch in der Lehre des Unendlichen*, Leipzig, Teubner, 1883.
Cantor, G., "Mitteilugen zur Lehre vom Transfinitten 1, II", *Zeitschrift für Philosophie und philosophische Kritik* 91 (1887), 81125, 252–270; 92, 250–265; reprinted in Cantor (1932), 378–439.
Cantor, G., *Gesammelte Abhandlungen mathematischen und philosophischen Inhalts* edited by E. Zermelo, Berlin, Springer, 1932.
Chihara, Charles, *Constructibility and mathematical existence*, Oxford, Oxford University Press, 1990.
Colyvan, Mark, *The indispensability of mathematics*, Oxford, Oxford University Press, 2001.
Dummett, Michael, "The philosophical significance of Gödel's theorem", *Ratio* 5 (1963), 140–155.
Dummett, Michael, *Frege: philosophy of language*, 2nd ed., Cambridge, Massachusetts, Harvard University Press, 1981.
Dummett, Michael, *Frege: philosophy of mathematics*, Cambridge, Massachusetts, Harvard University Press, 1991.
Euclid, *The thirteen books of Euclid's Elements*, translation and commentary by Thomas L. Heath, 2nd rev. ed., New York, Dover Publications, 1956.
Ewald, William, editor, *From Kant to Hilbert: a sourcebook in the foundations of mathematics*, Oxford, Oxford University Press, 1996.
Field, Hartry, *Realism, mathematics and modality*, Oxford, Blackwell, 1989.
Fine, Kit, "First-Order Modal Theories I—Sets", *Noûs* 15 (1981), 177–205.
Fine, Kit, "Relatively unrestricted quantification", in *Absolute generality*, edited by Agustín Rayo and Gabriel Uzquiano, Oxford, Oxford University Press, 2006, 20–44.
Gauss, Karl Friedrich, "*Briefwechsel mit Schumacher*", Werke, Band 8, 216 (1831).
Gödel, Kurt, "An interpretation of the intuitionistic sentential logic", in *The philosophy of mathematics*, edited by J. Hintikka, Oxford, Oxford University Press, 1969, 128–129.
Goodman, Nelson, and W.V. O. Quine "Steps toward a constructive nominalism", *Journal of Symbolic Logic* 12 (1947), 105–122.
Hellman, Geoffrey, *Mathematics without numbers: towards a modal-structural interpretation*, Oxford, Oxford University Press, 1989.
Hilbert, David, "Uber das Unendliche", *Mathematische Annalen* 95 (1925), 161–190; translated as "On the infinite", in *Philosophy of mathematics*, second edition, edited by Paul Benacerraf and Hilary Putnam, Cambridge, Cambridge University Press, 1983, 183–201.
Linnebo, Øystein, "Reference by abstraction" *Proceedings of the Aristotelian Society* 112 (2012), 45–71.
Linnebo, Øystein, "The potential hierarchy of sets", *Review of Symbolic Logic* 6 (2013): 205–228.
Linnebo, Øystein, Stewart Shapiro, and Geoffrey Hellman, "Aristotelian continuua", *Philosophia Mathematica* 24 (3) (2016): 214–24.
Linnebo, Øystein and Stewart Shapiro, "Actual and potential infinity", *Noûs* 53 (2019): 160–191.
Maddy, Penelope, *Realism in mathematics*, Oxford, Oxford University Press, 1990.

Parsons, Charles, "What is the iterative conception of set?", in *Logic, foundations of mathematics and computability theory*, edited by R. Butts and J. Hintikka, Dordrecht, Holland, D. Reidel, 1977, 335–367.

Plato, *The Republic of Plato*, translated by Francis Cornford Oxford, Oxford University Press, 1945

Proclus, *Commentary on Euclid's elements I*, translated by G. Morrow, Princeton, Princeton University Press, 1970.

Putnam, Hilary, "Mathematics without foundations", *Journal of Philosophy* 64 (1967): 5–22.

Shapiro, Stewart, *Philosophy of mathematics: structure and ontology*, New York, Oxford University Press, 1997.

Shapiro, Stewart and Crispin Wright, "All things indefinitely extensible", in *Absolute Generality*, edited by *Agustín Rayo and Gabriel Uzquiano*, Oxford, Oxford University Press, 2006, 255–304.

Zermelo, Ernst, "Uber Grenzzahlen und Mengenbereiche: Neue Untersuchungen über die Grundlagen der Mengenlehre", *Fundamenta Mathematicae* 16 (1930), 29–47; translated as "On boundary numbers and domains of sets: new investigations in the foundations of set theory", in Ewald (1996), pp. 1219–1233.

Chapter 26
MODAL SET THEORY*

Christopher Menzel

26.1 Modal set theory and traditional modal metaphysics

Set theory is the study of sets using the tools of contemporary mathematical logic. *Modal* set theory draws in particular upon contemporary modal logic, the logic of necessity and possibility. One simple and obvious motivation for modal set theory is the fact that, from a realist perspective that takes the existence of sets seriously, sets have philosophically interesting modal properties. For instance, perhaps the most notable and distinctive property of sets is their *extensionality*: sets a and b are identical if they have exactly the same members; formally, where we take variables from the lower end of the alphabet to range over sets:

$$\textbf{Ext} \qquad \forall a \forall b \Big(\forall x \big(x \in a \leftrightarrow x \in b \big) \rightarrow a = b \Big).$$

Intuitively, however, extensionality is not a contingent matter, a mere matter of happenstance. Rather, there simply *couldn't* have been distinct sets that shared all their members; there is no such possible world. That is, at a minimum, we want to be able to express that extensionality is a *necessary truth*:

$$\Box\textbf{Ext} \qquad \Box\forall a \forall b \Big(\forall x \big(x \in a \leftrightarrow x \in b \big) \rightarrow a = b \Big).$$

Clearly, however, there is much more than this to the modal connection between sets and their members. For note that, for all $\Box$ **Ext** tells us, one and the same set a could have vastly different members from one world to the next, so long as it remains the case that, in each world, no other set has exactly the same members as a in that world. Intuitively, however, the intimate connection between a set and its members is maintained *across* worlds; if a set has Angela Merkel, say, as a member, it could not possibly have failed to have her as a member. Sets, that is to say, have their members *essentially*; if x is a member of a, then it is a member of a in every world in which a exists, i.e., in every world in which something is identical to a; formally, letting $E!t$ abbreviate $\exists y\, y = t$, for terms t:

$$\textbf{E}\in \qquad \Box\forall a \forall x \Big(x \in a \rightarrow \Box\big(E!a \rightarrow x \in a \big) \Big),$$

* The author is exceedingly grateful to Øystein Linnebo, Neil Barton, and John Wigglesworth for helpful and illuminating comments on earlier drafts of this entry.

Likewise, *non*-membership; a set cannot "add" new members in one world that it lacks in another:

$$\mathbf{E}\notin \qquad \Box\forall a\forall x\Big(x\notin a\rightarrow\Box\big(E!a\rightarrow x\notin a\big)\Big).$$

A related philosophical issue is not settled by the preceding principles. Suppose x is a member of a set a here in the actual world and that a exists in some other possible world w. Then by **E∈**, x is a member of a in w. But nothing follows about x's *existence* in w. For all we know from **E∈**, all sets might exist necessarily, even those that have contingent members, members that might not themselves have existed. Hence, if that is so, the singleton set {Merkel}, for example, would have existed even if Merkel hadn't. However, on most conceptions of set, sets are *ontologically dependent* upon their members and, hence, could not themselves exist without their members existing; there could be no singleton set {Merkel} without Merkel. On such a conception, then, a further modal principle is needed to guarantee this:[1]

$$\mathbf{OD} \qquad \Box\forall a\forall x\Big(x\in a\rightarrow\Box\big(E!a\rightarrow E!x\big)\Big).$$

Given **OD** and the assumption that Angela Merkel is a contingent being, $\Diamond\neg\, E!m$, it now follows that the set {Merkel} too is contingent, as expected; it fails to exist at any Merkel-free possible world.

The preceding principles are motivated by the commonsense assumption that ordinary individuals like Merkel are indeed contingent, that they might have failed to exist. Surprisingly, at first blush anyway, this assumption is neither philosophically nor mathematically trivial. Regarding the latter, it is in fact a well-known theorem of the simplest and most straightforward system of modal predicate logic that there neither are nor could have been any contingent beings, i.e., that, necessarily, everything there is exists necessarily:[2]

$$\mathbf{Nec} \qquad \Box\forall x\Box E!x.$$

Avoiding this consequence requires choosing between restrictions (of varying severity) on one's logical system, each with its own virtues and liabilities.[3] However, some philosophers—so-called *necessitists*—embrace **Nec** (from which, of course, **OD** trivially follows) and choose instead to offer sophisticated philosophical explanations of the allegedly mistaken commonsense intuition that some things might not have existed.[4] The choice of a logic and the adoption of a philosophical standpoint about the metaphysics of sets are therefore interestingly interdependent.

It is not our purpose here to adjudicate these issues. Rather, the point of this initial section has been to illustrate one powerful motivation for modal set theory, namely, its usefulness as a tool for exploring quite traditional lines of inquiry in modal metaphysics concerning contingency, essentiality, ontological dependence, and the like that surface naturally in connection with the existence of sets.[5] The remainder of this chapter, however, will be devoted to the development of modal set theory with regard to a rather more directed inquiry into both the nature and *structure* of sets that is motivated in particular by the attractive prospect of a satisfying explanation of Russell's Paradox.

26.2 ZF and Russell's Paradox[6]

The years 1897–1903 saw the emergence of a string of related paradoxes concerning the notions of number, set, class, property, proposition, and truth.[7] Among those concerning sets,

Russell's Paradox is undoubtedly the best known and arguably the one most directly responsible for subsequent developments in the foundations of mathematics. The argument is well-known. Its heart, of course, is the principle of *naïve comprehension*, i.e., the principle that, for any property of things, there is the set of things that have the property. More formally (and somewhat anachronistically) expressed in the language of first-order logic, it is the principle that, for any predicate $\varphi(x)$, there is the set of things it is true of:

$$\textbf{NC} \qquad \exists a \forall x \left(x \in a \leftrightarrow \varphi(x) \right), \text{where '}a\text{' is not free in } \varphi(x).$$

Intuitively, at first sight anyway, the principle seems airtight. For a well-defined predicate unambiguously picks out some existing things (or perhaps no things at all), and what more could you need for the existence of a set than the existence of its purported members, the things that constitute it? For all its intuitive appeal, of course, **NC** is inconsistent: given the well-defined predicate '$x \notin x$', by **NC** we have the "Russell set" r containing exactly those things x that are not members of themselves, i.e.,

$$\textbf{R} \qquad \forall x \left(x \in r \leftrightarrow x \notin x \right).$$

Instantiating with r, the contradiction that $r \in r$ if and only if $r \notin r$ follows immediately.

The best known and most influential response to Russell's Paradox is of course that of Ernst Zermelo. To express Zermelo's ideas, it is useful to speak of mere *pluralities* of things, where such talk is to be thought of as "ontologically innocent". That is, talk of a plurality of things is not to be understood to refer to some *further* thing over and above the things we are talking about—a set or class or mereological sum that they constitute—but, rather, simply as a convenient way of talking about those things jointly, or collectively, as we seem freely to do when we use plural noun phrases in sentences like "It took *three men* to lift the piano" and "*The fans* went wild". The lesson of Russell's Paradox, then, in these terms, is that not all pluralities of things can *safely* be assumed to constitute a further thing, viz., a set that contains them; in particular, to assume without qualification, as **NC** would have it, that the things an arbitrary predicate is true of constitute a set can be logically catastrophic. At the same time, some pluralities seem clearly safe. Zermelo's brilliantly executed idea (Zermelo, 1908)[8]—implemented in his axiomatic set theory Z—was to *stipulate* the existence of some initial sets to get things going and then introduce a variety of sound "set-building" operations that lead safely from given objects or sets to new sets. We will describe Z in some detail.

Zermelo begins with the extensionality axiom **Ext**. His next axiom, the axiom of *elementary sets*, is actually a combination of an existence axiom and a set-building axiom. Specifically, he postulates the existence of the empty set $\varnothing$,

$$\textbf{ES} \qquad \exists a \forall x \, x \notin a,$$

and introduces the axiom of *Pairing*, which says, in effect, that any pair of (not necessarily distinct) objects x and y are jointly safe and hence constitute a set $\{x, y\}$:[9]

$$\textbf{Pr} \qquad \exists a \forall z \left(z \in a \leftrightarrow \left(z = x \vee z = y \right) \right).$$

Assuming extensionality, these two axioms alone already give us the power to prove the existence of the infinite series of *Zermelo numbers* $\varnothing$, $\{\varnothing\}$, $\{\{\varnothing\}\}$, $\{\{\{\varnothing\}\}\}$, …, so-called because they served as Zermelo's surrogates for the natural numbers (for convenience, abbreviate them,

respectively, as $\varnothing_0$, $\varnothing_1$, $\varnothing_2$, $\varnothing_3$, …). However, in addition to so-called "pure" sets like these that are "built up" solely from the empty set, Zermelo also made room in his theory for the existence of arbitrarily many *urelements*, that is, things that are not themselves sets—persons, planets, natural numbers, etc.—and, hence, by **Pr**, for the existence of "impure" sets built up from them. And although he didn't explicitly assume it as an axiom in 1908, it is in the spirit of his theory to take the urelements to constitute a set U of their own:[10]

$$\textbf{Ur} \qquad \exists a \forall x \Big(x \in a \leftrightarrow \big(x \neq \varnothing \wedge \forall y \, y \notin x \big) \Big).$$

Let *ZU* be Zermelo's theory Z with the additional axiom **Ur**.

NC is of course absent from Zermelo's theory, but a significant remnant of it remains in the form of a set-building principle, *Separation*. Given a predicate $\varphi(x)$, **NC** called sets into being *ex nihilo* from the things of which $\varphi(x)$ is true. Separation, by contrast, vouches only for the things $\varphi(x)$ is true of that are already members of some previously given set:

$$\textbf{Sep} \qquad \forall b \exists a \forall x \Big(x \in a \leftrightarrow \big(x \in b \wedge \varphi\big(x\big) \big) \Big), \text{where '}a\text{' is not free in } \varphi\big(x\big).^{11}$$

Zermelo's next two set-building axioms are *Union*,

$$\textbf{Un} \qquad \exists a \forall x \Big(x \in a \leftrightarrow \exists c \big(c \in b \wedge x \in c \big) \Big),$$

which says that the members of the members of a given set b constitute a set, and *Powerset*,

$$\textbf{Pow} \qquad \exists a \forall x \big(x \in a \leftrightarrow x \subseteq b \big),$$

which tells us that the subsets of a set safely constitute a set. Although formulated decades before the mature conception of the set theoretic universe in Zermelo (1930), these two axioms clearly anticipated it. For, given the initial set U of urelements—call it U_0—by **Pow** the set $\wp(U_0)$ of all of their subsets exists. By **Un** there is the set $U_1 = U_0 \cup \wp(U_0)$ consisting of all the members of U_0 and $\wp(U_0)$. Applying **Pow** again we have the set $\wp(U_1)$ of all the subsets of U_1 which we can then join with U_1 itself to yield the set $U_2 = U_1 \cup \wp(U_1)$. In general:

$$\textbf{D1} \qquad \begin{aligned} U_0 &= U \\ U_{n+1} &= U_n \cup \wp\big(U_n\big) \end{aligned}$$

Even in Zermelo's early work, then, the sets are naturally taken to have a structure that is *cumulative* and *hierarchical*, advancing "upwards" via iterations of the powerset and union operations, from an initial stock of urelements, in an ever-expanding series of *stages*, or *levels*, each successive level U_{n+1} consisting of everything in the preceding level together with all the sets that can be formed from them, as indicated in Figure 26.1.

Say that one level U_n is *higher* than another U_m just in case $n > m$ (equivalently, in light of their cumulative nature, just in case $U_m \subset U_n$) and that the *level* $\lambda(x)$ of an object x is the first level of the hierarchy in which it occurs. Since (a) we begin with a base level U_0 of urelements, (b) the hierarchy grows discretely from one level to the next, and (c) a set of level U_{n+1} is always constituted by objects in level U_n, it should also be clear that the sets on this conception are all

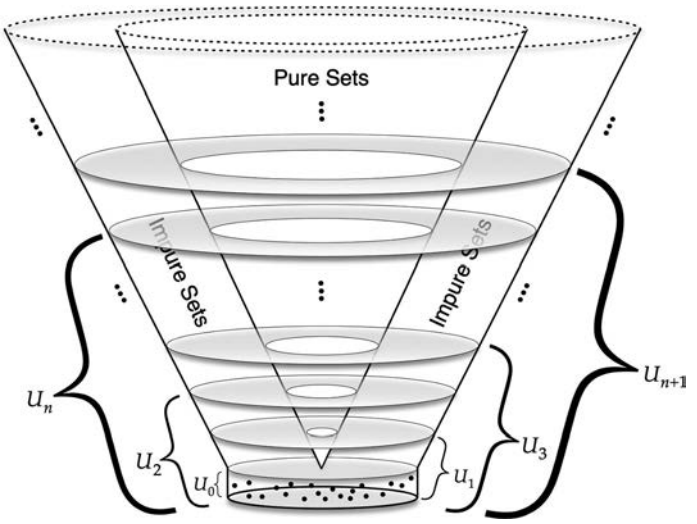


Figure 26.1 The finite levels of the cumulative hierarchy

well-founded: no set can be a member of itself and, more generally, there are no infinitely descending membership chains $\ldots \in a_{n+1} \in a_n \in \ldots \in a_1 \in a_0$. Since the aforementioned axioms do not explicitly rule out such structural impossibilities, it must be done independently by means of a separate principle; in Z, this is the axiom of *Foundation*, which requires every nonempty set a to have a member with which it shares no members:[12]

$$\textbf{Fnd} \qquad a \neq \varnothing \rightarrow \exists y\big(y \in a \wedge \forall z\big(z \in y \rightarrow z \notin a\big)\big).$$

We turn now to the critical Zermelian axiom of *Infinity*. What is particularly important about this axiom, especially for purposes here, is not merely that it asserts the safety of an infinite plurality—**Ur** will have already done that on the assumption that there are infinitely many urelements—but, rather, the safety of a plurality that is *unbounded* in our hierarchy of finite (i.e., finitely indexed) levels U_n. Consider, in particular, the Zermelo numbers $\varnothing_0, \varnothing_1, \varnothing_2, \ldots$ As $\lambda(\varnothing_i) = U_{i+1}$, for all natural numbers i, it follows that, for every finite level U_n, no matter how high, there is a Zermelo number ($\varnothing_n$, for example) that only first occurs in a higher level still; the Zermelo numbers are thus unbounded in the hierarchy of finite levels $U_0, U_1, U_2, \ldots$, and, hence, never constitute a set in any of them.

Such unbounded pluralities, then, are of a rather different sort structurally than any we've encountered hitherto. But, ultimately, from the realist's standpoint, at least, there doesn't seem to be any more reason to question their safety than there is to question the safety of the urelements that we sanction in **Ur** that get the hierarchy going in the first place or the plurality of subsets of a given level that we sanction in **Pow** that enable us, at any given level, to extend the hierarchy to the next level. For, just as in those cases, all of the things in question are *there* and hence *available* for collection into a set. Moreover, as the Zermelo numbers are unbounded in the hierarchy of finite levels and hence could not constitute a set at any such level, the set they would constitute would definitively not be a member of itself. So there also seems to be no hint of a Russell-style paradox in the offing.

Accordingly, Zermelo's Infinity axiom posits the collective safely of the Zermelo numbers the way that **Ur** does for the urelements and **Pow** does for the subsets of a given set. More exactly, it declares that there is a set that *contains* the Zermelo numbers:

$$\textbf{Inf} \qquad \exists a\Big(\varnothing \in a \wedge \forall x\Big(x \in a \rightarrow \{x\} \in a\Big)\Big).$$

The set of Zermelo numbers proper—call it ζ (zeta)—can then be derived straightaway by **Sep**.[13]

So the axiom of Infinity implies, not just that there are infinite sets, but that there are infinite sets whose members occur at arbitrarily high finite levels of the hierarchy. Hence, intuitively, the cumulative hierarchy $U_0, U_1, U_2, \ldots$, of cumulative levels must continue *beyond* the finite. For this to be provable, however, we must first show that they themselves form a set $\{U_0, U_1, U_2, \ldots\}$. Given their structural similarity to the Zermelo numbers $\varnothing_0, \varnothing_1, \varnothing_2, \ldots$, it certainly seems that they *should*. But **Inf** and the other axioms are not enough to guarantee this.[14] Thus, a further principle is needed, one typically attributed to Abraham Fraenkel,[15] the axiom schema of *Replacement*, the addition of which to Zermelo's theory Z (or ZU) gives us ZF (ZFU). Replacement captures the structural intuition that if the members of a set can be correlated one-to-one with a given plurality—each $\varnothing_n \in \zeta$ with U_n, for example—then that plurality also constitutes a set. More formally, where $\exists^! x\varphi$ as usual means that something is uniquely φ:

$$\textbf{Rep} \qquad \forall x \exists! y \psi\left(x,y\right) \rightarrow \forall b \exists a \forall y\Big(y \in a \leftrightarrow \exists x\Big(x \in b \wedge \psi\left(x,y\right)\Big)\Big).$$

Let $L(y)$ mean that y is one of our finite levels.[16] Letting $\psi(x,y)$ be '$x \in \zeta \wedge L(y) \wedge x \in y \wedge \forall z$ $((L(z) \wedge x \in z) \rightarrow y \subseteq z)$', i.e., "$x$ is a Zermelo number and y is its level". $\psi(x,y)$ correlates the members of ζ one-to-one with the levels $U_1, U_2, \ldots$ and, hence, by **Rep**, they constitute a set to which (by **Pr** and **Un**) we can add the initial level U_0 of urelements; so all the levels jointly constitute our desired set $\{U_0, U_1, U_2, \ldots\}$. By **Un** their union is a set, so all the members of all the finite levels do indeed form a set U_ω of their own, the first *transfinite* level of the hierarchy[17]—the result, as it were, of putting a "disk" atop the hierarchy of finite levels depicted in Figure 26.1 indicating its "completion" at a further level. That of course is not the end of the hierarchy but simply a new starting point for iterating the powerset and union operations to generate yet further levels $U_{\omega+1}, U_{\omega+2}, \ldots$, which (by **Rep**) jointly form a set and (by **Un**) hence constitute a new limit level $U_{\omega+\omega}$, and thus once again further levels $U_{\omega+\omega+1}, U_{\omega+\omega+2}, \ldots$ and so on, as depicted in Figure 26.2. In general, then, by including a limit clause representing the continual "completion" of these unbounded series of levels, we can define the entire transfinite cumulative hierarchy for all ordinal numbers, finite and transfinite alike:

$$\textbf{D2} \qquad \begin{aligned} U_0 &= U \\ U_{\alpha+1} &= U_\alpha \cup \wp\left(U_\alpha\right) \\ U_\gamma &= \textstyle\bigcup_{\alpha<\gamma} U_\alpha, \text{ for limit ordinals } \gamma \end{aligned}$$

This intuitive and deeply satisfying conception of the structure of the set theoretic universe yields a compelling explanation of Russell's Paradox: a plurality safely constitutes a set if and only if it is *bounded* in the full cumulative hierarchy, that is, if and only if there is a level of the hierarchy at or before which the plurality "runs out", that is, more formally: a level U_α such that, for everything x in that plurality, $\lambda(x)$ is no higher than U_α, in which case those things are "available" to be collected into a set at the next level $U_{\alpha+1}$. However, some predicates—notably, '$x \notin x$'—pick out *absolutely unbounded* pluralities, pluralities that never "run out" by any level[18]

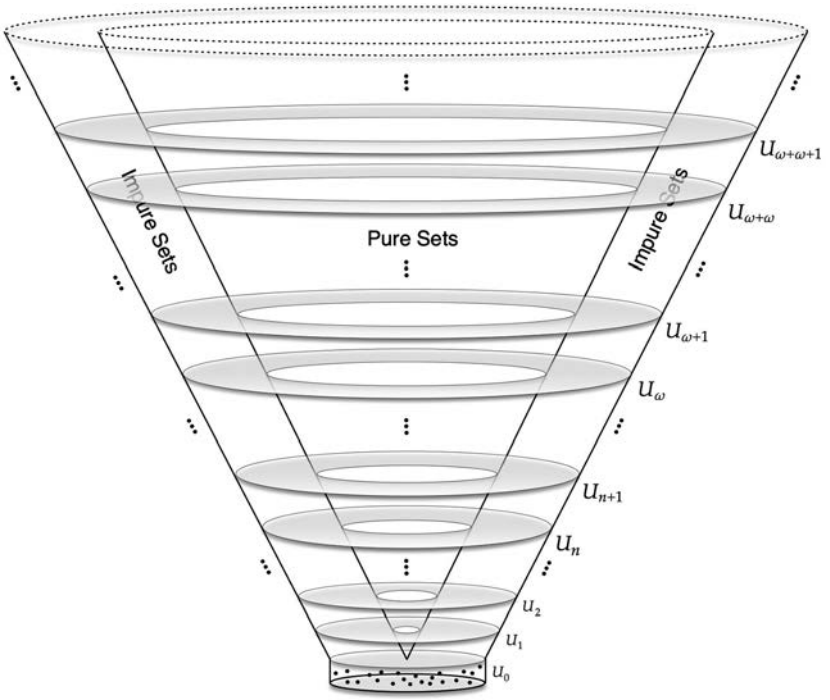


Figure 26.2 The transfinite cumulative hierarchy

and, hence, never constitute sets. Accordingly, the assumption that they do, as in Russell's Paradox, leads to contradiction.

26.3 Modal set theory and the completion problem

Unfortunately, as satisfying and illuminating as this explanation might be, a serious puzzle remains for the realist who wants to take the existence of sets seriously: the full cumulative hierarchy is itself a well-defined plurality; why is it not safe? That is, why does the hierarchy itself fail to be "completed" so as to constitute a set? Note the question is not: Why is there no universal set, i.e., no set containing all the urelements and all the sets? As we've just seen, the iterative conception of set provides a cogent answer to that question: only those pluralities that "run out" by some level of the cumulative hierarchy constitute sets at the next level and, obviously, the entire hierarchy is not such a plurality; there is no level at which the members of all the levels form a set. Rather, the question is: Why is the hierarchy only as "high" as it is? Why do all the urelements and sets that there actually are fail to constitute a *further* level that kicks off yet another series of iterations? For the same justification we had for "completing" the hierarchy of finite levels appears to apply no less to the full hierarchy: all of the urelements and sets in all the levels are *there*, as robustly as the members of the finite levels; moreover, since the set they would constitute—call it U_{Ω}—would be of a higher level than its members and, hence, would not be a member of itself, just as in the case of U_{ω}, no obvious threat of paradox looms. Call this *the completion problem*.

Granted, the completion problem does not appear to be as grave and immediate a threat to the coherence of set theory as Russell's Paradox. But it does raise disturbing questions for the

realist: if there is no answer to the completion problem, then there is an essential element of randomness to set existence. For once we acknowledge that there are pluralities that inexplicably fail to constitute sets, it is hard to see what grounds there are for picking and choosing between those that do and those that don't: in particular, if the same reasons for accepting that the finite levels U_0, U_1, U_2, … constitute a set seem to hold for all the levels of the hierarchy without their constituting a set, then what reason to do we have for accepting that even the finite levels U_0, U_1, U_2, … do? Or that *any* plurality does, for that matter? Without a solution to the completion problem, then, the *actual* structure of the hierarchy appears to be unknowable; any claims to knowledge of it would appear to be groundless, as the objects of the purported knowledge might well concern entities that simply do not exist.

Putnam (1967) was the first to argue explicitly that such questions are answered by taking the principles underlying the iterative conception of set to be essentially *modal* and, more specifically, by suggesting that a set is not to be understood in terms of the actual existence of a finished thing but as the *possibility* of its formation (p. 12):[19]

> [T]here is not, from a mathematical point of view, any significant difference between the assertion that *there exists a set of integers* satisfying an arithmetical condition and the assertion that *it is possible to select* integers so as to satisfy the condition. Sets, [to parody] John Stuart Mill, are permanent possibilities of selection.[20]

Parsons (1977) spells the idea out a little less metaphorically in a thesis—call it *Parsons's Principle*—that addresses the completion problem directly: any given plurality of things "*can* constitute a set, but it is not necessary that they do". Thus, necessarily, no matter how many cumulative levels there might be, the *absolutely unbounded* pluralities that don't in fact constitute sets in any level nonetheless *could have* constituted sets. The answer to the completion problem on this *potentialist* conception of sets, then, is simply that there neither is nor could be a "completed" cumulative hierarchy. Rather, instead of the completed stages of the cumulative hierarchy, we have a *potential* hierarchy, i.e., roughly speaking, an infinite hierarchy of *possibilities* where, given any possible completion of the hierarchy up to a given level U_α, there is always a more expansive possibility in which some of the mere pluralities of U_α constitute sets—in the "maximal" case, a possibility comprising the entire next level $U_{\alpha+1}$. The completion problem only arises on the assumption that all the levels—hence all the sets—that there could be (relative to an initial set U_0 of urelements) are already *actual* and, hence, that the hierarchy of sets is complete, that there is no more "collecting" of pluralities into sets that can be done. For only under that assumption—call it *actualism*—is it mysterious why the hierarchy is only as high as it is, why it (or indeed any absolutely unbounded plurality) fails to constitute a further set. The potentialist rejects the actualist assumption: the unbounded pluralities of one possible world always constitute sets in further, more comprehensive worlds.

What becomes of ZF on the potentialist conception? Thought of semantically, the potentialist conception suggests (roughly put) that an assertion to the effect that a certain set exists—and hence occurs at some level of the cumulative hierarchy—should be understood as an assertion that it is *possible* that such a set exists; likewise, assertions about all sets should be understood, not simply as assertions about the sets that *in fact* exist but, roughly speaking, about all the sets there could be, all the sets in any possible world.[21] Formalized, this insight yields what Linnebo (2013) calls the *potentialist translation* $\varphi^\Diamond$ of a sentence φ of ordinary set theory, viz., the result of replacing every existential quantifier occurrence $\exists$ in φ with its modalized counterpart $\Diamond\exists$ and every universal quantifier occurrence $\forall$ with $\Box\forall$. The idea, then, is that, if a statement φ of ordinary ZF set theory is purportedly true in the cumulative hierarchy, its modalized counterpart $\varphi^\Diamond$ will

be true in the potential hierarchy. This in turn might suggest that modal set theory will simply consist in formulating the theory $ZF^{\Diamond}$ that results from replacing the axioms of ZF with their potentialist translations. But that would not in and of itself be terribly interesting. Instead, modal set theorists like Parsons and Linnebo opt for the far more illuminating tack of taking the potentialist conception itself as primary and, hence, by axiomatizing its fundamental principles, *deriving* the axioms of $ZF^{\Diamond}$.

The first task toward that end is to identify the right propositional modal logic for the potentialist conception. Expressed in terms of possible worlds, the basic underlying intuition is that the universe of sets in a given world w can always be increased—for any world w there is an accessible world w' that includes not only everything already in w, but new sets whose members are mere pluralities in w.[22] This can be captured more formally by means of several intuitive constraints on world domains $D(w)$ and the accessibility relation R. Specifically, growth is reflected in the constraint that, if Rww', then $D(w) \subseteq D(w')$.[23] R itself should be a partial order (i.e., reflexive, transitive, and anti-symmetric)—each world w' accessible from a world w represents a way in which some of the mere pluralities of w constitute fully fledged sets in w', but different pluralities of w might constitute sets in different accessible worlds. Moreover, that the formation of new sets proceeds discretely is reflected in the requirement that R be weakly well-founded.[24] Finally, if w_1 and w_2 represent distinct expansions of the set theoretic universe of a world w—different "choices" of which mere pluralities of w to take to constitute sets—it should still be possible in each world that the sets constituted in the other exist; for, given the aforementioned constraint on domains, the pluralities that constitute sets in w_1 will still be available in w_2 and vice versa. Hence, a further natural condition is that accessibility be *directed*, that is, that for any two worlds w_1, w_2 accessible from a given world w, there is a third world accessible from both and, hence, one whose sets include all those formed in either world. The propositional modal logic determined by these conditions on accessibility is S4.2, the normal modal logic that includes the familiar axioms of the logic S4, viz.,

$$\mathbf{T} \quad \Box\varphi \rightarrow \varphi$$
$$\mathbf{4} \quad \Box\varphi \rightarrow \Box\Box\varphi$$

corresponding to reflexivity and transitivity, respectively, and the axiom

$$\mathbf{G} \quad \Diamond\Box\varphi \rightarrow \Box\Diamond\varphi$$

corresponding to directedness.[25] And to S4.2 is added classical quantification theory with identity and the axiom '$x \neq y \rightarrow \Box x \neq y$' expressing the necessity of difference.[26]

Now, as seen earlier, it is useful to express the axioms of ZF informally in terms of pluralities. To characterize the potentialist conception properly with the tools of modal logic, it is essential to quantify over them explicitly in order to identify the logical principles that govern their behavior. Accordingly, we introduce plural variables 'xx', 'yy', etc. and a new binary predicate $\prec$, where the formula '$y \prec xx$' indicates that y is one of the things xx. The inference rules for plural quantifiers parallel those for first-order quantifiers exactly.[27]

Several principles capture the existence and nature of pluralities. First, given the ontological innocence of plural quantification, the plural counterpart to **NC** is harmless:

$$\mathbf{PC} \quad \exists xx \forall y \big(y \prec xx \leftrightarrow \varphi(y)\big), \text{ where '}xx\text{' does not occur free in } \varphi(y),$$

that is, simply put, for any predicate $\varphi(y)$, there are the things it is true of.[28] Next, the modal properties of pluralities are captured in two axioms expressing the *stability* of the $\prec$ relation, that is, that, for any plurality xx and object y in a given world w, y will be among the things xx in an arbitrary accessible world w' if and only if it is among them in w:

$$\mathbf{Stb}^{+}_{\prec} \qquad y \prec xx \rightarrow \Box\big(y \prec xx\big)$$

$$\mathbf{Stb}^{-}_{\prec} \qquad y \not\prec xx \rightarrow \Box\big(y \not\prec xx\big).$$

However, these axioms don't rule out the possibility that xx *grows* in a further world w', that xx includes a new object z that only first comes to exist in w'. This possibility is ruled out by means of a schema that ensures that pluralities are *inextensible*:[29]

$$\mathbf{InEx}_{\prec} \qquad \forall y\big(y \prec xx \rightarrow \Box\varphi\big) \rightarrow \Box\forall y\big(y \prec xx \rightarrow \varphi\big).$$

Let *MPFO* be this system of plural first-order modal logic.

The next task is to extend MPFO to a basic modal set theory BMST. Like ZF, BMST axiomatizes the two fundamental structural properties of sets, viz., extensionality and foundation,[30] which are captured simply by adopting the axioms **Ext** and **Fnd** (hence also their necessitations). Recall from Section 26.1 in this chapter, however, that sets also have their members essentially, as expressed in the principles $\mathbf{E}\in$ and $\mathbf{E}\notin$.[31] Both of these principles, as well as the inextensibility of membership, are entailed by the following:

$$\mathbf{ED}_{\in} \qquad \exists xx\Box\forall y\big(y \prec xx \leftrightarrow y \in a\big).$$

Together with the stability and inextensibility principles for pluralities above, $\mathbf{ED}_{\in}$ says that, for any set a, one and the same plurality of things constitute a in every (accessible) possible world.

Recall that the general intuitive motivation for the naïve comprehension principle **NC** was that the existence of the members of a purported set should suffice for the existence of the set, which we can now express generally, and formally, in terms of plural quantification:

$$\mathbf{GNC} \qquad \forall xx\exists a\forall y\big(y \in a \leftrightarrow y \prec xx\big).$$

Given the innocent principle **PC**, **NC** (and catastrophe) follow immediately from **GNC**. As we've seen, both actualist and potentialist accounts have explanations of exactly where **GNC** goes wrong based intuitively on the iterative conception of set. On both accounts, in one sense or another, sets are "constructed" level by level without end, the new sets of one level constituted by the mere pluralities of previous levels. Hence, necessarily, the plurality of *all* the sets in all the levels, and unbounded pluralities generally, fail to constitute sets. Both accounts thus agree on where **GNC** gets it wrong: necessarily, there are pluralities that don't constitute sets. Additionally, however, the potentialist conception avoids the completion problem by explaining where **GNC** *almost* gets it right: not every plurality constitutes a set, but any plurality *could*. From the potentialist standpoint, then, **GNC** simply missed the implicit modality in claims of set existence; what we need is just its potentialist translation:

$$\mathbf{C} \qquad \Box\forall xx\Diamond\exists a\Box\forall y\big(y \in a \leftrightarrow y \prec xx\big).$$

That is, necessarily, any plurality of things *could* constitute a set consisting of exactly those things. This, of course, is a precise formalization of Parsons's Principle. Let BMST be the result of adding the axioms **Ext**, **Fnd**, $\mathbf{ED}_{\in}$, and **C** to MPFO.

It is a simple matter to show that BMST proves the potentialist translations of all of the ZF axioms except **Pow**, **Inf**, and **Rep**.[32] As these are by far the most powerful axioms of ZF, this shows that our basic intuitions about pluralities and sets as expressed in BMST—in particular, Parsons's Principle **C**—only get us so far. To see this with regard to **Pow**, suppose we have a plurality xx in some world w. By principle **C** there is a world u accessible from w in which they constitute a set b. Obviously, all the "subpluralities" of xx—all the pluralities $\gamma\gamma$ such that each thing in them is among the things xx—also exist in w. But neither **C** nor the structure of worlds and their domains in the underlying semantics provides any guarantee that any of them (other than xx) constitute a set in u, let alone all of them. Indeed, for all we know on the potentialist conception to this point, the subpluralities of xx might be *inexhaustible* in the sense that, for any world u accessible from w, there is always a further world v accessible from u at which some subplurality of xx only first constitutes a set (hence, a subset of b). If so, there is no world where all possible subsets of b exist, in which case the power set of b is impossible, contrary to $\mathbf{Pow}^{\Diamond}$, i.e., $\Diamond\exists a\Box\forall x(x \in a \leftrightarrow x \subseteq b)$.

To derive $\mathbf{Pow}^{\Diamond}$, then, an additional principle is required that rules out this sort inexhaustibility. Intuitively, this is accomplished most naturally by assuming that worlds more directly reflect the levels of the cumulative hierarchy; that is, by assuming, not only that the newly formed sets of a given world are mere pluralities of some preceding world, but that set formation is always *maximal*: that the newly formed sets of a world are *all* those that can be constituted from the mere pluralities of a preceding world.[33] This assumption can be expressed elegantly in a single axiom to the effect that, much like its members, the subsets of a given set a are constant across possible worlds:

$$\mathbf{ED}_{\subseteq} \qquad \exists xx\Box\forall\gamma\big(\gamma \prec xx \leftrightarrow \gamma \subseteq a\big).$$

$\mathbf{Pow}^{\Diamond}$ now follows straightaway from BMST + $\mathbf{ED}_{\subseteq}$.

Recall that **ES** and **Pr** alone suffice to generate the infinite plurality of Zermelo numbers $\varnothing_0$, $\varnothing_1$, $\varnothing_2$, … and it was left to **Inf** simply to sanction a set containing them. If we could prove the mere possibility that all the Zermelo numbers exist on the potentialist conception, we could immediately invoke principle **C** to prove the possible existence of a set containing them. However, the potentialist is in a slightly more fraught situation. For the potentialist principles to this point—the derived principles $\mathbf{ES}^{\Diamond}$ and $\mathbf{Pr}^{\Diamond}$ in particular—only yield a series of possible *initial segments* of the Zermelo numbers: by $\mathbf{ES}^{\Diamond}$ it is possible that $\varnothing_1$ exists; and by iterated applications of $\mathbf{Pr}^{\Diamond}$, it is possible that $\varnothing_0$ and $\varnothing_1$ exist and hence also that the Zermelo numbers $\varnothing_0$, $\varnothing_1$, and $\varnothing_2$ exist, and so on, but without the entire series of Zermelo numbers ever being "completed" in a single possibility.

To derive the possible existence of all the Zermelo numbers thus requires a further principle asserting, roughly, that whatever is true of the potential hierarchy as a whole, as expressed in the potentialist translation $\varphi^{\Diamond}$ of some proposition φ of set theory (hence a proposition containing no plural quantifiers[34]), is possible *simpliciter*:

$$\mathbf{Ref} \qquad \varphi^{\Diamond} \rightarrow \Diamond\varphi.$$

Thus, in particular, $\mathbf{ES}^{\Diamond}$ and $\mathbf{Pr}^{\Diamond}$ yield a proposition $\varphi^{\Diamond}$ expressing the infinite series of possibilities involving larger and larger initial finite segments of the Zermelo numbers.[35] By **Ref** that

series is reflected in a single possibility containing all—hence, by **PC**, the entire infinite plurality of—the Zermelo numbers, and so by Parsons's Principle **C**, it is possible that they constitute a single infinite set. $\mathbf{Inf}^{\Diamond}$ follows immediately by some simple modal logic. The potentialist translation $\mathbf{Rep}^{\Diamond}$ of the Replacement schema can be similarly proved by strengthening **Ref** to

$$\mathbf{Ref}^{+} \qquad \Diamond\forall x\left(\varphi^{\Diamond}(x) \rightarrow \varphi(x)\right).^{36}$$

Importantly, it can be shown that the modalized quantifiers $\Diamond\exists$ and $\Box\forall$ "behave proof-theoretically very much like ordinary quantifiers" (Linnebo, 2013, p. 213), thus explaining why they are not found in ordinary set-theoretic practice—mathematicians can, in effect, talk about the potentialist hierarchy *as if* it were actual.

26.4 Concluding philosophical postscript

This chapter has focused chiefly on the technical development of (one prominent variety of) modal set theory and its intuitive motivations without any close critical attention paid to surrounding philosophical questions. In closing, we note briefly that a problem parallel to the completion problem threatens to arise for the potentialist. To see this, first, instead of restricting ourselves, relative to any given possible world, only to those possibilities that represent growth of the set theoretic universe, as we do in characterizing the potentialist hierarchy, let us broaden our metaphysical perspective to one where we are considering all possibilities on a par. From this standpoint, we see that, on the potentialist conception, at least the vast majority of pure sets are metaphysically contingent beings—in particular, for any given possible world w, those pure sets in any accessible world w' whose members, jointly, are mere pluralities in w. If this is in fact the sober truth about the metaphysics of sets, then set existence is modally capricious—two possible worlds can be in all respects identical but for the fact that there are pure sets in one that simply and inexplicably fail to exist in the other. Though not identical to the completion problem, the apparently inexplicable contingency of set existence on the potentialist conception, taken literally, seems to raise questions parallel to those arising from the apparently inexplicable nonexistence of certain sets, as noted in the completion problem.

Perhaps in response to this difficulty—though neither explicitly says so—both Parsons (1977, §IV) and Linnebo (2013, pp. 207–8) suggest that the modality of the potential hierarchy is more semantic than metaphysical: at any given time, one's conception of the "height" of the set theoretic universe, hence the range of one's quantifiers, is determined by one's strongest large cardinal assumptions.[37] Once convinced of the existence of a larger cardinal still, pluralities that (relative to the earlier conception) had been absolutely unbounded constitute sets under the stronger assumptions and the range of one's quantifiers broadens accordingly. Thus, Linnebo (2013, p. 208):

> A claim is possible, in this sense, if it can be made to hold by a permissible extension of the mathematical ontology; and it is necessary if it holds under any permissible such extension. Metaphysical modality would be unsuitable for our present purposes because pure sets are taken to exist of metaphysical necessity if at all.

However, if after all (pure) sets exist as a matter of metaphysical necessity (so that, in particular, any pure sets that *could have* existed *actually* exist), as Linnebo appears to suggest here, then the completion problem threatens once again to rear its head with all its original force: why are

there only the pure sets there actually are? If, necessarily, all pure sets exist of metaphysical necessity, what explains the fact that there *couldn't* have been more, the fact that there are pluralities of things such that it is not even metaphysically possible that they constitute sets?[38]

These and related metaphysical questions prompted by the potentialist conception of set point to a fertile area for exploration in the philosophical foundations modal set theory.

Notes

1 The three principles above are still jointly consistent with the possibility of sets a that contain members x that cannot coexist with a. This (rather perverse) possibility can be ruled out by replacing **OD** with a principle asserting that the membership relation is *existence entailing*: $\Box\forall a\Box\forall x\Box(x \in a \rightarrow (E!a \wedge E!x))$. Given **E∈**, both **E∉** and **OD** follow (assuming the propositional modal logic S5).

2 See Menzel (2014a, §2) for a formal proof and a discussion of the surrounding philosophical issues.

3 The system Q of Prior (1968) is among the most severely restricted of well-known systems, abandoning in particular the interdefinability of $\Box$ and $\Diamond$ and most familiar principles of propositional modal logic. For examples of less severely restricted systems, see Kripke (1963), Menzel (1991), and Fine (1978).

4 The term "necessitism" and its cognates was coined in Williamson (2010), although the view was in large measure anticipated and developed in detail in Linsky and Zalta (1994). See also Salmon (1987) for an influential precursor.

5 Van Cleve (1985) is a fine exploration of the issues raised in this section. See also Fine (1981) for a detailed and rather more formal study.

6 The exposition in this section and portions of the following is similar to that found in Sections 26.3–26.4 of Menzel (2018), which was written largely in parallel with the current entry.

7 See Cantini (2014) for an excellent overview.

8 Translated as Zermelo (1967); see also the informative introductory note by Felgner that accompanies the translation of this paper in the polyglot edition (Zermelo, 2010) of Zermelo's collected works, pp. 160–89. Zermelo's theory included the important but controversial axiom of *Choice*, though it will play no part here.

9 We will make free use of the common $\{x_1,\ldots,x_n\}$ notation for finite sets without defining it formally.

10 The existence of U is not entirely unproblematic, as it could turn out to be inconsistent with the other axioms if there are "too many" urelements. See, e.g., Nolan (1996) and Menzel (2014b). Zermelo himself wasn't sure how to work urelements into his theory until over two decades later; see Zermelo (1930) and Kanamori's informative introductory note to its translation in Zermelo (2010), pp. 390–430.

11 **Sep** renders **ES** otiose, since it is a truth of (classical first-order) logic that something x exists, $\exists x E!\, x$, from which **ES** follows directly from **Pr** and the instance of **Sep** where $\varphi(x)$ is $x \neq x$.

12 Zermelo did not include Foundation in his 1908 axiomatization but, as non-well-founded sets were not defined and studied in any systematic way until Mirimanoff (1917), and the iterative conception was at most only beginning to take shape in Zermelo's mind, it seems likely that he did not at the time recognize any pressing need for the axiom.

13 Specifically, by letting b be the set given by **Inf** and letting $\varphi(x)$ be $\forall\, y[(\varnothing \in y \wedge\ \forall\, z(z \in y \rightarrow \{z\} \in y)) \rightarrow x \in y]$, i.e., the predicate "$x$ is in every set y that contains $\varnothing$ and the singleton of any of its members".

14 To see this, very briefly: where V_ω is the set of hereditarily finite pure sets, let $W_o = V_\omega \cup U$ and $W_{n+1} = W_n \cup \wp(W_n)$, for $n \in \omega$, and let $W = \bigcup_{n\in\omega} W_n$. It is easy to see that W is a model of Z and that $U_i \in W$ for all $i \in \omega$ but that $\{U_0,U_1,U_2,\ldots\} \notin W$. The author thanks Noah Schweber and Joel David Hamkins for this construction.

15 Skolem independently identified the need for Replacement, and his explicitly first-order formulation of the principle is essentially the one that is mostly used today. See Fraenkel (1922) and Skolem (1922); an English translation of the latter can be found in van Heijenoort (1967, pp. 290–301).

16 $L(y)$ is definable without any mention of finitude as: $\forall a[(U \in a \wedge\ \forall\, b(b \in a \rightarrow b \cup\ \wp(b) \in a)) \rightarrow y \in a]$, i.e., "$y$ is in every set that contains the set U of urelements and also contains $b \cup\ \wp(b)$ whenever it contains b, for any set b".

17 ω is the first transfinite ordinal number, the first "counting number" after the natural numbers. It is also the first *limit* ordinal, i.e., the first ordinal $\alpha > 0$ such that, if $\beta < \alpha$, then $\beta + 1 < \alpha$. See Devlin (1991, ch. 3) for a good introduction to transfinite ordinals and cardinals and their arithmetic.

18 In the cumulative hierarchy, where all sets are well-founded, no set is a member of itself so '$x \notin x$' is in fact true of everything.

19 See Hellman (1989) for a detailed development of Putnam's ideas and Roberts (2019) for a rigorous critique of Hellman.

20 There are intimations of Putnam's idea in Cantor, notably in an 1897 letter to Hilbert : "I say of a set that it can be thought of as *finished* … if it is possible without contradiction (as can be done with finite sets) to think of *all its elements as existing together*, and so to think of the set itself as *a compounded thing for itself* ; or (in other words) if it is *possible* to imagine the set as *actually existing* with the totality of its elements" (Ewald, 1996, p. 927).

21 My exposition from this point draws heavily upon Linnebo (2010, 2013). This work in turn was strongly influenced by Parsons (1983).

22 For fairly obvious reasons, possible world semantics can only be considered a useful heuristic, as Kripke models are themselves by definition *sets*. Hence, no Kripke model could purport to represent (in a one-to-one fashion) *all possible* sets without opening the door once again to Russell's Paradox. This does, however, raise the further question of what it means for the potentialist translation of a theorem of ZF to be *true*, if it is not true in virtue of facts about genuinely existing sets in genuinely existing possible worlds.

23 This will of course have the consequence that, for any object x, including urelements, it will be a logical truth that x exists necessarily, $\Box E!x$. That is the right result, as the constraints in question are only meant intuitively to capture possible ways of *extending* the set theoretic universe of a given world w; nothing in w "goes away" as new sets are added in accessible worlds.

24 R is weakly well-founded if every nonempty set S of worlds contains an R-minimal world, i.e., a world $w \in S$ such that, for all other $u \in S$, not-Ruw. (R is well-founded if it is a strict partial order — irreflexive and transitive — that satisfies this definition.)

25 There is no axiom corresponding to antisymmetry because it is not definable in propositional modal logic. See, e.g., Blackburn, de Rijke, and Venema (2001, §4.5). Well-foundedness can be axiomatized in a bimodal extension of propositional modal logic with both "forward-looking" and "backward-looking" necessity operators. See Studd (2013).

26 The necessity of identity is a well-known theorem of this logic.

27 See Linnebo (2014) for a useful overview of the logic of plural quantification and Oliver and Smiley (2016) for a comprehensive treatment of plural logic.

28 As Linnebo points out (2013, p. 210), **PC** entails the existence of an "empty" plurality. This can be avoided at the cost of some inconvenience. As to **PC**'s harmlessness, note that a Russell-style argument to a contradiction from **PC** with φ = '$\gamma \prec \gamma$' (assuming that construction is even permitted syntactically) breaks down, as it is in general invalid to infer from the fact that all things (distributively) have a property F, that any things (collectively) have F. Hence, from $\forall\gamma\varphi(\gamma)$ we cannot legitimately infer $\varphi(xx)$; in particular, we can't infer $xx \prec xx \leftrightarrow xx \not\prec xx$ from $\forall\gamma(\gamma \prec xx \leftrightarrow \gamma \not\prec \gamma)$. For more on distributive vs. collective quantification, see, e.g., Scha (1984) and McKay (2006).

29 To understand how this works to ensure inextensibility it is useful to note that $\mathbf{InEx}_{\prec}$ is just the Barcan formula (BF)—$\forall\gamma\Box\varphi \rightarrow \Box\forall\gamma\varphi$—restricted to pluralities xx. The semantic effect of unrestricted BF is that, so to say, *existence* is inextensible, i.e., that there couldn't have been anything other than the things that just happen to exist in fact, a notoriously problematic metaphysical claim (see Menzel, 2014a, esp. Sections 1 and 2). For suppose there could have been something a distinct from everything that exists in fact. Then $\forall\gamma\Box(\gamma \neq a)$. But then by BF, $\Box\forall\gamma(\gamma \neq a)$, contradicting the assumption that such an a could have existed. Restricted as in $\mathbf{InEx}_{\prec}$, however, the semantic effect is simply that, for a given plurality xx, there couldn't have been anything a among *them* that isn't already among them, i.e., that *pluralities* are inextensible.

30 Foundation follows semantically from the weak well-foundedness of the accessibility relation R and the fundamental potentialist assumption that the members of any set that first comes to be in any world w is constituted by objects that exist in some world u such that Ruw. See Linnebo (2013, pp. 216–17).

31 Since $\Box E!a$ is a simple theorem of MPFO, the antecedents $E!a$ of the conditionals in the consequents of these two principles are superfluous in the context of BMST.

32 See Linnebo (2013, p. 221) and Parsons (1983, Appendix 2). The addition of the urelement axiom **Ur**, which neither Linnebo nor Parsons includes in their discussions, would involve no significant complications to these results.

33 Linnebo (2013, p. 209 n. 7) suggests that maximality entails that the accessibility relation is linear but that would follow only if worlds are individuated extensionally by their domains (which is plausible if, as in Linnebo (2013), the existence of urelements is not assumed). Linearity can be forced axiomatically by any of a variety of axioms; see Chellas (1980). The system resulting from adding a linearity axiom to S4 in place of **G** is S4.3 though, as Parsons notes (1983, pp. 319–20), S4.3 seems to add nothing over S4.2 for the purpose of deriving set-theoretic consequences.

34 Allowing plural quantifiers in φ without restriction would in fact lead to Russell's Paradox.

35 Specifically, $\varphi^{\Diamond}$ is $\Diamond\exists a\forall x(x \not\in a) \wedge \Box\forall x\Diamond\exists y\Box\forall z(z \in y \leftrightarrow z = x)$, and hence $\Diamond\varphi$ is $\Diamond(\exists a\forall x(x \not\in a) \wedge \forall x\exists y\forall z(z \in y \leftrightarrow z = x))$.

36 Again, where φ contains no plural quantifiers. Parsons (1983, p. 323) relies on a rather different principle to derive $\mathbf{Inf}^{\Diamond}$ and $\mathbf{Rep}^{\Diamond}$. See Linnebo (2013, p. 223 n. 28) for further discussion.

37 See also Fine (2006).

38 Hewitt (2015, §§2.4–2.6) raises related problems.

References

Blackburn, P., de Rijke, M., & Venema, Y. (2001). *Modal Logic*. Cambridge: Cambridge University Press.

Cantini, A. (2014). Paradoxes and Contemporary Logic. In E. N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy*. URL: https://plato.stanford.edu/archives/fall2014/entries/paradoxes-contemporary-logic.

Chellas, B. F. (1980). *Modal Logic: An Introduction*. Cambridge: Cambridge University Press.

Devlin, K. (1991). *The Joy of Sets* (2nd ed.). New York: Springer-Verlag.

Ewald, W. (Ed.). (1996). *From Kant to Hilbert: A Source Book in the Foundations of Mathematics* (Vol. II) Oxford: Clarendon Press.

Fine, K. (1978). Model Theory for Modal Logic, Part I – The De Re/De Dicto Distinction. *Journal of Philosophical Logic*, 7(2), 125–156.

Fine, K. (1981). First-Order Modal Theories—Sets. *Nous*, 15(2), 177–205.

Fine, K. (2006). Relatively Unrestricted Quantification. In A. Rayo & G. Uzquiano (Eds.), *Absolute Generality* (pp. 20–44). Oxford: Oxford Clarendon Press.

Fraenkel, A. (1922). Zu den Grundlagen der Cantor-Zermeloschen Mengenlehre. *Mathematische Annalen*, 86, 230–237.

van Heijenoort, J. (Ed.). (1967). *From Frege to Gödel: A Source Book in Mathematical Logic, 1879–1931*. Cambridge, MA: Harvard University Press.

Hellman, G. (1989). *Mathematics Without Numbers: Towards a Modal-Structural Interpretation*. Oxford: Oxford University Press.

Hewitt, S. (2015). When Do Some Things Form a Set? *Philosophia Mathematica*, 23(3), 311–337.

Kripke, S. (1963). Semantical Considerations on Modal Logic. *Acta Philosophica Fennica*, 24, 83–94.

Linnebo, Ø. (2010). Pluralities and Sets. *The Journal of Philosophy*, 107(3), 144–164.

Linnebo, Ø. (2013). The Potential Hierarchy of Sets. *The Review of Symbolic Logic*, 6(2), 205–228.

Linnebo, Ø. (2014). Plural Quantification. In E. N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy* (Fall 2014 Edition). Stanford, CA: Stanford University.

Linsky, B., & Zalta, E. N. (1994). In Defense of the Simplest Quantified Modal Logic. In J. Tomberlin (Ed.), *Philosophical Perspectives 8: Logic and Language* (pp. 431–458). Atascadero: Ridgeview Publishing Co.

McKay, T. J. (2006). *Plural Predication*. Oxford: Clarendon Press.

Menzel, C. (1991). The True Modal Logic. *Journal of Philosophical Logic*, 20(4), 331–374.

Menzel, C. (2014a). Actualism. In E. N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy* (Summer 2014 Edition). Stanford, CA: Stanford University.

Menzel, C. (2014b). Wide Sets, ZFCU, and the Iterative Conception. *Journal of Philosophy*, 111(2), 57–83.

Menzel, C. (2018). The Argument from Collections. In T. Dougherty & J. Walls (Eds.), *Two Dozen or so Arguments for God: The Plantinga Project*. Oxford: Oxford University Press.

Mirimanoff, D. (1917). Les antinomies de Russell et de Burali-Forti et le problème fondamental de la théorie des ensembles. *L'Enseignement Mathematique*, 19, 37–52.

Nolan, D. (1996). Recombination Unbound. *Philosophical Studies*, 84(2/3), 239–262.

Oliver A., and T. Smiley (2016). *Plural Logic* (2nd enlarged ed.). Oxford: Oxford University Press.
Parsons, C. (1977). What is the Iterative Conception of Set? In R. E. Butts & J. Hintikka (Eds.), *Logic, Foundations of Mathematics, and Computability Theory* (pp. 335–367). DordrechtSpringer.
Parsons, C. (1983). Sets and Modality. In C. Parsons (Ed.), *Mathematics in Philosophy: Selected Essays* (pp. 298–341). Ithaca: Cornell University Press.
Prior, A. N. (1968). Tense Logic for Non-Permanent Existents. In A. N. Prior (Ed.), *Papers on Time and Tense* (pp. 145–160). Oxford: Oxford University Press.
Putnam, H. (1967). Mathematics without Foundations. *Journal of Philosophy*, 64(1), 5–22.
Roberts, S. (2019). Modal Structuralism and Reflection. *Review of Symbolic Logic*, 12(4), 823–860.
Salmon, N. (1987). Existence. In J. E. Tomberlin (Ed.), *Philosophical Perspectives* (Vol. 1, pp. 49–108). Atacadero, CA: Ridgeview Publishing Company.
Scha, R. J. H. (1984). Distributive, Collective, and Cumulative Quantification. In J. A. G. Groenendijk, T. M. V. Janssen and M. B. J. Stokhof (Eds.), *Truth, Interpretation and Information* (pp. 131–158). Dordrecht: Foris.
Skolem, T. (1922). Einige Bemerkungen zur axiomatischen Begründung der Mengenlehre. In *Mathematikerkongressen I Helsingfors* (pp. 217–232). Helsinki: Akademiska bokhandeln.
Studd, J. P. (2013). The Iterative Conception of Set—A (Bi)-Modal Axiomatisation. *Journal of Philosophical Logic*, 42, 697–725.
Van Cleve, J. (1985). Why a Set Contains Its Members Essentially. *Nous*, 19(4), 585–602.
Williamson, T. (2010). Necessitism, Contingentism, and Plural Quantification. *Mind*, 119(475), 657–748.
Zermelo, E. (1908). Untersuchungen über die Grundlagen der Mengenlehre, I. *Mathematische Annalen*, 65, 261–281.
Zermelo, E. (1930). Über Grenzzahlen und Mengenbereiche. *Fundamenta Mathematicae*, 16, 29
Zermelo, E. (1967). Investigations in the Foundations of Set Theory. In J. van Heijenoort (Ed.), *From Frege to Gödel: A Source Book in Mathematical Logic, 1879–1931* (pp. 199–215). Cambridge, MA: Harvard University Press.
Zermelo, E. (2010). *Collected Works*, Vol. I. (H.-D. Ebbinghaus, C. G. Fraser, & A. Kanamori, Eds.). New York: Springer.

Chapter 27

THE LOGIC OF METAPHYSICAL MODALITY

Bob Hale

27.1 Introduction

Modal words—adverbs such as 'necessarily' and 'possibly', modal adjectives such as 'necessary', 'possible' and 'contingent', and verbs such as 'must', 'may', 'might', 'can', 'could', etc.—can be used, both in philosophy and elsewhere, to express different sorts of necessity and possibility, or different kinds of modal claim. Besides expressing *alethic* modal claims about what must be, or might have been, *true*, modal words (and especially the modal verbs 'must', 'may' and 'might') may figure in *deontic* claims about what *ought* to be or *may* be done (what is morally obligatory or permissible), and in *epistemic* modal claims (e.g., 'He must have missed the train' and 'She may be coming by car' in their most likely uses). Here we shall be concerned exclusively with alethic modalities—that is, roughly, with modalities having to do with the kind or mode of truth attaching to propositions.

In their epistemic uses, modal words clearly involve a kind of relativity—roughly, what (epistemically) must (or may) be so is what is necessary (or possible), relative to what we know. At least some kinds of alethic modality are likewise plausibly, and commonly, taken to be relative. Thus, what is physically or, more generally, naturally necessary may be thought not to be necessary outright and without qualification, but to be what must be so, given the laws of physics or nature. By contrast, *logical* necessity, of which the various forms of relative necessity are usually taken to be relativizations, may be thought to be non-relative or absolute.[1]

One question we can ask about different kinds of necessity and possibility concerns their logic: to what logical principles do they conform? We might expect answers to vary, depending upon which kind of modality is in question. For example, if we are concerned with an alethic modality, we should expect necessity to entail truth, and truth to entail possibility. But with deontic modalities, matters are clearly otherwise: what ought to be so (i.e., what is deontically necessary) may very well not be so, and what is so may very well not be permissible (deontically possible). Further, we may expect differences between different alethic modalities. For example, if we take physical necessity to be necessity relative to the (actual) laws of physics, we may think that, absolutely speaking, those laws might have been different, so that what is physically necessary might not have been so. By contrast, if we take the laws of logic to be absolutely necessary, we shall think that what is logically necessary could not have failed to be so. Here, we focus on the question: what is the logic of *metaphysical* modality? The next two sections provide some

background explanations: first, of what is meant by metaphysical modality, and second, of the basic ideas of the logic of modality and alternative modal logics. We shall then present some arguments for the view that the logical principles governing metaphysical necessity and possibility are those of the strongest normal modal logic, the system known as *S5*. This is followed by discussion of some arguments for the opposing view, that the correct modal logic for metaphysical modality must be a weaker system.

27.2 Modal logics

In modal logic, the notions of necessity and possibility are expressed by the modal operators $\Box$ and $\Diamond$, which may be read as 'it is necessary that … ' or 'Necessarily …' and 'It is possible that …' or 'Possibly …' respectively. Like negation ('It is not the case that …'), they are sentence-forming operators on sentences. As such, they may be iterated so as to express complex modalities. Thus, as well as $\Box A$ and $\Diamond B$ for any given sentences A and B, we have well-formed sentences such as $\Box\Diamond A$, $\Box\Box B$, $\Box\Diamond\Box A$ and so on. Usually, $\Box$ and $\Diamond$ are taken to be interdefinable by one of the equivalences $\Diamond A \leftrightarrow \neg\Box\neg A$ and $\Box A \leftrightarrow \neg\Diamond\neg A$, which assert that a proposition is possible just in case its negation is not necessary and is necessary just in case its negation is not possible.

27.2.1 Standard deductive systems

A modal logic is obtained from an underlying non-modal logic, such as classical propositional logic or first-order quantification logic, by adding the modal operators to the formal language and some rules or axioms governing them.[2] Assuming the non-modal basis is formulated as an axiomatic system, and that $\Box$ is taken as primitive and $\Diamond$ defined as earlier, the very weak modal logic, *K*, is obtained by adding a single modal axiom ensuring that $\Box$ distributes across the conditional:

$$K \textit{ axiom}: \qquad \Box(A \rightarrow B) \rightarrow (\Box A \rightarrow \Box B)$$

together with the *Rule of Necessitation*:

$$RN: \qquad \text{From } \vdash A \text{ infer } \vdash \Box A$$

which allows us to necessitate A whenever it has been established as a theorem of the system (so that, for example, the necessitation of any theorem of the underlying non-modal logic is also a theorem of *K*).[3,4]

The weakness of *K* shows itself partly in the fact that we are not even able to prove as a theorem that if a proposition is necessary, then it is true (i.e., $\nvdash_K \Box A \rightarrow A$). This implication is known as the Law of Necessity. As noted in the Introduction, we would expect it to be a theorem of any *alethic* modal logic. It is a theorem of the system *T*, which extends *K* by adding it as an axiom.

While *T* is a better candidate to capture the logic of some alethic modalities, it is still a quite weak system. In particular, it settles no questions about iterated modalities, such as $\Box\Box$ and $\Diamond\Box$. Clearly $\vdash_T \Box\Box A \rightarrow \Box A$, since this is just a special case of the *T*-axiom, and it is almost immediate that $\vdash_T \Box A \rightarrow \Diamond\Box A$, since this is an instance of $A \rightarrow \Diamond A$, which is an easy consequence of the *T*-axiom.[5] But what of the converse implications, $\Box A \rightarrow \Box\Box A$ and $\Diamond\Box A \rightarrow \Box A$? Might what is necessary have failed to be necessary, or is it *necessarily* necessary? Might what is

only contingently true have been necessary, or is only what is necessary *possibly* necessary? Logic cannot by itself answer these questions. Neither is provable in *T*. What it can do is explore the consequences of accepting these principles, by developing stronger modal systems in which they hold. Adding the first principle, $\Box A \rightarrow \Box\Box A$ (or its equivalent $\Diamond\Diamond A \rightarrow \Diamond A$), to *T* as an additional axiom gives the system *S4*. Adding the second, $\Diamond\Box A \rightarrow \Box A$ (or its equivalent $\Diamond A \rightarrow \Box\Diamond A$), to *T* as an additional axiom gives the system *S5*.

S5 is stronger than *S4* in the sense that every theorem of *S4* is provable in *S5*, but not conversely. In particular, $\vdash_{S5} \Box A \rightarrow \Box\Box A$ but $\nvdash_{S4} \Diamond\Box A \rightarrow \Box A$. Thus, *S4* lies between *T* and *S5*. But it is not the only intermediate system of interest. The system *B* results from *T* by the addition of the Brouwersche principle $A \rightarrow \Box\Diamond A$ as an axiom, which is likewise a theorem of *S5* but not of *S4*.

If the various modal principles characteristic of *K*, *T*, *B*, *S4* or *S5* are added to an underlying system of first-order quantification theory, we obtain a system of quantified modal logic. The combination of modal operators with quantifiers gives rise to some intriguing questions. One which is especially interesting from a philosophical point of view concerns the principles $\forall x\Box A(x) \rightarrow \Box\forall x A(x)$ and its converse, $\Box\forall x A(x) \rightarrow \forall x\Box A(x)$, known as the Barcan and Converse Barcan Principles, respectively. If we assume the underlying non-modal quantification theory is classical, Converse Barcan is provable in quantified *K*, and Barcan in quantified *B* (and so, of course, in quantified *S5*, but not in quantified *S4*).[6]

27.2.2 World semantics

In the model-theoretic semantics for modal logics developed in the late 1950s and early 1960s, an interpretation is based on a set of indices—usually called '(possible) worlds'. Sentences are interpreted not as true or false *simpliciter*, but as true or false *relative to*, or *at*, a world.

For a *propositional* modal language (whose sentences are just sentence variables and compounds built from them by means of sentential operators, including $\Box$ and $\Diamond$), a model consists of a set *W* (usually thought of as a set of possible worlds) together with a binary relation *R* defined on *W* and an assignment of truth-values relative to worlds to the sentence variables.[7] Intuitively, Rw_iw_j says that w_j is *possible relative to* w_i. This is sometimes expressed: w_j is *accessible from* w_i. A model may include further constraints on the accessibility relation *R*, such as that it is to be reflexive, or symmetric, or transitive, etc. Truth-functional compounds are then assigned values relative to worlds by the obvious relativizations of the usual clauses for $\neg$, $\wedge$, etc.—for example, $\neg A$ is true at a world w_i iff *A* is false at w_i; $A \wedge B$ is true at w_i iff *A* is true at w_i and *B* is true at w_i, etc. Modal sentences are assigned values according to the rules: $\Box A$ is true at w_i iff *A* is true at every w_j such that Rw_iw_j, and $\Diamond A$ is true at w_i iff *A* is true at some w_j such that Rw_iw_j—necessity at a world is truth at all worlds possible relative to that world, and possibility at a world truth at some relatively possible world.

For a *first-order* modal language, things are inevitably a little more complicated. Basically, we need to assign a set of individuals to each world, from which the bound individual variables take their values, and to assign extensions to any predicate constants relative to each world. Further details need not concern us here.[8]

As usual, a sentence is *valid* iff it evaluates as true at every world in every admissible model. If no further constraints are imposed on *R*, the foregoing semantics validates the *K*-principle, but none of the axioms for *T*, *B*, *S4* or *S5*. Consider, for example, the *T*-axiom. Let $\mathfrak{M}$ be a model with just one world *w*, which does not access itself, so that there are no worlds accessible from *w*. Suppose that *A* is false at *w*. Then the *T*-axiom will be false in $\mathfrak{M}$, since $\Box A$ will be true but *A* false at *w*. $\Box A$ requires that *A* be true at every w' accessible from *w*—but this condition is

vacuously met, since there *are* no worlds accessible from w. We can similarly construct countermodels to the other axioms.

Requiring that each world is accessible to itself (i.e., that R be reflexive) suffices for the T axiom. Similarly, requiring that R be symmetric (i.e., that Rw_jw_i whenever Rw_iw_j) suffices for the B axiom, requiring transitivity (i.e., that Rw_iw_k whenever Rw_iw_j and Rw_jw_k) secures the *S4* axiom and requiring both symmetry and transitivity secures the *S5* axiom.[9]

27.3 Some arguments for *S5*

27.3.1 The logic of absolute necessity

As we saw, some kinds of modality are commonly taken to be relative. On the standard account,[10] relative necessity is analyzed in terms of logical necessity—where $\Box$ is read 'it is logically necessary that', 'It is physically necessary that A', for example, is defined as $\Box(\Phi \rightarrow A)$, where Φ is a conjunction of physical laws. Logical necessity is taken to be absolute, at least in the sense that it is not to be itself analyzed as a form of relative necessity. But it may well be thought to be absolute in another, and rather stronger, sense—in which to be absolutely necessary is to be necessary without qualification or completely unconditionally. It is not clear that this notion of absoluteness can be explained without circularity. In terms of worlds, the obvious characterization is that it is absolutely necessary that A iff A is true at absolutely every possible world, without restriction or exception—absolute necessity is explained by absolutely unrestricted quantification over possible worlds.[11,12]

Given this explanation, we can mount a quite simple argument for the conclusion that the modal logic of absolute necessity is *S5*. If it is to be absolutely necessary that A, it must be true that A at absolutely every possible world without restriction. In terms of admissible models, this means that if $\Box_{abs}A$ is to be true at a world w_j in a model $\mathfrak{M}$, A must be true at each world w_k in $\mathfrak{M}$'s set of worlds W—in particular, there must be no $w_k \in W$ inaccessible from w_j. For if there were such a world, $\Box A$ might be true at w_j even though there is a possible world at which A is false. But then $\Box A$ would be at best only relatively, not absolutely, necessary. But if each world in the set W of worlds in any model $\mathfrak{M}$ is accesible from every world in W, it follows that the accessibility relation R must be an equivalence relation, i.e., reflexive, symmetric and transitive. And that, as noted previously, suffices for the *S5* axioms.

If metaphysical necessity is a kind of absolute necessity, it follows that its modal logic is likewise *S5*. But is the antecedent of that conditional true?

One consideration in favor of taking it to be so is that Kripke, whose work has done more than any other to secure acceptance of a metaphysical notion of necessity, explicitly states that the kind of necessity he has in mind is 'necessity in the highest degree' (Kripke 1980: 99; see also p. 164).

Another argument relies on a certain conception of the relation between logical and metaphysical modality. On one view, metaphysical lies between logical and physical necessity, weaker than the first but stronger than the second. On another, metaphysical necessity is a kind of relative necessity. On another view still, logical necessity is a species of metaphysical necessity, so that while logical necessity is stronger than metaphysical in the purely technical sense that the former implies but is not implied by the latter, in another sense metaphysical necessities are every bit as necessary as logical necessities. In particular, if logical necessity is absolute, then so is metaphysical necessity in general. This last view is especially natural and plausible if, as suggested earlier, metaphysical necessity is explained as what holds true in virtue of the nature of things in general, and logical necessities are those things which are true in virtue of the natures of just the logical

entities—the various logical functions such as negation, conjunction, etc. For there would then be no reason to regard logical necessities as stronger than metaphysical necessities in general, except in the purely technical sense already noted—the difference would just be that logical necessities hold in virtue of the natures of just logical entities, whereas metaphysical necessities hold in virtue of the natures of things in general.[13]

27.3.2 Williamson's argument

If the argument just rehearsed is sound, it shows that the logic of metaphysical modality is *at least* as strong as *S5*, the strongest normal modal logic. It does *not* show that *S5* is *exactly* the right logic, since for all the argument shows, that might be some stronger, non-normal logic. An argument given by Timothy Williamson at least closes that gap, and if successful, provides an independent reason for taking *S5* to be the correct logic.

Williamson's argument focuses on propositional modal logic.[14] The question is: which modal logic captures all and only those truths which are *metaphysically universal*—those truths about metaphysical modality which are sufficiently general in the following sense. Let A be any formula of our propositional language. Then, by its universal generalization is meant the result of substituting distinct propositional variables for distinct atomic sentences throughout A, and prefixing universal quantifiers on each variable. For example, the universal generalization of $\Box p \rightarrow \Box\Box p$ is $\forall X(\Box X \rightarrow \Box\Box X)$. A is metaphysically universal iff its universal generalization is true on its intended interpretation, which treats the truth-functional operators as usual and interprets $\Box$ as metaphysical necessity, which Williamson glosses as 'what could not have been otherwise, what would have obtained whatever had obtained'.[15] The conclusion of the argument (q.v. Williamson 2013: 110–115) is that 'if possibility and necessity are non-contingent, then MU [the system which is sound and complete for metaphysical universality] is exactly the system *S5*'. Possibility and necessity are non-contingent iff whatever is possible or necessary is necessarily so, so that the antecedent amounts to assuming that the axioms $\Box A \rightarrow \Box\Box A$ (*S4*) and $\neg\Box A \rightarrow \Box\neg\Box A$ (*E*)[16] are metaphysically universal. Williamson assumes an axiomatization of *S5* with modus ponens as the sole rule of inference and as axioms all formulae of the form $\Box A$ where A is a truth-functional tautology, all instances of the *K*, *T*, *S4* and *E* schemas (hence avoiding the need for a primitive rule of Necessitation). No one doubts that MU is at least as strong as *T*. Thus, given the antecedent assumption, MU is at least as strong as *S5*.

The argument that it is no stronger, and is therefore exactly *S5*, turns on a theorem of Schiller Joe Scroggs (Scroggs 1951), which asserts that every proper extension of *S5* is the logic of a single finite frame. Williamson uses this result to show that any proper extension of *S5* will have a theorem which is not metaphysically universal. In essence, the argument runs as follows.[17] Let $S5^{+}$ be a proper extension of *S5* and its single frame be the pair $\langle W,R\rangle$, where W has just n worlds. Then $S5^{+}$ has the theorem:

$$\mathrm{Alt}_{n} \qquad \Box P_1 \vee \Box(P_1 \rightarrow P_2) \vee \ldots \vee \Box((P_1 \wedge \ldots \wedge P_n) \rightarrow P_{n+1})$$

There can be no countermodel to Alt_n, because a countermodel would require at least $n + 1$ worlds. A countermodel would have to make each disjunct of Alt_n false. To falsify the first disjunct, we would need a world w_1 at which P_1 is false. To falsify the second, we would need world w_2 at which $P_1 \rightarrow P_2$ is false, and hence at which P_1 is true but P_2 false, so that $w_2 \neq w_1$. Generally, to falsify the i^{th} disjunct we shall require a world w_i distinct from each of the worlds falsifying the preceding disjuncts. But Alt_n has $n + 1$ disjuncts, so a countermodel would require $n + 1$ worlds. But *S5*+ has only models with n worlds. So, there is no *S5*+ countermodel for Alt_n. However,

Alt_n is not metaphysically universal. For under an interpretation on which each P_i says that there are at least i donkeys, each disjunct of Alt_n is false, since it is metaphysically possible that there should have been fewer than i donkeys, so that Alt_n itself is false and so not metaphysically universal.[18]

Clearly this argument, if sound, is enough to close the gap left open by the argument of the preceding section. Does it do more than that? Williamson grants that the assumption that (metaphysical) necessity and possibility are non-contingent is controversial. It might seem that in the absence of a supplementary argument for the *S4* and *E* principles, we have at best a limiting result: MU is no stronger than *S5*. Williamson provides no direct argument for these principles.[19] However, he does present a kind of challenge-argument to those who would stop somewhere short of *S5*. We may assume that no one will propose anything weaker than *T*. It is unclear how anyone who proposes *T* or a system between it and *S5* can provide a general argument to show that every metaphysically universal formula is a theorem of their favored system. There is nothing comparable to Scroggs's result for systems weaker than *S5*, and no other way apparent to secure the desired conclusion. Of course, Williamson grants, one might propose a weaker system and try to deal piecemeal with proposed counterexamples—i.e., try to show that any purported false instance of a universal generalization of a theorem of their system is in fact true, and that any purportedly true non-theorem of their system has a false instance. That is, he grants, 'better than nothing, but a systematic, principled argument for the equation would be far more satisfying' (Williamson 2013: 115).

27.4 Some arguments against *S5*

The principal targets of those who have thought the right logic for metaphysical modality must be one weaker than *S5* have been the *S4* and *B* principles (which, in the presence of the *K* and *T* principles plus the usual rules, suffice for *S5*). In this final section, we comment on the main arguments levelled against them.

27.4.1 Against S4

Nathan Salmon (1989 and earlier work cited therein) argues[20] that the logic of metaphysical necessity cannot be *S4* (or, therefore, *S5*), on the ground that while a certain table could not have been composed of matter significantly different from that from which it is actually composed, it could have been made from slightly different matter, and had it been made from slightly different matter, it would have been possible that it should have been made from significantly different matter from that from which it is actually composed, even though that is not actually possible—so that something that is possibly possible is not possible *simpliciter*, contradicting the *S4* principle that $\Diamond\Diamond A \rightarrow \Diamond A$ (and its equivalent $\Box A \rightarrow \Box\Box A$). Salmon is obviously relying on a version of the principle of necessity of origin, which, in the form Kripke propounds it (Kripke 1980: 114, fn. 56), runs: "If a material object has its origin from a certain hunk of matter, it could not have had its origin in any other matter". Kripke himself hints that he thinks that some qualifications to the principle may be needed, and explicitly mentions the vagueness of the notion of a hunk of matter as one source of problems. Thus, it may be thought that he would endorse a modification along the lines which others have thought necessary or desirable, so that the principle asserts that if a material object originates from a certain portion of matter, it could have originated from a slightly different portion of matter, but could not have originated from a substantially different portion of matter. It is some such version of the principle that Salmon's objection assumes.

Since it is by no means obvious that the presupposed principle—whether in its original or some qualified form—is true, it would be open to a defender of *S5*, especially given an independent argument in favor of *S5* as the correct modal logic for metaphysical necessity, to view Salmon's objection as telling against the origin principle rather than against *S5*. However, Salmon's argument may be resisted without jettisoning the necessity of origin. For the argument evidently relies on the complex counterfactual claim that if the table had been composed of matter slightly different from that from which it is actually composed, it would have been possible that it should have been composed of matter differing only slightly from that, but more substantially from that of which it is actually composed. And it is certainly open to a proponent of the necessity of origin to deny *this* conditional, by insisting that the necessity of origin principle should be understood as requiring that the table could not have been composed of matter more than slightly different from that from which it is *actually composed*, where the phrase 'that from which it is actually composed' refers rigidly. In this case, it would be fallacious to argue:

> Suppose that the table is actually composed of matter m, but that, as might have been the case, it had been composed of slightly different matter m'. Then it would be possible that it should have been composed of slightly different matter from that, i.e., from m'.[21]

27.4.2 Against B

Perhaps the most discussed argument against the *B* principle that $A \rightarrow \Box\Diamond A$ is due to Michael Dummett, responding to Kripke's claim that, contrary to what might be thought to be plain common sense, it is not true, given that there neither are nor ever have been any animals of the sort described and pictured in the stories and paintings which form parts of the myth of unicorns, that there might have been unicorns.[22]

Dummett's argument appears to run as follows:

Although there are no unicorns (that is: there are no animals conforming to the pictures and descriptions of unicorns in the tales and myths), there might have been such animals. As the pictures and descriptions do not settle whether unicorns would belong to the order Artiodactyla (even-toed ungulates, like deer) or to the order Perissodactyla (odd-toed ungulates, like horses), such animals might have been of either of these two orders. In the one case, they would have been of one species, in the other, of another quite distinct species. In world terms, as Dummett puts it:

> there are no unicorns in the actual world w, but there is a possible world u in which there are unicorns, which belong to the order Artiodactyla, and another possible world v in which there are also unicorns, which in that world belong to the order Perissodactyla… . In world u, any animal, to be a unicorn, must have the same anatomical structure as the unicorns in u, and hence, in particuar, must belong to the order Artiodactyla. It follows that the world v is not possible relatively to u, and, conversely, that u is not possible relatively to v. How about the actual world w—is that possible relatively to either u or v? … u is a world in which it holds good that unicorns are necessarily of the order Artiodactyla, whereas in w it is possible for unicorns to be of the order Perissodactyla. Since a proposition necessarily true in u is possibly false in w, w cannot be possible relatively to u, although u is possible relatively to w. The relation of relative possibility (accessibility) is therefore not symmetrical.

Since the B principle, $A \rightarrow \Box\Diamond A$, holds only if the accessibility relation is symmetrical, it follows that the B principle is false.

Ingenious as it is, and plausible as it may seem, Dummett's argument is certainly not irresistible. It evidently requires that there be a proposition p, expressible by the words 'Unicorns belong to the order Artiodactyla', which is such that both $\Diamond\Box p$ and $\Diamond\neg p$. But, starting from the point that the pictures and descriptions constituting the myth do not settle even the order to which animals conforming to them would belong, it may be argued that there is no such proposition. For it follows from this that the pictures and descriptions do not suffice for the introduction of 'unicorn' as a term for a certain animal species.[23] Of course, that does not prevent us from employing 'unicorn' as a general descriptive term, meaning 'horse-like creature with a single spiraling horn projecting from its forehead'. Let us use $U(\xi)$ for this predicate. Then it is evidently (metaphysically) possible that there should have existed creatures answering to this description, i.e., $\Diamond\exists x U(x)$, even if there are not in fact any such, i.e., $\neg\exists x U(x)$. Further, 'ξ belongs to the order Artiodactyla' (briefly $A(\xi)$) is a predicate in good standing, as is 'ξ belongs to the order Perissodactyla' (briefly, $P(\xi)$). And it is further evidently possible both that there should have been creatures satisfying both $U(\xi)$ and $A(\xi)$ and that there should have been creatures satisfying both $U(\xi)$ and $P(\xi)$, i.e., $\Diamond\exists x(U(x) \wedge A(x))$ and $\Diamond\exists x(U(x) \wedge P(x))$—as we might less perspicuously (and potentially quite misleadingly) put it, there might have been unicorns of the order Artiodactyla and there might have been unicorns of the order Perissodactyla. Further still, we may agree that had there been creatures satisfying these complex predicates, they would have necessarily satisfied $A(\xi)$, or $P(\xi)$, depending upon which possibility was realized. In world terms, the proposition that there are horse-like animals which are necessarily of the order Artiodactyla is possibly true at the actual world w (because true at u which is possible relative to w). And the proposition that there are horse-like animals of the order Perissodactyla (and so not of the order Artiodactyla) is likewise possibly true at w (because true at v which is also possible relative to w). But we cannot conclude from this that there is a proposition which is necessarily true at u and so possibly necessarily true at w, but also false at v, and so possibly false at w—from which it would indeed follow that w cannot be possible relative to u, so that, since u *is* possible relative to w, the relation of relative possibility cannot be symmetric. In purely modal terms, what is true is only that $\Diamond\exists x(U(x) \wedge \Box A(x))$ and $\Diamond\exists x(U(x) \wedge \neg A(x))$. There is nothing here that conflicts with the B principle, or with the symmetry of relative possibility, as there would be if we had a pair of true propositions $\Diamond\Box p$ and $\Diamond\neg p$.[24]

Notes

1 Whether and in what sense it is so, and whether, if it is so, logical necessity is the *only* kind of absolute necessity, are contested questions—see, for example, Field (1989): 237 n. 8—on which we shall say a little more later.

2 For the sake of simplicity, we shall here assume that the underlying non-modal logic is classical, but it should be remembered that we could instead adopt a non-classical basis, e.g., intuitionistic and/or free logic.

3 *K* is not, strictly speaking, the weakest modal system, but it is the weakest 'normal' modal logic. We shall assume that the logic of metaphysical modality is at least as strong as *K*, so that weaker non-normal systems can be set aside here. For an explanation of the difference between normal and non-normal systems, see, e.g., Priest (2008: ch. 4) and Hughes and Cresswell (1968: 274–275, 1984: 4). The distinction comes from Kripke (1965).

4 Modern texts usually present modal logics as natural deduction or tableaux systems, in part because they make the construction of proofs easier than it is in axiomatic systems, but comparison of different systems is simpler if they are presented axiomatically.

5 Substituting $\neg A$ for A in the T-axiom gives $\Box\neg A \rightarrow \neg A$, whence by contraposition and double negation elimination, $A \rightarrow \neg\Box\neg A$. Whence by the equivalence $\neg\Box\neg A \leftrightarrow \Diamond A$, it follows that $A \rightarrow \mathrm{sw}\Diamond A$.

6 The same holds even if the underlying quantification logic is intuitionistic. It fails if the underlying quantification logic is, say, a negative free logic, in which it is allowed that names may fail of reference, so that the premises for the inferences usually known as universal instantiation (from $\forall xA(x)$ to $A(t)$) and existential generalization (from $A(t)$ to $\exists xA(x)$) must each be strengthened by an additional premise asserting that t exists. For discussion of this point, see Hale (2013: section 9.2). The Barcan and Converse Barcan principles are named after Ruth Barcan (later Marcus), who published the first system of quantified modal logic in Barcan (1946). For a detailed and very illumination discussion, see Williamson (2013: ch. 2).

7 In the now-standard terminology, a *model* is an ordered triple $\langle W,R,V\rangle$, where W is a non-empty set, $R \subseteq W \times W$, and V—the assignment function—assigns to each sentence variable a function from W into $\{1,0\}$. Models with the same W and R, differing only over V instantiate the same *frame*, i.e., ordered pair $\langle W,R\rangle$.

8 See, for example, Hughes and Cresswell (1968: ch. 8), Priest (2008: ch. 14, 15) or Williamson (2013: section 3.4) for more detail.

9 Note that to validate the *system S5*, as opposed to just its distinctive or characteristic axiom, we need to validate the T axiom, so we need R to be reflexive as well as both symmetric and transitive; similarly, to validate the system B, we need R to be reflexive as well as symmetric, and to validate the system S4, we need R to be reflexive as well as transitive.

10 See Smiley (1963). There are difficulties with the standard account (q.v. Humberstone 1981), in view of which some revision is probably required. For further discussion, see Hale and Leech (2017). These complications do not substantially affect the argument discussed here.

11 This might be thought circular because 'possible world' needs to be understood as 'absolutely possible world'. I don't think this is obviously correct but can't discuss it here.

12 Those who, like me, do not think analyses in terms of worlds can be philosophically illuminating might explain absoluteness in terms of absolutely unrestricted quantification over propositions: $\Box_{abs}A \leftrightarrow \forall q(q \Box\!\!\rightarrow A)$, where q is understood to range over absolutely all propositions, whether expressible in our language or not. For discussion of this and other alternative explanations, see Hale (2015) or Hale (2013: ch. 4).

13 An argument for the absoluteness of logical necessity is given in Hale (2013: 105ff). The view of logical necessity as a species of metaphysical necessity is put forward by Kit Fine (1994) and further defended by Hale (2013: section 4.5). For useful discussion of other views about the status of metaphysical necessity as absolute or otherwise, see Nolan (2011); Rosen (2006).

14 This is not a significant limitation, given that quantified modal logics differ from propositional ones only over the underlying non-modal logic, i.e., they add no new modal principles.

15 See Williamson (2013: 3, 93) for the characterization of metaphysical necessity.

16 Equivalently, $\Diamond A \rightarrow \Box\Diamond A$. Williamson is working in a system with $\Box$ primitive and $\Diamond$ defined.

17 See Williamson (2013: 111) for a fuller statement.

18 As Williamson observes, there are two complications. First, Scroggs's result requires extensions to be quasi-normal (i.e., extends K and is closed under modus ponens and uniform substitution, but need not satisfy the rule of necessitation) and second, it does not extend to logical consequence. Williamson points out that the first gap is merely apparent (Williamson 2013: 111 n. 37) and gives a separate argument that the usual $S5$ consequence relation is right for metaphysical modality (112–114).

19 He does, however, refer to the main arguments against the $S4$ and B principles, indicating that he thinks them to have been convincingly answered. See Williamson (2013: 44 n. 18).

20 In Salmon (1989: 4–8). See also Salmon (1981: 229–252). Salmon's argument is based on one given in Chandler (1976).

21 This objection to Salmon's argument is made in Hale (2013: 128 n. 18). Essentially the same objection is made by Sonia Roca-Royes (2006: section 3.2.1).

22 See Kripke (1980): 23–24, 156–157; Dummett (1993): especially 346–347.

23 To be clear, the reason why 'unicorn' cannot, on the basis of the myth, function as a term for a determinate species is not that there exist no creatures conforming to the descriptions given in the stories, but simply that those descriptions are insufficiently definite to settle the order of ungulates—Arteriodactyla or Perissodactyla—to which unicorns belong; but any species of ungulates must belong to one or the other order. The point is accordingly independent of the widely held view (cf. Wiggins 1993) that species terms are "extension-involving".

24 This response seems to me to fit well with Kripke's remark: "Perhaps according to me the truth should not be put in terms of saying that it is necessary that there should be no unicorns, but just that we can't say under what circumstances there would have been unicorns" (Kripke 1980: 24). A quite closely related, but distinct, criticism of Dummett's argument is given by Marga Reimer (1997). Another, significantly different, criticism, drawing upon a reformulation of Gareth Evans's distinction between "deep" and "superficial" contingency (and necessity) (q.v. Evans 1979: 187–189, reprinted in Evans 1985: 211–213) is made by Ian Rumfitt (2010: 60–64), who argues that Dummett's argument requires the premise 'Necessarily, animals of which the term 'unicorn' (meaning what it actually means) is true are unicorns', which Kripke can and should reject. The line of argument presented earlier might be advanced on the basis of a view about species terms of the kind defended by David Wiggins (1993, reprinted in Clark and Hale 1994), according to which such terms are extension-involving, i.e., lack a sense unless there are (or have been) exemplars of the relevant species. Indeed, Rumfitt tells me, and Wiggins confirms, that Wiggins once suggested (in conversation) a diagnosis of the error in Dummett's argument very similar to mine. But as already emphasized (see note 23), my argument does not rely on this view.

A related argument against *B* given in Stephanou (2000) is criticized in Gregory (2001); see also Hayaki (2005).

References

Barcan, R. (1946) "A Functional Calculus of First-Order Based on Strict Implication," *Journal of Symbolic Logic* 11: 1–16.

Chandler, H. (1976) "Plantinga and the Contingently Possible," *Analysis* 36: 106–109.

Clark, P. and Hale, B. (eds.) (1994) *Reading Putnam*. Oxford: Basil Blackwell.

Dummett, M. (1993) "Could There Be Unicorns?" in Dummett, M. (ed.), *The Seas of Language*. Oxford: Oxford University Press, 328–348.

Evans, G. (1979) "Reference and Contingency," *The Monist* 62: 161–189.

Evans, G. (1985) *Collected Papers*. Oxford: Clarendon Press.

Field, H. (1989) *Realism, Mathematics and Modality*. Oxford: Basil Blackwell.

Fine, K. (1994) "Essence and Modality," *Philosophical Perspectives* 8: 1–16.

Gregory, D. (2001) "B Is Innocent," *Analysis* 61: 225–229.

Hale, B. (2013) *Necessary Beings: An Essay on Ontology, Modality, and the Relations between Them*. Oxford: Oxford University Press.

Hale, B. (2015) *Necessary Beings: An Essay on Ontology, Modality, and the Relations between Them* (2nd ed.). Oxford: Clarendon Press (Paperback).

Hale, B. and Leech, J. (2017) "Relative Necessity Reformulated," *Journal of Philosophical Logic* 46: 1–26.

Hayaki, R. (2005) "The Transience of Possibility," *European Journal of Analytic Philosophy* 1: 25–36.

Hughes, G.E. and Cresswell, M.J. (1968) *An Introduction to Modal Logic*. London: Methuen.

Hughes, G.E. and Cresswell, M.J. (1984) *A Companion to Modal Logic*. London: Methuen.

Humberstone, I.L. (1981) "Relative Necessity Revisited," *Reports on Mathematical Logic* 13: 33–42.

Kripke, S. (1965) "Semantical Analysis of Modal Logic II: Non-Normal Modal Propositional Calculi," in Henkin, L., Tarski, A., and Addison, J.N. (eds.), *The Theory of Models*. Amsterdam: North Holland, 206–220.

Kripke, S. (1980) *Naming and Necessity*. Oxford: Basil Blackwell.

Nolan, D. (2011) "The Extent of Metaphysical Necessity," *Philosophical Perspectives* 25: 313–339.

Priest, G. (2008) *An Introduction to Non-Classical Logic* (2nd ed.). Cambridge: Cambridge University Press.

Reimer, M. (1997) "Could There Have Been Unicorns?" *International Journal of Philosophical Studies* 5: 35–51.

Roca-Royes, S. (2006) "Peacocke's Principle-Based Account of Modality: 'Flexibility of Origins' Plus S4," *Erkenntnis* 65: 405–426.

Rosen, G. (2006) "The Limits of Contingency," in MacBride, F. (ed.), *Identity and Modality*. Oxford: Oxford University Press, 13–39.

Rumfitt, I. (2010) "Logical Necessity," in Hale, B. and Hoffman, A. (eds.), *Modality: Metaphysics, Logic, and Epistemology*. Oxford: Oxford University Press, 35–64.

Salmon, N. (1981) *Reference and Essence*. Princeton, NJ: Princeton University Press.

Salmon, N. (1989) "The Logic of What Might Have Been," *Philosophical Review* 98: 3–34.

Scroggs, S.J. (1951) "Extensions of the Lewis System S5," *Journal of Symbolic Logic* 16: 112–120.
Smiley, T. (1963) "Relative Necessity," *The Journal of Symbolic Logic* 28: 113–134.
Stephanou, Y. (2000) "Necessary Beings," *Analysis* 60: 188–191.
Wiggins, D. (1993) "Putnam's Doctrine of Natural Kind Words and Frege's Doctrines of Sense, Reference, and Extension: Can They Cohere?" in Moore, A.W. (ed.), *Meaning and Reference*. Oxford: Oxford University Press, 192–207.
Williamson, T. (2013) *Modal Logic as Metaphysics*. Oxford: Oxford University Press.

Chapter 28

MODALITY AND THE PLURALITY OF LOGICS

Otávio Bueno

28.1 Introduction

Modality, the realm of the possible and the necessary, has played a significant role in the study of logical consequence. Arguably, when it was first formulated, the concept of logical consequence was understood in modal terms: in today's vocabulary, we say that, in a valid argument, necessarily the truth of the premises guarantees the true of the conclusion, or, even better, without semantic ascent, in a valid argument, the conjunction of the premises and the negation of the conclusion is impossible.

In what follows, I discuss the role of modality in current debates in the foundations of logic. In the background, two families of views regarding the metaphysics of modality are relevant. According to reductive views, the modal, in any of its many guises, can be reduced to the non-modal, whether the implementation of such reduction is achieved in terms of concrete worlds (Lewis 1986) or abstract objects (Menzel 1990, 2018). In contrast, according to non-reductive views, the modal cannot be reduced to the non-modal and ultimately needs to be taken as primitive. These are modalist views (Forbes 1989; Fine 1994; Shalkowski 1994; Lowe 2008; and Scott Shalkowski's chapter on "Modalism" in this *Handbook*). Similarly to reductive proposals, non-reductive approaches come in different forms: some assign an important role to essences, others do not; some invoke metaphysically robust entities, others do not. What they have in common is the emphasis on the need for primitive modality in philosophical theorizing.

In this chapter I examine the role of modality and of modalism in the examination of logical consequence. After briefly distinguishing some forms of modalism and specifying the metaphysically deflationary version I favor, I mention a few arguments against modalism, briefly respond to them, and offer some arguments for the view. I then consider logical pluralism, the view according to which there is a plurality of logical consequence relations (a plurality of logics); different logics are adequate to different domains, and, in some cases, several logics are adequate to the same domain. I then consider some arguments against logical pluralism, respond to them, and provide some arguments in support of the view. Since modalism plays a key role in these latter arguments, I conclude that the combination of both modalism and logical pluralism provides an attractive, integrated view. The second half of the chapter examines the connection between primitive modality and logical space, and I indicate considerations to the effect that a single underlying logical space is needed to support the modal content of logical consequence

and to make sense of genuine logical disagreement. But I note that logical space need not (and, in fact, should not) be understood in any metaphysically robust way.

28.2 Modalism

If modalism is conceptualized in an inclusive way, as a view according to which primitive modality has a significant role in philosophical inquiry, a variety of distinctive conceptions and research programs can be categorized as being modalist. *Metaphysical modalism* articulates the key role of primitive modality in terms of primitive modal facts and facts about essences, even if essences may be taken to be more basic than modality (Fine 1994). As the familiar example goes, it is part of the essence of the singleton of Socrates (the set whose only member is Socrates) that it necessarily has Socrates as a member, but it is not part of the essence of Socrates that he is a member of his singleton. Presumably, Socrates would remain whoever he is (whatever that turns out to be) even if the singleton of Socrates did not exist. Socrates would be whoever he is, whatever kind of being he is, given his essence. Since such essence arguably exists completely independently of Socrates's singleton and Socrates's necessary inclusion in it, essence has metaphysical priority over necessity. In this respect, essences are taken to be more basic than modality: facts about Socrates's essence are more basic than modal facts about Socrates (and his singleton). Essences typically play a prominent role in preventing variations in the properties of an object that would lead to the loss of that object. Provided that the essence or, at least, the essential properties of the object are intact, the object will remain. Interestingly, however, in order for essences to be invoked in these contexts, they ultimately need to be articulated modally: Socrates *would not* be Socrates, if his essence *were* different. This raises the issue of what exactly is gained by invoking essences as something more basic than modality if one's grip on the former is ultimately understood in terms of the latter. Considerations along these lines seem to reinforce the role played by primitive modality even in essentialist views (see also Lowe 2008).

Non-metaphysical modalism avoids increasing the ontology by avoiding the need for essences or metaphysically robust modal facts as part of the articulation of the role that primitive modality plays in philosophical theorizing (Bueno and Shalkowski 2009, 2013, 2015a, 2015b). A distinguishing feature of this approach to primitive modality is that it is modality without additional ontological commitments.

Considerations against modalism have been raised. Reductivists about modality insist that the content of modal discourse is best characterized without recourse to any primitive modal concepts. It is enough to posit concrete worlds (Lewis 1986). It is unclear, however, whether reductionism is ultimately able to account properly for modality, as the introduction of worlds *seems* to change the subject matter. Avid runners might be interested in figuring out whether it is possible for them to run 200 miles. That someone in some other world runs such a distance will not offer *them* much reassurance.

There are, however, better arguments for modalism. A particularly significant one concludes that primitive modality is ultimately presupposed even by reductionist accounts of modal discourse, given that the very structure of logical space requires it (Shalkowski 1994). The logical space, the space of all logical possibilities, has a structure; modal reductionists need to ensure that the reduction of the modal to the non-modal is extensionally adequate so that no impossibilities are displayed in the logical space and all possibilities are exhibited there. Without primitive modality, it is unclear how such adequacy conditions can be achieved. Considerations in order to rule out overgeneration (counting as possible what is, in fact, impossible) and undergeneration (failing to count as possible what is indeed possible) require primitive modality. One needs to establish a proper fit between, on the one hand, the non-modal posits in terms of which the

reduction is implemented (such as worlds or abstract objects) and, on the other, what is possible (or impossible). This, in turn, demands a notion of possibility regarding the very structure of the logical space, so that only what is possible figures in the logical space and nothing impossible does. But these possibilities, which determine the adequacy of the very reductions, on pain of circularity, cannot themselves be characterized in terms of worlds or abstract entities. Absent the prospect of inspecting the domain of worlds to verify the exactness of the reduction, primitive modality needs already to be in place.

A related argument for primitive modality emerges from an examination of the model-theoretic account of logical consequence. According to this account, an argument is valid if, and only if, every interpretation in which the premises are true, the conclusion is also true (for a critical discussion, see Etchemendy 1990). The modal strength of the consequence relation is allegedly captured by the fact that all possibilities are, in principle, covered by variations of all interpretations. Clearly, this is a significant idealization. On the model-theoretic approach, interpretations are set-theoretic constructions and have built-in representational resources in terms of what they can, or cannot, represent. For instance, when they are applied to set theory, interpretations have a domain that, one would have expected, should be a set: presumably, the set of all sets. But there is no such set (if set theory is consistent), and so it is unclear that the representational devices provided by set theory are indeed adequate to account for all possibilities, since not even the domain of the theory is properly captured by the model-theoretic framework (see also Field 1989). Once again, to ensure that set-theoretic representations of the relevant situations neither overgenerate nor undergenerate, a primitive modal concept is needed: only what is possible can be represented set-theoretically, and nothing impossible can be so represented. What is invoked here is not a set-theoretic notion of possibility, given that the validity of arguments should not be tied to what can (or cannot) be represented in set theory, and should be prior to and independent of such representations (see Shalkowski 2004).

28.3 Plurality of logics

One of the most salient features of logic from the twentieth century on is the plurality of non-classical logics, such as constructive, paraconsistent, and quantum logics, among others. Any philosophical account of the foundations of logic and mathematics need to make sense of such a plurality. Despite that, there are those who maintain that one should not adopt any form of pluralism about logic, but rather defend a logical monist perspective, according to which there is only one, properly called, logic: one right logic.

An argument to this effect was presented early on by W.V.O. Quine (1970, 80–81), who insisted that any change of logic (from classical to some other logic) amounts to a change of subject: if one were to insist that the conjunction of a given statement and its negation can be true, one would no longer be talking about negation. Not only would this lead to triviality, given classical logic, but any attempt to deny the principle of non-contradiction would only result in a change of subject matter.

But the argument, of course, does not go very far. Faced with a contradiction of the form *A* and *not-A* in classical logic, triviality results: everything follows. Typically, however, this is not the case in a paraconsistent logic, in which an inference from a contradiction to an arbitrary sentence of the relevant language is blocked. Moreover, it simply begs the question to insist that one is no longer considering logic if an inference that is valid in classical logic is contested. After all, it is then just assumed that classical logic is the correct—and only possible—logic.

For some logical monists, however, the one right logic is other than classical. Graham Priest, for instance, favors logical monism while not favoring classical logic (Priest 2006, ch. 12).

He considers that perhaps the monist logic may be the overlap of what all logics have in common. As it turns out—and Priest himself acknowledges this—the resulting logic may be exceedingly weak. In fact, it is so weak that it may be unrecognizable as a minimally acceptable logic, with perhaps only conjunction elimination being adopted. One is moving perilously close to logical nihilism (according to which there is no right logic). It is unclear at what cost it is worth keeping logical monism, if the result systematizes almost no valid patterns of inference.

In contrast, there are better arguments in support of the plurality of logics favored by logical pluralists. By acknowledging such a plurality, one can easily make sense of, account for, and accommodate the variety of logics other than so-called classical logic. Such a pluralism informs and shapes so much of contemporary logic research. One can also avoid the threat of logical nihilism that certain versions of logical monism face, given that the logical pluralist acknowledges and recognizes that each logic, whether classical, constructive, paraconsistent, or quantum (among others), has a legitimate place in the overall theory of valid inference.

There are different versions of logical pluralism, depending on how the plurality in question is characterized. According to *logical pluralism by cases* (Beall and Restall 2006), an argument is valid provided that in *every case* in which the premises are true, the conclusion is true (for a discussion, and a different formulation of logical pluralism, see Shapiro 2014). Which cases should be considered? They are understood in terms of situations. Consistent and complete situations yield classical logic. Consistent and incomplete situations generate constructive logic. Inconsistent and complete situations yield paraconsistent logic.

This approach, however, faces several difficulties. The proposal's key feature makes it unable to capture quantum logic (Bueno and Shalkowski 2009). Moreover, this conception leads straight into logical nihilism: if indeed *all* cases are considered (rather than cases of a particular kind), one ends up with no logic at all. Finally, given that logical pluralism by cases relies on the model-theoretic conception, it inherits the representational limitations of that conception, and an underlying notion of possibility needs to be in place to display the relevant constraints on logical space.

Modalist logical pluralism avoids these difficulties (Bueno and Shalkowski 2009, 2013). On this view, an argument is valid provided that the conjunction of the premises and the negation of the conclusion is *impossible*. Which (primitive) possibilities should be considered? Consistent and complete possibilities lead to classical logics. Consistent and incomplete possibilities yield constructive logics. Inconsistent and complete possibilities generate paraconsistent logics. Inconsistent and incomplete possibilities form non-alethic logics. Furthermore, possibilities involved in non-distributive situations generate quantum logics. Different logics are, thus, linked to different possibilities. (As will become clear in this chapter, possibilities are different from cases in that they are connected with the relevant properties of the objects and relations under consideration.)

As an illustration, consider the Russell set, the set of all sets that are not members of themselves; that is, $R = \{x: x \notin x\}$. This set cannot be studied in classical set theories: if they are consistent, there is no such set, and there cannot be one, given classical logic. However, in a paraconsistent set theory, whose underlying logic is paraconsistent, it is perfectly possible to study this set. Despite having inconsistent properties, the Russell set is not a trivial object: it does not have all properties (da Costa, Krause, and Bueno 2007).

Modalist logical pluralism has a number of attractive features. It accounts for the *objectivity* of logics. Once a logic is adopted in a given domain, it is not up to us what follows (or does not follow) from it. This is a fact about the consequence relation that is invoked and the sentences in the language in question, not a fact whose outcome we can stipulate.

The proposal also accounts for the *normativity* of logics. A logic is adequate to a given domain if the inferences it licenses preserve the proper relations among the objects in the relevant domain. For instance, in a domain involving inconsistencies, a logic that licenses *explosion* (an inference to the effect that everything follows from a contradiction) is clearly inadequate: one cannot reason productively about inconsistent objects (objects for which both *P* and *not-P* hold, such as the Russell set), given the resulting triviality. In contrast, a paraconsistent logic would be adequate to such a domain. If a logic is adequate to a given domain, it is a *mistake* to draw an inference that this logic does not license, such as explosion.

Moreover, this approach can make room for *logical disagreement*. Since different logics can be about the same domain, there is disagreement among them. Classical and paraconsistent logics overlap: in fact, they license the same inferences in consistent domains. The disagreement emerges only in inconsistent domains, since paraconsistent logic recognizes as a possibility—contradictions without triviality—something that classical logic precludes.

Finally, modalist logical pluralism can accommodate the possibility of *logical revision*. If a logic is inadequate to a given domain, it can be changed, and a different, more adequate, one can be adopted instead. For modalist logical pluralists, given finitely many patterns of reasoning, there is always more than one logic adequate to them. Logical revisions can offer interesting outcomes, since they highlight the dependence of certain mathematical results on the logic used for its derivation. After all, changes in logic often lead to changes in certain results. For example, given Gödel's incompleteness theorem, classical arithmetic is incomplete. However, if the underlying logic is infinitary and has an ω-rule, which allows one to infer from all of the numerical instances of a formula its universal quantification, then arithmetic is complete. If the underlying logic is paraconsistent and allows for inconsistent models, arithmetic, again, is also complete. One can study complete arithmetic by shifting to a non-classical logic, provided that logical revision is possible. Logical pluralists can accommodate this whereas logical monists cannot.

28.4 Modalism and logical space

Given these considerations, for the modalist logical pluralist, there are different adequate logics in different domains of inquiry, and even different adequate logics in the same domain. After all, several paraconsistent logics can be adequate for the same inconsistent domain (da Costa, Krause, and Bueno 2007; Bueno and Shalkowski 2009). This raises the issue of the connection between the plurality of logics and a logical space. Is there a single underlying logical space, or is there a plurality of such spaces? In the remainder of this chapter, I argue that a single logical space is required to capture properly the content of logical pluralism. Otherwise, (i) the modal force of logical consequence is significantly weakened, and (ii) logical disagreement becomes impossible.

After all, regarding (i), it is with respect to a logical space that the space of possibilities for a logic is constituted. In this sense, modal space and logical space are one and the same; moreover, logical space has modal force because it is already a modal space. If there were a plurality of logical spaces, the modal force of an inference would be relativized to the corresponding underlying space, and since an inference's modal force emerges from the logical space, its modal force would be correspondingly relativized.

Regarding (ii), if each logic is tied to a particular logical space, different logics would not disagree with one another. Each logic would be relativized to its own logical space and would be arguably adequate to the logical space it is associated with. Given this relativity and the corresponding adequacy of a logic to its logical space, logical disagreement would not emerge. Each logic would sanction precisely those inferences sanctioned by its logical space, being adequate to

that space. To generate genuine logical disagreement, a single logical space is required, with respect to which different logics are characterized. None of these unwanted outcomes emerges if the plurality of logics obtains over a single logical and modal space.

28.5 What is a logical space?

There are different ways of conceptualizing a logical space (Wittgenstein 1922; Rayo 2013; Turner 2016). It is important, first, to be clear about what a logical space is and what it is used for. A logical space is the space of all logical possibilities, and as such it is crucial that it is both "sound" and "complete", i.e., no impossibilities are in the space and all possibilities are in it. Things have properties and bear relations to other things (where properties and relations are *not* understood in any metaphysically inflated way, such as in terms of abstract objects). Some of these properties are modal: certain things are fragile, other are flammable. Some of these relations are modal: certain plants are edible by us, but not by other animals. The logical space records all of these possibilities. It is relative to these possibilities, conveyed in the logical space, that logical consequence relations, characteristic of the various logics, are constituted. If the conjunction of the premises and the negation of the conclusion of a logically valid argument is impossible, this is a fact about the relevant possibilities, as displayed in the logical space, in which the configurations of all possibilities are included.

It is thus important that the logical space be ontologically independent from the mental and linguistic activities of those who use the many logics that can be characterized relative to such a space. After all, the objectivity of the various consequence relations, the fact that once a logic is specified, it is not up to us what follows (or not) from that logic, emerges from the possibilities exhibited in the logical space. '*A or not-A*' is not logically valid if it is possible for *A* to have an indeterminate truth-value. '*A and not-A*' is not logically invalid if it is possible for *A* to have overdetermined truth-values. All of these outcomes ultimately depend on the possibilities exhibited in the logical space. If fictions are incomplete, the excluded middle fails when applied to them; if the Liar sentence ('This sentence is not true') is asserted, the law of non-contradiction is violated. Classical logics are, of course, highly idealized, so that valid inferences were explicitly not defined for such instances of the sentential variables, rendering them, on this count alone, inadequate as a tool for modeling valid inference in general.

The development of a new logic often highlights new possibilities that were not recognized as such before. Certain regions of the logical space that were not salient before are now emphasized. New possibilities are singled out when a new logic is advanced.

What is the nature of a logical space? Is it a metaphysical structure, one that is ontologically independent from us, or is it something constructed by us? To answer these questions is to provide an interpretation of logical space. It is unclear, however, how to settle this issue. Arguments in favor of one conception of logical space over others are likely to end up begging the question against rival views. For instance, the view according to which we construct the logical space emphasizes that we build such a space as new logics are advanced. The logics are then used to understand the structure of the logical space. But this assumes that the logical space is not a metaphysical structure that exists independently of us: it is being constructed as we go along, while describing possibilities that are not of our own making. In turn, those who favor the metaphysical characterization of logical space will insist on the ontological independence of such a space, which exists quite independently of any features we may invoke to construct or describe the space in question. But this assumes that the logical space is not constructed.

My own inclination is not to settle this issue (of the nature of logical space). The logical space is the space of possibilities that are supposedly independent of us, without, however, having to

be metaphysically robust: one need not specify whether such a space is a universal or a particular, abstract or concrete, dispositional or fully actualized. After all, whatever proposal is offered on the metaphysical view of logical space, the same traits can be obtained on the constructed approach, provided that the constructions stand for possibilities that are taken to be independent of us. This is analogous to what happens in some versions of nominalism in mathematics. Mathematical concepts (understood in a deflationary way, not as abstract objects) are human constructions. Yet, once the concepts are introduced, principles connecting these concepts are formulated, and a logic is adopted, it is no longer up to us what follows or not from the principles in question. In this sense, there is independence of mathematical results from us, despite the fact that we made up the concepts and introduced the principles. Even if the logical space is constructed, what is possible or not, as expressed by such a space, is not up to us.

However the logical space is characterized, what is significant is that possibilities constrain the various logics in the form mentioned earlier. The logical space is the space of what is possible (all possible configurations), and it is relative to that that logical inferences are then characterized. Logical space marks distinctions: what is possible and what is not; what is necessary and what is not; what is the case and what is not. The contrast to what is achieved by a logical space is the complete lack of distinctions or differentiations: an undifferentiated motley of stuff in which everything is true, and that is trivialism.

28.6 Trivialism and logical space

There are at least two forms of trivialism: robust trivialism and frail trivialism. Robust trivialism is the view according to which everything is true (Priest 2006; Estrada-González 2012). This has been the real target of all accounts of logical consequence: if a logical consequence relation is such that everything follows from a contradiction (and, thus, faced with a contradiction, everything is true), then there is no distinction between what does, and what does not, follow from what. In this case, no logic is then characterized; logical theorizing—attempting to draw proper distinctions and determine suitable logical consequence relations—would be pointless.

Frail trivialism is only verbally a form of trivialism: it posits a consequence relation according to which every sentence in the language follows. In a metalanguage, however, it is then expressed that this consequence relation is indeed a consequence relation (a certain algebra defined over a set of propositions) rather than something else, such as a topological space or a rabbit. Since frail trivialism introduces into the metalanguage just the kind of distinction that the trivialist (in the object language) avoids, frail trivialists must justify embracing inferring everything in the object language while demurring from doing so in the metalanguage. Without a justification for these differing treatments of logical consequence relations, robust trivialism is the only trivialist game in town.

Trivialism *should* be resisted precisely because it undermines the very grounds of logical theorizing. Yet, how to argue effectively against it is by no means obvious. Any criticism raised against the view will be promptly accepted by the trivialist! Priest (2006) attempts to argue not that trivialism is false—with which the trivialist agrees (but also disagrees!)—but rather that there are no trivialists. An action requires a choice, and, Priest notes,

> such an action is incompatible with believing everything. It follows that I cannot but reject trivialism. Phenomenologically, it is not an option for me. This does not show that trivialism is untrue …. it is quite possible that everything is the case; but not for me—or for any other person.
>
> (Priest 2006, p. 70)

Given that we need to act, we cannot accept trivialism even if it were true. This is correct but ineffective, since trivialists will not disagree. Furthermore, the argument fails to establish the nonexistence of trivialists, just as an analogous argument similarly fails to establish the nonexistence of dialetheists (those, such as Priest, who believe in the existence of true contradictions). The analogous argument replaces, in Priest's quotation, "believing everything" with "believing in true contradictions", "everything is the case" with "some contradictions are true", and "trivialism" with "dialetheism". One cannot thereby establish that dialetheists do not exist. In the end, Priest's considerations are not successful (Bueno 2015).

28.7 Logical pluralism and logical space

Logical pluralism, as noted, is the view according to which there is a plurality of adequate logics in different domains, and several adequate logics even in the same domain. A domain is a region in the logical space. A logic is adequate as long as the logical consequence relation it imposes on the various statements properly matches the relations among the various objects in the domain. For instance, given a certain understanding of quantum mechanics, a distributive logic seems inadequate to capture the logical relations among quantum objects. Different logics, then, are adequate to different regions of the logical space. As we saw, consistent and complete possibilities are adequately tracked by classical logic. Inconsistent and complete possibilities are adequately tracked by paraconsistent logics. Consistent and incomplete possibilities are adequately tracked by constructive logics.

Different logics sanction different possibilities: certain contradictory states (of the form *A* and *not-A*), which are impossible according to classical logic, become possible in paraconsistent logics. Since the scope of the possible in the logical space is broader in paraconsistent logics, inferences that are taken to be valid in classical logic become invalid in a paraconsistent setting, such as *explosion*. As a result, on the view favored here, logic is not topic neutral: depending on the domain under consideration, depending on the region of the logical space at stake, different logics are adequate. (This, of course, contrasts with the received view of logic according to which logic is involved with the most general truths; see Russell 1903, ch. 1–3.)

The *modal force* of logical consequence emerges from the logical space with respect to which logical consequence is characterized. As noted, I favor a modal conception of logical consequence: in a valid argument, the conjunction of the premises and the negation of the conclusion is impossible. The impossibility emerges from the fact that no region in the logical space corresponds to the situation represented by the conjunction of the premises and the negation of the conclusion. Since the logical space is "sound" and "complete", the fact that there are no regions in the space corresponding to the situation in question grounds the corresponding impossibility. A single logical space underlies this fact.

With a single logical space, the various logics are assessed. Some would be adequate relative to certain regions of the space and inadequate elsewhere. Several would be adequate relative to the same regions of the space and inadequate elsewhere. None would be adequate relative to the entirety of the space. Note, however, that this is not logical nihilism, given that several logics, such as classical, constructive, paraconsistent, or quantum, are adequate at least relative to some regions of the logical space. For the nihilist, none are, at least none that are minimally strong to be considered a logic.

In order to make sense of *logical disagreement*, that is, the disagreement between defenders of rival logics, it is crucial to have a *single* logical space with respect to which the logics in question are assessed. If each logic specified its own logical space, and was assessed relative to it, different logics would not disagree: they would concern different logical spaces, relative to which each of them would be adequate. Logical rivalry emerges when the *same* logical space is at stake

whenever different logics are compared and contrasted. Different regions of the same logical space are then considered, and different logics specify different, and in some cases overlapping, regions of the space, relative to which their adequacy emerges.

28.8 Conclusion

The considerations presented in this chapter suggest that, in order for logical pluralism to be philosophically satisfying, it is important that a *single* logical and modal space be in place. Otherwise, it is unclear how the modal force of logical consequence relations could be properly obtained, and logical disagreement would become impossible. Logical space, nevertheless, need not, and should not, be thought of in reified terms. Although a robust metaphysical understanding of logical spaces is available, there is no need to adopt it: a more deflationary understanding of such space, which resists the need for establishing the metaphysical nature of logical spaces, while acknowledging the independence of the possible, is readily available. In the end, with a single logical space, modalist logical pluralism becomes, at least, a more satisfying view.

References

Beall, JC, and Restall, G. (2006) *Logical Pluralism*, Oxford: Oxford University Press.

Bueno, O. (2015) "Truth, Trivialism, and Perceptual Illusions", in J.-Y. Béziau, M. Chakraborty, and S. Dutta (eds.), *New Directions in Paraconsistent Logic*, Dordrecht: Springer, 465–476.

Bueno, O., and Shalkowski, S. (2009) "Modalism and Logical Pluralism", *Mind* 118: 295–321.

Bueno, O., and Shalkowski, S. (2013) "Logical Constants: A Modalist Approach", *Noûs* 47: 1–24.

Bueno, O., and Shalkowski, S. (2015a) "Modalism and Theoretical Virtues: Toward an Epistemology of Modality", *Philosophical Studies* 172: 671–689.

Bueno, O., and Shalkowski, S. (2015b) "Empirically Grounded Philosophical Theorizing", in Chris Daly (ed.), *The Palgrave Handbook of Philosophical Methods*, Hampshire: Palgrave, 231–257.

da Costa, N.C.A., Krause, D., and Bueno, O. (2007) "Paraconsistent Logics and Paraconsistency", in D. Jacquette (ed.), *Philosophy of Logic*, Amsterdam: North-Holland, 791–911.

Estrada-González, L. (2012) "Models of Possibilism and Trivialism", *Logic and Logical Philosophy* 21: 175–205.

Etchemendy, J. (1990) *The Concept of Logical Consequence*, Cambridge, MA: Harvard University Press.

Field, H. (1989) *Realism, Mathematics and Modality*. Oxford: Blackwell.

Fine, K. (1994) "Essence and Modality", in J. Tomberlin (ed.), *Philosophical Perspectives 8: Logic and Language*, Atascadero, CA: Ridgeview Publishing Company, 1–16.

Forbes, G. (1989) *Languages of Possibility*, London: Basil Blackwell.

Lewis, D. (1986) *On the Plurality of Worlds*, Oxford: Blackwell.

Lowe, E. J. (2008) "Two Notions of Being: Entity and Essence", in R. Le Poidevin (ed.), *Being: Developments in Contemporary Metaphysics, Royal Institute of Philosophy Supplement*, vol. 62, pp. 23–48.

Menzel, C. (1990) "Actualism, Ontological Commitment, and Possible Worlds Semantics", *Synthese* 85: 355–389.

Menzel, C. (2018) "Actualism", in Zalta, E. N. (ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2018 edition), URL: https://plato.stanford.edu/archives/sum2018/entries/actualism/.

Priest, G. (2006) *Doubt Truth to Be a Liar*, Oxford: Clarendon Press.

Quine, W.V.O. (1970) *Philosophy of Logic* (2nd ed., 1986), Cambridge, MA: Harvard University Press.

Rayo, A. (2013) *The Construction of Logical Space*, Oxford: Oxford University Press.

Russell, B. (1903) *The Principles of Mathematics*, Cambridge: Cambridge University Press.

Shalkowski, S. (1994) "The Ontological Ground of the Alethic Modality", *Philosophical Review* 103: 669–688.

Shalkowski, S. (2004) "Logic and Absolute Necessity", *The Journal of Philosophy* 101: 55–82.

Shapiro, S. (2014) *Varieties of Logic*, Oxford: Oxford University Press.

Turner, J. (2016) *The Facts in Logical Space: A Tractarian Ontology*, Oxford: Oxford University Press.

Wittgenstein, L. (1922) *Tractatus Logico-Philosophicus*, London: Routledge and Kegan Paul.


Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

PART 7

Modality in the history of philosophy


Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Chapter 29

ANCIENT GREEK MODAL LOGIC

Robin Smith

Formal logic began in classical Greece. Aristotle's *Prior Analytics* contains a fully developed theory of logical inference, with mathematical proofs for the validity of forms of inference and some metalogical results. While no earlier logical work has survived, Aristotle's contemporaries and immediate precursors appear also to have studied logical inference, and we can trace a separate line of development for logical theory from the "Megarians" that is clearly flourishing a generation later with Diodorus Cronus and the "Dialectical School." From this descended the logic of the Stoics, among whom Chrysippus became the most celebrated logician of the Hellenistic period. Unfortunately, our knowledge of this alternative tradition is very fragmentary.

Modal logic, that is, the study of inferences in which necessity, possibility, and impossibility play a role, was pursued in both these traditions. The theory of inference in Aristotle's *Prior Analytics* I.1–7 (usually called the *syllogistic*) has been shown by modern studies to be a masterful development. However, his modal extension of this in *Prior Analytics* I.8–22 is far more problematic and has often been seen as incoherent. By contrast, we know far less about the study of modal logic in the alternative tradition represented by the Stoics, since very little has survived apart from indirect accounts. In what follows, I first discuss the treatment of necessity and possibility in the broad historical context of ancient Greek logic and philosophy, and then turn to the details of Aristotle's system.

First, however, there is an important point of perspective. Modern formal logic takes the bearers of truth and falsehood to be propositions that are fully determinate. Utterances of declarative sentences express propositions, but which proposition a sentence expresses may vary with time and other contexts of utterance: "It's raining here" uttered on January 1, 2000, in Chicago expresses a different proposition from that same sentence uttered in Cairo on July 1, 1900. In either case, the proposition expressed is either timelessly true or timelessly false. Ancient Greek logicians instead took propositions themselves to change their truth-value with the passage of time. Thus, the sentence "It's raining in Chicago" expresses the same meaning uttered yesterday, today, and tomorrow, but its truth-value changes with the weather. This has consequences for understanding modally qualified propositions. The standard procedure for modal semantics in modern logic is to appeal to possible worlds (alternative ways things might have been). By contrast, Greek logicians tended to associate necessity with omnitemporality, so that a proposition is

necessarily true if it is *always* true. For the Greeks, then, necessity is closely connected with unchangeability and possibility with changeability.

This takes on greater importance against the background of early Greek thought, where the distinction between what is necessary and what is possible was often linked with a metaphysical distinction, that between what is changeable and what is not. Parmenides (active in the early fifth century BCE) argued that since it is a contradiction to say that what is also is not, change cannot be real: if there were change, then what is not would have to be, which (on his view) is contradictory. Similarly, there cannot be a multiplicity of things, since that would require that there is something which *is not* some other thing, and thus something which both is and is not. However we interpret Parmenides' arguments, they had an enormous impact on subsequent Greek philosophy. We find explicit responses to them in Anaxagoras, Democritus, Plato, and Aristotle. Plato instead developed a metaphysics involving two distinct realms, the realm of being (which is necessarily as it is, unchanging with time) and the realm of becoming, where we find that which "both is and is not." Aristotle also tells us that a somewhat obscure group of philosophers, the Megarians, held that only what is actual is possible. If you are now sitting, then it is not possible that you are now standing, since it would be a contradiction to say that you are now sitting and now not sitting. If you attempt to refute this by standing up and thus showing that you *can* stand, the Megarians reply, that only shows that you can stand *now*, not that you could have stood *then*, when you were seated (see Aristotle, *Metaphysics* Θ.3).

The necessity of the past and Diodorus' "master" argument

This Megarian position can be given a sharper focus if we combine it with the principle that whatever has been true in the past is now necessarily true: if it is true that it rained yesterday in Chicago, then it is now necessarily true that it rained yesterday in Chicago. The unchangeability of the past is close to proverbial in ancient Greek thought:

> For of this alone even a god is deprived,
> To make undone those things that have been done.
> (Agathon, quoted by Aristotle, *EN* 1139b10–11)

In *On Interpretation 9*, Aristotle criticizes an argument for the necessity of everything that happens resting on this principle and the principle of bivalence. Consider the proposition "There will be a sea-battle tomorrow," interpreted as temporally definite, e.g. "There will be a sea-battle tomorrow, January 1, 2000." Suppose, in addition, that every declarative sentence is, when uttered, necessarily either true or false. Therefore, this proposition, when uttered on December 31, 1999, was necessarily either true or false. However, if it was true when uttered, then since the past is now unchangeable, it is now necessarily true that it was true, and hence now necessary that there will be a sea-battle tomorrow. Moreover, whether or not the sentence in question was actually uttered on the day in question, all that matters is whether *things were such as to make it true*. If they were, then it is now necessarily true that they were, and hence now necessary that there will be a sea-battle tomorrow.

Aristotle thinks the conclusion of this argument is unacceptable, since it implies that there is no choice and that deliberation is pointless. However, though it is clear that he wants to reject it, it is far less clear just how he proposes to do so: an enormous literature has arisen on its interpretation. His remarks are contained in an extended discussion of what he calls "contradictions"

(*antitheseis*), i.e., pairs of sentences one of which affirms of its subject exactly what the other denies of that subject. While he holds that in general the members of any such contradiction must be one true and the other false, he argues that this may fail to hold for contradictions consisting of future contingent statements such as "There will be a sea-battle tomorrow" and "There will not be a sea-battle tomorrow." While this much is obvious, there has been a long-standing controversy about exactly how he proposed to avoid it and about whether his view is coherent (among the vast literature on the subject, see Anscombe 1956; Hintikka 1964; Frede 1970; Whitaker 1996).

Aristotle does not name those who advanced the argument he is attacking, but, as noted earlier, in *Metaphysics* Θ 3 he attributes a shorter argument with the same purpose to "Megarians." The association with the Megarians is significant, since a more fully developed version was advanced by Diodorus Cronus (died ca. 284 CE) who was probably influenced by Megarian logic, if not himself a member of the Megarian School (see Sedley 1977, 2018; Döring 1989; Ebert 2008). This argument, which Diodorus called the "Master" argument (*kurieuôn logos*), argued that these three propositions are an inconsistent triad:

(1) What is past is necessary.
(2) The impossible does not follow (from) the possible.
(3) There is something possible which neither is nor will be true.

Diodorus' own response to the "Master" was to reject (3), leading to his definitions of the possible as "that which either is or will be true" and the necessary as "that which is true and never will be false."

The Stoic school that arose in the latter half of the third century BCE took its logic largely from Diodorus and others of the Dialectical School, and developing a response to the "Master" became a major concern of Stoic logicians, though we know relatively little of the details. We do have at least a sketch of the response of Chrysippus (ca. 280–ca. 207 BCE), regarded in antiquity as the greatest Stoic logician. He accepted the mutual incompatibility of its three propositions but argued that the one to be rejected is "The impossible does not follow from the possible" by appealing to Stoic propositional semantics. He considered these propositions:

(1) If Dion is dead, then *he* [indicating Dion] is dead.
(2) Dion has died.
(3) *He* [indicating Dion] has died.

Using the criterion that a conditional proposition is true if and only if its antecedent is incompatible with the denial of its consequent, proposition (1) is true. Therefore, (3) does follow from (2). Moreover, since Dion is not immortal, (2) is possible. However, Chrysippus held that a proposition containing an indexical term ("he" in this case) "perishes" (ceases to exist) when the object ostended ceases to exist. Consequently, when Dion has died, (3) does not exist, and since it is therefore false at all times (i.e. whenever it exists), it is impossible. We therefore have something impossible that follows from something possible. Unfortunately, since our knowledge of Chrysippus' argument relies on summaries by (sometimes hostile) others, we cannot say with confidence what further use he made of it. Since the Stoics were determinists, he may have been offering a response to the complaint that determinism makes everything necessary by displaying something that is possible but neither is nor will be true.

Aristotle's modal logic

Aristotle's modal logic is an extension of his basic logical system (the *syllogistic*), and so an account of it must begin with that basic system as presented in *Prior Analytics* I.1–7. The syllogistic is a theory of the inferential relationships among four types of *categorical sentence*:

Universal affirmative: Every human is two-footed
Particular affirmative: Some human is two-footed
Universal negative: No human is two-footed
Particular negative: Some human is not two-footed / Not every human is two-footed

Aristotle adopts a somewhat artificial way of expressing these *categorical* sentences, as they are usually called: "Every/Some/No A is B" becomes "B is predicated of every/some/no A" (this language is not at all ordinary Greek). For convenience, I abbreviate these as follows:

AaB A is predicated of every/all B
AeB A is predicated of no B
AiB A is predicated of some B
AoB A is not predicated of some B /A is not predicated of every B

He then investigates all possible combinations of pairs of premises having one term in common to see which "syllogize," that is, entail as conclusion a categorical sentence using the other two terms in the premises. He first argues that four forms are valid on the basis of their semantics:

AaB, BaC |- AaC (traditionally called Barbara since the 13th century)
AeB, BaC |- AeC (Celarent)
AaB, BiC |- AiC (Darii)
AeB, BiC |- AoC (Ferio)

He characterizes these as "complete" or "perfect" (*teleios*), which for him appears to mean that their validity is evident from the meanings of the categorical sentences alone without requiring additional deductive steps. To prove the validity of further forms, he argues that certain *conversion* inferences hold (*An. Pr.* I.2):

AeB |- BeA
AiB |- BiA
AaB |- BiA

With these conversion rules plus a form of indirect proof ("proof through the impossible"), Aristotle then gives proofs for the validity of a further ten forms of argument, a type of natural deduction (see Corcoran 1974):

AaB, AeC |- BeC (Camestres)
AeB, AaC |- BeC (Cesare)
AeB, AiC |- BoC (Festino)
AaB, AoC |- BoC (Baroco)
AaC, BaC |- AiB (Darapti)
AeC, BaC |- AoB (Felapton)
AiC, BaC |- AiB (Disamis)

AaC, BiC |- AiB (Datisi)
AoC, BaC |- AoB (Bocardo)
AeC, BiC |- AoB (Ferison)

Aristotle also shows that combinations of premise forms other than these "do not syllogize" by offering countermodels: two triplets of terms such that, when substituted for the terms in the premise pair in question, they satisfy the premises but yield respectively a true putative universal affirmative and a true universal negative "conclusion." Modern studies have shown that the resulting system is logically complete (Corcoran 1972). In addition to these methods, Aristotle sometimes notes that alternative proofs are possible using what he calls *ekthesis*, relying on these further rules:

AiB |- for some C, AaC and BaC
AoB |- for some C, AeC and BaC

While his system is complete without proofs through *ekthesis*, they are indispensable in the modal syllogistic.

I have described this theory as applying to a certain class of propositions, but Aristotle himself treats it as the theory of inference *tout court*, since he also believes, and believes he can show, that all valid inferences can be "reduced" to the "arguments in the figures," as he calls his system (*An. Pr.* I 30).

The modal syllogistic

The modal syllogistic is an expansion of this system that adds two additional forms of predication: "necessarily belongs" and "possibly belongs." Aristotle considers what happens when "belongs" is replaced by "belongs necessarily" or "belongs possibly" in one or both of the premises of the established forms. In the majority of cases, he follows the same pattern of proof he used in the corresponding non-modal case to see whether they yield a conclusion in the corresponding modal case. This of course requires modal versions of the conversion rules. Proof through impossibility raises special difficulties in the modal syllogistic, and Aristotle sometimes uses a modal form of *ekthesis* in its place (see later in this chapter for further discussion of this point). As in the non-modal case, he proves invalidity by providing countermodels, though these differ in form from those for the non-modal syllogistic. The modal syllogistic resulting from these additions, presented in *Prior Analytics* I.8–22, is considerably more complicated than the non-modal syllogistic. It is also far more problematic to interpret, as commentators since ancient times have noted.

Modal conversion rules

Aristotle says that necessity premises convert in the same way as non-modal premises:

Necessarily AeB |- Necessarily BeA
Necessarily AiB |- Necessarily BiA
Necessarily AaB |- Necessarily BiA

His arguments for these in *Prior Analytics* I.3 are parallel to his arguments for the non-modal cases in I.2. For necessary-e, he says:

> If it is necessary for A to belong to no B, then it is also necessary for B to belong to no A. For if it were possible to some, then A would be possible to some B.
>
> (25a29–32)

For necessary-a and necessary-i, he says:

> If A belongs of necessity to every or to some B, then it will be necessary for B to belong to some A. For if it is not necessary, then neither would A belong of necessity to some B.
>
> (29a32–34)

The first argument assumes "possibly B to some A" is the denial of "necessarily B to no A," so Aristotle is apparently following the (now-standard) understanding that "not necessarily" is equivalent to "possibly not" and treating "necessarily" and "possibly" as sentential operators:

$$\sim\Box(\mathrm{BeA}) \leftrightarrow \Diamond \sim (\mathrm{BeA}) \leftrightarrow \Diamond(\mathrm{BiA})$$

However, in his subsequent discussions of the conversions of possibility sentences (*Prior Analytics* I.13), Aristotle adopts an account of possibility that is inconsistent with this (I will return to the latter point later), In addition, as interpreters since the Middle Ages have noted, "Necessarily A belongs to B" may be taken to mean either "The property of being necessarily A belongs to B" (the *de re* reading) or "The proposition that A belongs to B is necessary" (the *de dicto* reading). This distinction is a matter of the scope of the modal operator. In what follows, I will use parentheses to indicate it, as follows:

N(AaB) = □ ∀x(Bx↠Ax) (*wide scope, de dicto*)
NAaB = ∀x(Bx↠□ Ax) (*narrow scope, de re)*
N(AiB) = □ ∃x(Bx&Ax) (*de dicto*)
NAiB = ∃x(Bx&□ Ax) (*de re)*
N(AeB) = □ ∀x(Bx↠~Ax) (*de dicto*)
NAeB = ∀x(Bx↠□ ~Ax) (*de re)*
N(AoB) = □ ∃x(Bx&~Ax) (*de dicto*)
NAoB = ∃x(Bx&□ ~Ax) (*de re)*

Aristotle's conversion rules are valid on the *de dicto* reading, since they follow from the conversion rules for non-modal sentences, e.g. since AeB |- BeA, **N**(AeB) |- **N**(BeA). However, they are not valid on the *de re* reading. (NA)eB, for instance, says that nothing that is B is necessarily A, from which it follows that nothing that is necessarily A is B; however, the latter does not entail "Nothing that is A is necessarily B."

The two understandings of possibility premises

The conversion rules Aristotle gives for premises asserting possibility differ radically from the corresponding non-modal rules. In *Prior Analytics* I.13, he first takes note of two ways "possible" may be understood: as "not necessarily not" (a sense he also recognizes in *On Interpretation*) and as "neither necessarily true nor necessarily false." He then chooses the latter rather than the former as his preferred definition, so that it is equivalent to "Possibly A does not belong to B." This might be expressed in the following schema:

$$\Diamond \mathrm{Ax} \leftrightarrow \Diamond \sim \mathrm{Ax}$$

From this, Aristotle concludes that all possibility sentences are equivalent to their *opposites*; that is, each affirmative sentence is equivalent to the corresponding negative sentence:

Possibly A belongs to every B -|- Possibly A belongs to no B
Possibly A belongs to some B -|- Possibly A does not belong to some B

Combined with the previous principle, these make sense if we understand possibility sentences *de re*. For the universal case, we have:

$$\text{Possibly A belongs to every B} = \forall x\left(Bx \rightarrow \Diamond Ax\right) = \left(\mathbf{P}A\right)aB$$

But since ◊Ax entails ◊~Ax, we have:

$$\left(\mathbf{P}A\right)aB = \forall x\left(Bx \rightarrow \Diamond Ax\right) \leftrightarrow -\forall x\left(Bx \rightarrow \Diamond \sim Ax\right) = \left(\mathbf{P}A\right)eB$$

Similarly,

$$\left(\mathbf{P}A\right)iB = \exists x\left(Bx \,\&\, \Diamond Ax\right) \leftrightarrow -\exists x\left(Bx \,\&\, \Diamond \sim Ax\right) = \left(\mathbf{P}A\right)oB$$

Consequently, says Aristotle, all possibility premises are "affirmative in form" and follow the conversion rules corresponding to affirmative non-modal sentences. However, they also follow the additional a-e and i-o rules indicated earlier.

This has major consequences for Aristotle's modal logic. First, given his understanding of possibility, we no longer have the equivalences between "possibly" and "not necessarily not" or between "necessarily" and "not possibly not": "A possibly belongs to B" is inconsistent both with "A necessarily belongs to B" and with "A necessarily does not belong to B." Consequently "x is not necessarily A" is not equivalent to "x is possibly not A" but rather to the disjunction "either x is necessarily not A or x is possibly not A" (where the latter is equivalent to "either x is necessarily not A or x is possibly A"). Similarly, the denial of "x is possibly A" is "either x is necessarily not A or x is necessarily A." These disjunctions are not expressible as categorical sentences. Aristotle is aware of this, as is clear from his treatment of the modal forms of Baroco and Bocardo (see later in this chapter). They also prevent him from using proofs through impossibility in the same way as in the non-modal case. Instead, he substitutes proofs by *ekthesis* for syllogisms with necessary premises, while in the case of possible premises he does use a necessary premise as the *reductio* hypothesis but then notes (correctly) that the conclusion asserts "possibility not in accordance with our definition" (see later in this chapter). He never makes use of a *premise* asserting "possibility not in accordance with our definition."

Given his definition of possibility, Aristotle expands his modal syllogistic with all the forms that result from substitutions of possible-e for possible-a and of possible-o for possible-i premises and conclusions.

Aristotle's modal proofs

Aristotle then proceeds to investigate premise combinations involving at least one modal premise. His approach is to consider only those combinations he has already included in the non-modal syllogistic, determining whether the same categorical type of conclusion follows in the modal case and, if so, with what modality (though he does include a number of additional forms that result from the substitution of possibility premises as noted earlier). He considers, in order: (1) cases with two necessary premises; (2) one necessary and one assertoric

(non-modal) premise; (3) two possible premises; (4) one possible and one assertoric premise; (5) one necessary and one possible premise. In general, he provides proofs, where possible, that follow the patterns used in the non-modal case but with the application of modal conversion rules. As before, he provides countermodels to show the invalidity of invalid forms, but these are now usually single term-triples showing, not that no conclusion follows, but instead that a conclusion of a certain modality does not follow. The details are both complex and difficult to interpret. In what follows, I will describe a particular syllogistic form by giving its medieval name followed by a series of letters indicating the modalities of the two premises and conclusion: for instance, "Barbara NAN" means "A syllogism in the form Barbara with a necessary major premise, an assertoric minor premise, and a necessary conclusion."

Aristotle begins with a quick treatment of all the NN cases: for each valid form in the non-modal syllogistic, the corresponding NNN form is valid and, with two exceptions, proofs are exactly analogous to those for the non-modal cases, since analogous conversion rules hold for N sentences (*Prior Analytics* I.8). The exceptions are the two cases requiring proofs through impossibility in the non-modal case (that procedure is not available because of Aristotle's somewhat complicated views on possibility sentences). A proof by *ekthesis* for Baroco NNN would be as follows:

Necessarily BaA	premise
Necessarily BoC	premise
Necessarily CaX	2, N-o-ecthesis
Necessarily BeX	2, N-o-ecthesis
Necessarily XeB	4, N-e-conversion
Necessarily XeA	1, 5 Barbara NNN
Necessarily AeX	6, N-e-conversion
Necessarily AoC	3, 8, Felapton NNN

Steps 5 and 7 here rely on necessary-e conversion, which as noted previously requires a *de dicto* reading. Steps 3 and 4 require a modal version of e-ecthesis:

If necessarily BoC, then for some part of C (say X), necessarily BeX and necessarily CaX.

Since Aristotle's proofs of the conversion rules of necessary premises appear to require the *de dicto* reading, so will all these proofs. However, his treatment of AN and NA cases appears to be inconsistent with this. He says that for first-figure syllogisms, a necessity conclusion follows when it is the major premise that is necessary, but that only an assertoric conclusion follows when it is the minor premise that is necessary:

Necessarily AaB, BaC |- Necessarily AaC
AaB, Necessarily BaC |- AaC only
Necessarily AeB, BaC |- Necessarily AeC
AeB, Necessarily BaC |- AeC only

He defends the first of each pair (NAN) as follows:

> Since A belongs, or does not belong, of necessity to every B and C is one of the Bs [τι των B], it is obvious that it [A] will be of necessity in one or the other of these relations [ἔσται θάτερον τούτων] to C.
>
> (30a21–23)

This language strongly recalls that used in I.4 to justify Barbara AAA and Celarent AAA and appeals directly to his definition of "belongs to all/belongs to none." He rejects the corresponding ANN cases as follows:

1. Suppose AaB, Necessarily BaC |- Necessarily AaC
2. Necessarily AaC, Necessarily BaC |- Necessarily AiC (*Darapti* with necessity premises, already established)
3. AaB, Necessarily BaC |- Necessarily AiC (1 and 2)
4. However, 3 is invalid (shown by a counterexample)
5. Therefore, 1 is invalid (since with 2, known to be valid, it yields the invalid 3)

We can preserve Aristotle's distinction between the NAN and the ANA cases here if we adopt a *de re* reading for the necessary premise (and conclusion)/explanation for Aristotle's distinction between "the two Barbaras" is that he treats "necessarily A" and "A" as distinct predicates. Using "**N**A" for "necessarily A," we then have:

(**N**A)aB, BaC |- (**N**A)aC: valid (Barbara with terms **N**A, B, and C)
AaB, (**N**B)aC |- (**N**A)aC: invalid since it contains four distinct terms (**N**A ≠ A)

However, this interpretation calls into question the validity of Barbara NNN:

(**N**A)aB, (**N**B)aC |- (NA)aC : invalid, since it contains four terms

We might remedy this by supposing that **N**B implies B, which Aristotle appears to accept elsewhere, though he does not mention it in this context. However, the problem remains that Aristotle's acceptance of Barbara NAN seems inconsistent with the *de dicto* reading, while his conversion rules for necessary propositions seem inconsistent with the *de re* reading. This problem—the "two Barbaras" problem—is the most widely discussed issue in interpreting Aristotle's modal syllogistic.

There are other related problems. Aristotle says that both Baroco ANN and Baroco NAN are invalid. For Baroco ANN, he uses the same countermodel he used earlier for Camestres NAN**,** concerning which he says this:

> Let A be animal, B human, C white, and let the premises be taken similarly. For it is possible for animal to belong to no white. Then human will not belong to any white either, but not of necessity, for it is possible for there to be a white human (though not as long as animal belongs to nothing white).

The example, as given for Camestres ANN, is as follows:

> Suppose a situation in which animal belongs to no white (i.e. that there are no white animals). Then, since every human is (of necessity) an animal, it would follow that in that situation nothing white is a human. However, it is not the case that in that situation, it would be necessary that nothing white is a human.

This counterexample is quite different in form from Aristotle's usual counterexamples, which consist of two sets of three terms such that, when substituted for the terms in a putative syllogism, result in a true universal (and necessary, in the case of modal premises) affirmative

conclusion in one case and a true universal (and necessary, in the case of modal premises) negative conclusion. Here, he simply appeals to the claim that "it is possible for there to be a white human." One might object as follows: in a situation in which there are no white animals, everything that is white will be something other than an animal; but surely, whatever is not an animal is *necessarily* not a human (e.g. a white stone is necessarily not a human, even if it is not necessarily white). Therefore, in a situation in which there are no white animals, each thing that is white is something that is necessarily not a human.

Furthermore, proofs using ecthesis can be given for both Baroco ANN and Bocardo NAN in the same way as for the NNN cases. The proof for Baroco ANN:

BaA	premise
NBoC	premise
CaX	2, ecthesis
NBeX	2, ecthesis
NXeB	4, conversion
NXeA	1, 5 Barbara NAN
NAeX	6, conversion
XiC	3, conversion
NAoC	7, 8, Ferio NAN

However, Aristotle rejects Baroco ANN and Baroco NAN with countermodels, which contradicts this.

In the case of Camestres NAA, Aristotle offers two additional arguments that a necessity conclusion does not follow. First (30b21–24), he follows the non-modal proof and notes that the corresponding form Celarent ANN is not valid, though Celarent NNN and NAA are valid. This is not so much a proof of invalidity as an argument that a specific proof for validity cannot be given. Second (30b24–31), he argues that if Camestres NAN were valid, then Felapton ANN would be valid, but (as he shows by a countermodel) it is not. For suppose **N**AaB, AeC |- **N**BeC to be valid:

NAaB	premise
AeC	premise
NBeC	1,2, Camestres NAN (assumed to be valid)
NCeB	3, conversion
NBiA	1, conversion
NCoA	4,5 Ferio NNN

Thus, we can deduce AeC, **N**AaB |- **N**CoA (Felapton ANN) using Camestres NAN. However, says Aristotle, "nothing prevents something having been taken as A such that C possibly belongs to every A." He illustrates this with a counterexample: if we suppose that there are no white animals, it still might be the case that every animal is *possibly* white (that is, we can suppose that every animal in such a situation has the possibility of becoming white).

Syllogisms with possible premises

Aristotle's account of syllogisms with possibility premises is both much more complex and more problematic. In general, he treats "possibly A" and "A" as distinct predicates (I will represent "possibly A" here as **P**A), which seems congenial to a *de re* reading. However, given his definition

of possibility, **P**A is logically independent from A: something may be possibly A whether or not it is A, and something may be A whether or not it is possibly A. Consequently, whatever is necessarily A is both not possibly not A *and also not possibly A*, in Aristotle's defined sense, and likewise what is not possibly A is both not possibly A *and not possibly not A*. In this section, I will use ***P***(*Q*) (for example, ***P***(AaB)) to represent a proposition expressing possibility "not according to the definition," that is, a denial of a necessary proposition.

To begin with, Aristotle accepts the following inference forms as valid:

PAaB, **P**BaC |- **P**AaC
PAaB, BaC |- **P**AaC
AaB, **P**BaC |- **P***(*AaC) (the conclusion here is "possible not in accordance with our definition")

He says the first two are "complete" ("evident from the definition of possible"). On a *de re* interpretation, Barbara PAP should be valid for the same reason as Barbara NAN. However, in the case of Barbara PPP, the reasoning that could be applied to Barbara NNN does not apply, since **P**(*p*) does not entail *p*. Instead (and remarkably), Aristotle adds the stipulation that "A possibly to every B" is to be understood as "A belongs to everything to which B *possibly* belongs." On this understanding, Barbara PP amounts to **P**Aa**P**B, **P**Ba**P**C |- **P**Aa**P**C, which is valid as a non-modal Barbara. However, in the case of Barbara PNP, Aristotle instead appears to treat the major premise as **P**AaB:

> For since C is below B and A is possible to every B, it is evident that it will be possible to every C.
>
> (33b34–36)

Because of this, the interpretation of Aristotelian possibility sentences is unclear: does he prefer the ampliated reading, or the unampliated reading, or does he switch between the two—and if so, on what principles?

Aristotle's treatment of Barbara AP is considerably more complex, and indeed it is perhaps the most intriguing, as well as the most difficult, section of his modal syllogistic. He begins by arguing for the following principle:

> If when A is B then it is necessary for B to be, then if A is possible B will also of necessity be possible.
>
> (34a5–7)

This might be translated into modern formal logic in two ways:

$P \rightarrow Q$ |- $\Diamond P \rightarrow \Diamond Q$
$\Box (P \rightarrow Q)$ |- $\Diamond P \rightarrow \Diamond Q$

This is the only place in the syllogistic where Aristotle uses letters as sentential variables rather than predicate variables, and it also seems to countenance arguments with a single premise. In fact, Aristotle takes note of the latter point immediately, explaining that we should really understand *P* here as the premises of a syllogism (since, on his view, nothing follows from just one premise). It is also the only passage in the modal syllogistic to which Aristotle makes any reference outside *Prior Analytics* I.8–22 itself: he states it in *Metaphysics* Θ 3, in arguing against the Megarian position that only what is necessary is possible.

Having stated his principle, Aristotle offers a proof for it (34a7–12) and then puts it to use in arguing for Barbara AP. His argument is an unusually structured proof through impossibility. He first derives another principle from it:

> If we assume as true something which is false but not impossible, then whatever follows from that assumption may be false but will not be impossible.
>
> (34a225–27)

He then uses this principle to construct a proof for Barbara AP:

> With these distinctions made, then, let A belong to every B but let B be possible to every C. Then it is necessary for A to be possible to every C. For suppose not, and suppose B as belonging to every C (this is false but not impossible). Then since A is not possible to every C but B belongs to every C, A will not be possible to every B (for a syllogism comes about through the third figure). But it was assumed to be possible to belong to every B. Then necessarily A will be possible to every C, since when something false but not impossible was assumed the result was impossible.
>
> (34a34–b2)

As Aristotle notes, the conclusion here is the denial of the *reductio* hypothesis "A is not possible to every C," which he interprets as NAeB; consequently, it does not express "possibility in accordance with our definition" but rather the denial of NAeB, that is M(AeB). Both the nature of the principle Aristotle uses here and the structure of his argument for APM are exceptionally difficult to interpret (see Waterlow 1982, Fine 2011, Rini 2011, Malink & Rosen 2013).

Models of the modal syllogistic

An extensive literature on the interpretation of Aristotle's modal logic has developed over the past eight decades. I note here only a selection of these. The earliest such interpretation (Becker 1933) argued that Aristotle in effect employed the distinction between *de re* and *de dicto* readings of modal sentences without realizing it, leading to an inconsistency in his system (Becker also saw other inconsistencies). McCall (1963) gives a formal interpretation of a subset of Aristotle's results based on a non-formal interpretation advanced by Nicholas Rescher; Johnson (1994) gives a semantics for it. Nortmann (1996) models Aristotle's system in predicate logic and argues that its necessity is S4 necessity. Patterson (1995) gives a comprehensive but less formal reconstruction the foundation of which is an interpretation of Aristotelian modal sentences as containing "modal copulas." While most of these models make corrections in Aristotle's system to obtain coherence or limit themselves to a subset of his results, some more recent reconstructions aim at reproducing as much as possible and at modeling his proofs as well as his results. Rini (2011) does so by explaining apparent inconsistencies in Aristotle's understanding of modals on the basis of a semantic distinction of terms. Malink (2013) succeeds in reproducing every detail of Aristotle's results, though at the cost of much greater complexity.

References

Anscombe, Gertrude Elizabeth. 1956. "Aristotle and the Sea Battle," *Mind* 65: 1–15.

Becker, Albrecht, 1933. *Die Aristotelische Theorie der Möglichkeitsschlüsse*. Berlin: Junker und Dunnhaupt.

Bobzien, Susanne. 1993. "Chrysippus' Modal Logic and Its Relation to Philo and Diodorus," 63–84 in K. Döring and T. Ebert, eds., *Dialektiker und Stoiker* (Stuttgart: F. Steiner).

Bobzien, Susanne. 2011. "Dialectical School." in Edward N. Zalta, ed., *The Stanford Encyclopedia of Philosophy* (Fall 2011 edition), URL: https://plato.stanford.edu/archives/fall2011/entries/dialectical-school/.

Corcoran, John. 1972. "Completeness of an Ancient Logic," *Journal of Symbolic Logic* 37: 696–705.

Corcoran, John. 1974. "Aristotle's Natural Deduction System". in J. Corcoran, ed., *Ancient Logic and Its Modern Interpretations*. Dordrecht: D. Reidel, 85–131.

Döring, Klaus. 1972. *Die Megariker: kommentierte Sammlung der Testimonien*. Amsterdam: B. B. Grüner Verlag.

Döring, Klaus. 1989. "Gab es eine Dialektische Schule?" *Phronesis* 34: 293–310.

Ebert, Theodor. 2008. "In Defence of the Dialectical School." in F. Alesse, ed., *Anthropine Sophia. Studi di filologia e storiografica filosofica in memoria di Gabriele Giannantoni*. Naples: Bibliopolis, 275–293.

Fine, Kit. 2011. "Aristotle's Megarian Manoeuvres," *Mind* 120: 993–1034.

Frede, Dorothea. 1970. *Aristoteles und die "Seeschlacht." Das Problem der contingentia futura in De Interpretatione 9*. Göttingen: Vandenhoeck & Ruprecht.

Hintikka, Jaakko. 1964. "The Once and Future Sea Fight: Aristotle's Discussion of Future Contingents in *De Interpretatione* IX," *Philosophical Review* 73: 461–492.

Hintikka, Jaakko. 1973. *Time and Necessity: Studies in Aristotle's Theory of Modality*. Oxford: Clarendon Press.

Johnson, Fred. 1994. "Apodictic Syllogisms: Deductions and Decision Procedures," *History and Philosophy of Logic* 16: 1–18.

Malink, Marko. 2013. *Aristotle's Modal Syllogistic*. Cambridge, MA: Harvard University Press.

Malink, Marko, and Jacob Rosen. 2013. "Proof by Assumption of the Possible in *Prior Analytics* I.15," *Mind* 122*:* 953–986.

McCall, Storrs. 1963. *Aristotle's Modal Syllogisms*. Amsterdam: North Holland.

Nortmann, Ulrich. 1996. *Modale Syllogismen, mögliche Welten, Essentialismus: eine Analyse der aristotelischen Modallogik*, Berlin: De Gruyter.

Nortmann, Ulrich. 2002. "The Logic of Necessity in Aristotle: An Outline of Approaches to the Modal Syllogistic, Together with a General Account of *de dicto*- and *de re*-Necessity," *History and Philosophy of Logic* 23: 253–265.

Patterson, Richard, 1995. *Aristotle's Modal Logic: Essence and Entailment in the Organon*. Cambridge: Cambridge University Press.

Prior, Arthur. 1955. "Diodoran Modalities." *Philosophical Quarterly* 5: 205–213.

Rini, Adriane. 2011. *Aristotle's Modal Proofs: Prior Analytics I.8–22 in Predicate Logic*. Dordrecht: Springer.

Sedley, David. 1977. "Diodorus Cronus and Hellenistic Philosophy," *Proceedings of the Cambridge Philological Society* 203: 74–120.

Sedley, David. 2018. "Diodorus Cronus," in Edward N. Zalta, ed., *The Stanford Encyclopedia of Philosophy* (Winter 2018 edition), URL: https://plato.stanford.edu/archives/fall2018/entries/diodorus-cronus/.

Thom, Paul. 1996. *The Logic of Essentialism. An Interpretation of Aristotle's Modal Syllogistic*. Dordrecht: Kluwer.

Waterlow, Sarah. 1982. *Passage and Possibility: A Study of Aristotle's Modal Concepts*. Oxford: Oxford University Press.

Whitaker, C. W. A. 1996. *Aristotle's De Interpretatione: Contradiction and Dialectic*, Oxford: Clarendon Press.

Chapter 30

MODALITY IN MEDIEVAL PHILOSOPHY

Stephen Read

The metaphysics of modality

The synchronic conception of possibility and necessity as describing simultaneous alternatives is closely associated with Leibniz and the notion of possible worlds.[1] The notion seems never to have been noticed in the ancient world, but is found in Arabic discussions, notably in al-Ghazali in the late eleventh century, and in the Latin west from the twelfth century onwards, elaborated in particular by Scotus. In Aristotle, the prevailing conception was what Hintikka (1973: 103) called the "statistical model of modality": what is possible is what sometimes happens.[2] So what never occurs is impossible and what always occurs is necessary. Aristotle expresses this temporal idea of the possible in a famous passage in *De Caelo* I 12 (283b15 ff.), and in *Metaphysics* Θ4 (1047b4–5):

> It is evident that it cannot be true to say that this is possible but nevertheless will not be.
>
> (Aristotle 2006: 5)

However, Porphyry. in his *Introduction* (*Isagoge*) to Aristotle's logic (which he took to be Plato's), included the notion of inseparable accidents, giving the example of the blackness of crows and Ethiopians.[3] Neither, he says, is necessarily black, since we can conceive of a white crow, even if all crows are actually black. Whether this idea is true to Aristotle is uncertain.[4] Nonetheless, Aristotle seems to have held that once a possibility is realized, any alternatives disappear. So past and present are fixed and necessary, and what never happens is impossible. If the future alone is open—that is, unfixed and possible—then once there is no more future, nothing is possible. Hintikka (1973: 156) quotes Aristotle's observation in the famous ninth chapter of *De Interpretatione* (18b12–15):

> But if it was always true to say of what now is that it was so, or would be so, it could not not be so, or not be going to be so. But if something cannot not happen … it is necessary for it to happen.
>
> (Aristotle 1963: 50–1)

Whether or not this is a fallacy, for Hintikka it serves to explain Aristotle's endorsement of the statistical model.[5] The idea is that the instantiation of contrary potentialities cannot be simultaneous but must be successive. Hence if one potency is instantiated forever, the other can never be. Whatever is possible (or potential) must be realized at some time. One might worry that talk of something existing (or having existed) forever would be anathema to Aristotle, who famously rejected the completed infinite, but this is how he talks in *De Caelo* I 12 (281b25): "Hence everything which exists forever is absolutely undestructible" (Aristotle 1939: 115).

Knuuttila (2016) suggests that the conception of a simultaneous alternative present started to take hold in the Latin west from the twelfth century, under the influence of Augustine's views on divine power and freedom. Augustine and Boethius were the main bridges from the ancient world, which came to an end in Europe in the sixth century, to the revival of learning in the twelfth. Augustine, writing in the fourth and early fifth centuries, was largely uninfluenced by Aristotelianism. Boethius set out to translate Aristotle's works into Latin (the Roman intellectual world being till then for the most part linguistically Greek) and comment on it—a project cut short after his barely completing it for Aristotle's logic alone when put to death in 524 CE. Knuuttila (2008: 507–31, esp. 517ff.) sees the idea of synchronic alternatives to the present latent in both Augustine and Boethius, ideas which became explicitly recognized in theological discussion only in the twelfth century. Augustine, for example, claimed that some possibilities open to God are never realized, using the slogan "He could but he did not wish it".[6] Boethius, however, in his second commentary on Aristotle's discussion of future contingents in *De Interpretatione* Chapter 9, recognizes only the simultaneous possibilities of the open future:

> Things in the past and the present do indeed have a definite and established outcome; for those that have come about cannot not have come about; and as regards those that are happening now, it cannot happen that while they are happening they are not happening. As regards those that have to do with the future and are contingent, however, something can both happen and not happen.
>
> (Boethius 1998: 148)

Knuuttila (2008: 517) concedes, "Boethius did not develop a theory of simultaneous synchronic possibilities which remain intact even when diachronic possibilities have vanished". He claims that it was not fully recognized in the Latin west until Scotus spelled it out in the early fourteenth century. Scotus wrote:

> I do not call something contingent because it is not always or necessarily the case, but because the opposite of it could be actual at the very moment when it occurs.[7]

In the Arabic world, however, the concept of synchronic alternatives was recognized in the eleventh century, if not earlier. Kukkonen (2000) finds it in al-Ghazali's discussion of the eternity of the world in his *The Incoherence of the Philosophers*. The clash between Aristotle's arguments for the eternity of the world and religious doctrines of the creation arose in both the Islamic and the Christian worlds. But they surfaced earlier in the Islamic tradition since their program of translation of Aristotle's works occurred in the ninth century, whereas Boethius's similar exercise was not resumed in the Latin west until the twelfth century and completed only in the thirteenth. Al-Ghazali's claim of "incoherence" (*tahafut*, stumble or bankruptcy[8]) was directed specifically at Avicenna, who had repeated and elaborated Aristotle's arguments in a number of

works written in the early eleventh century. In 1095, al-Ghazali set out to show the error in this reasoning. Take the third proof that al-Ghazali discusses (cited in Averroes 1952: 57): suppose the world were created at some instant. Then it was possible that it be created, and had always been possible. So it was possible that it was created earlier. Indeed, for any time at which it might have been created, it might have been created earlier. But then we have an actual infinity of earlier possibilities, which is impossible. Hence it must have existed forever, and so is eternal.

The argument goes back, once again, to Aristotle's *De Caelo* I 12 (283a11–18):

> Why generated now, when for an infinite time [the world] had not existed? If there is no reason at all, and the possible points of time are infinite in number, then clearly there existed for an infinite time something susceptible to generation … If then we suppose its powers to be realized, both opposites will be present to it simultaneously.
>
> (Aristotle 1939: 123)

The nub of the argument is that if the world was created at some time, it could have been created earlier (given God's free choice), but possibility is actuality at some time, so it must have been created earlier. Contradiction. Thus, it has existed forever. The argument depends crucially on the temporal or statistical interpretation of possibility, that there is a single series of possibilities. Al-Ghazali sensed this, and his rebuttal turns on rejecting this conception and envisaging alternative synchronous possibilities. Just as the world could have been larger than it is, it could have been created at a different time. These possibilities do not aggregate into a single eternal past but represent separate options between which God can choose. Al-Ghazali preserves the revealed truth of creation by rethinking the nature of counterfactual possibility.

Returning to Scotus, this synchronous conception of alternative possibilities is what Knuuttila (1993: 143) claims justifies the title "possible world semantics". His argument turns on Scotus's talk of compossibility and incompossibility. "According to Scotus, that is logically possible the formulation of which does not contain a contradiction, and possibilities are partitioned into groups of possible states of affairs on the basis of relations of compossibility" (Knuuttila 1993: 138). These possibilities are objective and independent: God chooses between them rather than creating them from nothing:

> This logical possibility could remain separately in power by its own nature even if there were, *per impossibile*, no omnipotence to which it would be an object.
>
> (Scotus 1963, dist. 36: 296)

Wyatt (2000) contests this attribution to Scotus of the recognition of a domain of possible worlds constituting a semantics against which to interpret modal sentences. Nonetheless, she concedes (2000: 210) that Scotus maintains that not only is the future open, but so too is the present. Aristotle's claim that "Everything which is, when it is, is necessary" (*De Interpretatione* ch. 9, 19a23) is ambiguous. Aristotle commits a fallacy of composition and division, says Scotus (1963, dist. 39: 419), in claiming that the present is fixed and necessary. This fallacy was, ironically, one that Aristotle himself described in *De Sophisticis Elenchis*:[9]

> "A man can walk while sitting and can write while not writing". For the meaning is not the same if one divides the words and if one composes them in saying "It is possible to walk-while-writing".
>
> (Aristotle 1928: 166a23)

Aristotle's statement in *De Interpretatione* Chapter 9 is similarly ambiguous. The fallacy is to move from the necessity of the inference (*necessitas consequentiae*) to that of the consequent (*necessitas consequentis*). Necessarily, if it is, then it is; but that does not warrant the inference that if it is, then it necessarily is:

> So no true sense of this proposition ["Everything which is, when it is, is necessary"] means that being something—at the moment when it is—is necessary, but only that it is necessary relative to when it is, since at the moment when it is, it is still simply contingent, and consequently at that moment its opposite could still be.
>
> (Scotus 1963 dist. 39: 423)[10]

So although I cannot be actually doing anything inconsistent with what I am doing, I could be.

Scotus considers a further argument for the necessity of the present which appeals to a particular rule in obligational disputations—a genre peculiar to the late middle ages, developed throughout the thirteenth century and prevalent in numerous philosophical works in the fourteenth. An obligational disputation starts with a *casus*, a hypothetical situation, and a particular proposition, the *positum*, usually false in the given *casus*. The opponent puts a series of propositions to the respondent, who must respond by granting or denying them respectively, if they are relevant in either following from or being inconsistent with the *positum* and other propositions previously granted (or the negations of those denied); if not relevant, granting and denying them if they are known to hold or not hold in the *casus*, while doubting them if their truth is unknown.[11] The argument Scotus considers appeals to the rule that if the *positum* is false, it must be denied that now is the present moment (since the respondent must consistently maintain the truth of the *positum*).[12] Scotus's rejection of the rule again shows that he rejects the necessity of the present:

> The proof is not valid, for although what is posited should be maintained as true, yet it can be maintained for that instant without denying that the instant is one for which it is false, because the inference does not hold, "It is false for the instant, therefore it is impossible".
>
> (Scotus 1963: 423)

Knuuttila (1993: 150) argues that the series of propositions granted by the respondent starts to describe a possible world. However, this series of responses is often brief, hardly answering to Wyatt's demand (2000: 212) that reason be given to suppose that Scotus envisages anything like a complete description of how things are—i.e., a possible world. Rather, the *casus* which is nearly always present in an obligational disputation makes a better argument for Knuuttila's claim. The *casus* is envisaged as an entire situation, alternative to how things actually are, since the *positum* is often in fact true but taken to be false in the *casus*. The anonymous author of the *Ars Obligatoria* in Kretzmann and Stump (1985), in giving three reasons why "*casūs* are to be posited", observes that

> the third reason is so that we may admit something false [but] possible while finding out what follows in order that we may know what to do and how to respond when things are in fact as the false *casus* indicates.
>
> (Kretzmann and Stump 1985: 251–2)

The false *casus* makes a much better candidate for an alternative possible world, representing all that the respondent knows to be the case on some false assumption, and used as it is to evaluate the truth of irrelevant propositions proposed by the opponent.

Even if Wyatt is right to question whether Scotus's observations amount to a "detailed [semantic] theory" (Knuuttila 2016), George Hughes (1989: 97) found it implicit in Buridan's system of modal logic, developed in the 1330s:

> I want to suggest here … that we might understand what [Buridan] says in terms of modern "possible world semantics". Possible world theorists are quite accustomed to talking about possible worlds in which there are more horses than there are in the actual world. And then, if Buridan assures us that by "Every horse can sleep" he means "Everything that is or can be a horse can sleep", we could understand this to mean that for everything that is a horse in any possible world, there is a (perhaps other) possible world in which it is asleep. It seems to me, in fact, that in his modal logic he is implicitly working with a kind of possible worlds semantics throughout.

The crucial point is that nothing that is not a horse could be a horse. "Horse" is a substance term. The only possible and non-actual horses are pure possibilia.

Hughes concluded his article with an invitation "to try to give a Kripke-style possible worlds semantics for Buridan's modal system" (Hughes 1989: 108). That was recently done in Johnston (2015). But, as Hodges and Johnston (2017: 20) say, to infer from the fact that this semantics fits his logical system that it correctly reflects Buridan's intentions is dangerous. Rather, what that fact can do is add weight to what Hughes inferred from Buridan's own metaphysical views.

The logic of modality

Buridan was one of the few logicians in the Middle Ages who developed his own system of modal logic, distinct from and independent of Aristotle's system. Most others felt constrained to follow Aristotle's lead, even if they gave it their own semantic basis, and struggled in the attempt. Indeed, Buridan's system was so successful that it dominated discussion of modal logic for the next two centuries, replacing Aristotle's hegemony.[13] Another major medieval logician who developed his own system was Peter Abelard, writing in the early twelfth century, before Aristotle's treatise *Prior Analytics* was recovered in the Latin west through Boethius's translation. Thus, Abelard could develop his ideas independently of Aristotle, though many of these were eclipsed when Aristotle's account became known. A third major thinker who developed a modal system independent of Aristotle's was Avicenna, writing in what is now Iran, on the edge of the Islamic world, in the early eleventh century, though he was well aware of Aristotle's modal logic, and of its shortcomings.

Aristotle's modal system is described in Chapters 8–22 of *Prior Analytics*, chapters obviously inserted later, presumably by Aristotle himself, into an existing work: the start of Chapter 23 clearly follows straight on from the end of Chapter 7 (and Chapter 3 is a later insert, discussing modal conversions). Aristotle takes his theory of non-modal (so-called "assertoric" or "categorical") syllogisms and considers which syllogistic moods remain valid when one or more premises and the conclusion are modified by the modes "necessary" and "possible". A complication is that he distinguishes two senses of "possible", so-called "one-way" and "two-way" possibility, what we now call respectively "possible" (the dual of "necessary", that is, what is not necessarily not so) and "contingent" (what is neither necessary nor impossible). Moreover, he seems to allow "possible" in the sense of "contingent" to modify either the premises or the conclusion (or both), but considers "possible" only as the dual of "necessary" to modify the conclusion, not the premises.

In *Prior Analytics* Chapter 8, Aristotle observes that if the premises of any assertoric, two-premise syllogism are modified by "necessary", then a conclusion modified by "necessary" may be validly inferred in all cases where the corresponding assertoric conclusion follows from the unmodified premises. In Chapters 9–11, he considers mixed premises, one of necessity, one assertoric, in each figure, summarizing the results in Chapter 12. In Chapter 13 he discusses possibility, noting the validity of complementary conversion, namely, that any affirmative statement modified by "contingent" is equivalent to the corresponding negative statement so modified; e.g., "Every *A* is contingently *B*" converts to "No *A* is contingently *B*", and "Some *A* is contingently *B*" converts to "Not every *A* is contingently *B*".[14] In subsequent chapters he considers what follows modally or categorically from mixtures of premises of necessity, of contingency, and assertoric.

Even Aristotle's immediate successors in the Lyceum, e.g., Theophrastus, head of the school after Aristotle's death, found Aristotle's account of the modal syllogism problematic in ways that his non-modal system was not. Was there an understanding of the modal terms which supported the claims of validity he had made, or should those claims be revised? There seemed to be many inconsistencies, but the most immediate concerned his claim in Chapter 9 that a conclusion of necessity follows in the first figure from mixed premises, one of necessity, the other assertoric, if the premise modified by "necessary" is the major premise (that is, the premise containing the predicate of the conclusion), but not if it is the minor premise. Although the point applies to all four valid first-figure moods, the problem has come to be called the "problem of the two Barbaras", that is, inferring a universal affirmative of necessity from modally mixed universal affirmatives. Theophrastus went so far as to deny any such validity, advocating the so-called "ad peiorem" rule, that the conclusion can never be modally stronger than the weaker premise.[15]

Much ink was spilled over the next two thousand years trying to make sense of this,[16] but the boldest thinkers broke free of Aristotle's yoke and developed a modal system *ab initio*. First, we need to clarify what is covered by "mode" in "modal logic". Aristotle, as we saw, deals only with the aletheic modes "necessary, "possible" and "contingent". There are also temporal modes, "past", "present" and "future", and the epistemic modes, such as "know", "believe", "doubt". Buridan considers the epistemic modes in Q40 of the first book of his *Questions on Aristotle's Prior Analytics*. He starts by distinguishing compound from divided uses of modals:[17]

> I call "compound" ones like "I know that a man is seated", "That a man is a stone is believed, or apparent" and so on; but I call "divided" those like "A man I know to be an animal", "An ass appears to me to be a man", "God is believed to be triune" and so on.
>
> (Buridan n.d.)

Only trivial inferences follow with compound modals, he says: for one can know (or believe, or doubt) the premises of a (valid) syllogism without knowing (etc.) the conclusion. But since knowledge entails truth, if one knows the premises, the conclusion will certainly be true. Moreover:

> If you have arranged the premises into a syllogism, and you have considered the conclusion, and you know the syllogism is a necessary consequence, it follows that you will immediately know the conclusion.
>
> (*loc. cit.*)

Turning to the divided sense, Buridan claims that any valid first- or third-figure syllogism remains valid if the major premise and conclusion are modally qualified in the same way, and the minor is left assertoric:

> A modal conclusion follows by the *dictum de omni et nullo*, e.g., "Every *B* is supposed to be *A*, every *C* is *B*, so every *C* is supposed to be *A*.
>
> (*loc. cit.*)

Recall that Aristotle had made the same claim where the major and conclusion are modified by "necessary". Buridan denies this for the aletheic modal cases with universal major, for he takes universal propositions of necessity to be ampliated to the possible. Again, like the distinction between compound and divided senses, the notion of ampliation was suggested by Aristotle himself but was developed and taken much further by the medievals. At *Prior Analytics* 32b25ff., Aristotle had observed that

> "this possibly belongs to that" may be understood in two ways—either of what that belongs to, or of what that may belong to.
>
> (Aristotle 2009: 19)

The medievals were divided on how far to extend this idea. Buridan thought the subjects of universal affirmatives and negatives both of necessity and of possibility (and of contingency) were ampliated by the modal predicate to the possible. Thus, "Every *A* is necessarily *B*" means that every actual or possible *A* is necessarily *B*—and since he followed Aristotle[18] in accepting that the actual was *ipso facto* possible (in the sense of "not impossible"), that amounts to the claim that every possible *A* is necessarily *B*; and similarly, "No *A* is possibly *B*" (that is, "Every *A* is necessarily not *B*") means that no possible *A* is possibly *B*. Thereby, Buridan ensured that "No *A* is possibly *B*" is the contradictory of "Some *A* is possibly *B*" and that "Every *A* is necessarily *B*" is the contradictory of "Not every *A* is necessarily *B*".[19]

Others disagreed. For example, William Ockham followed Aristotle more closely:

> According to Aristotle's opinion, the argument "Every white thing is of necessity a swan, no ass is a swan, so every ass is of necessity not white" is not valid, because if the only white thing is a swan, the premises are true and the conclusion false. For "Every white thing is of necessity a swan" would be true, because each singular would be true, so it is true. And the major is definitely true, and the conclusion is manifestly false. So the argument is not valid.
>
> (Ockham 1974, III-1.32: 448)

Thus, Ockham takes the truth of "Every *A* is of necessity not *B*", that is, "No *A* is possibly *B*", to follow from the truth of all singular propositions "This cannot be *B*", where "this" denotes an actual *A*. This seems to have been Aristotle's view, too,[20] and also that of Abelard for propositions taken *de re* (Thom 2003: 53). Thom (2003: 66, 68) claims Avicenna held that not only are the subjects of modal propositions ampliated to the possible, but also both subject and predicate of assertorics. However, this seems to be based on a misreading, for El-Rouayheb (2016: 75) notes:

> Avicenna was understood [by his successors] to have rejected ampliation to the possible and to have upheld the view that the subject term should be understood to include only that of which it is true in actuality (past, present, or future).[21]

Avicenna was, perhaps, the most original and ambitious logician of the middle ages, not least in developing what Hodges (in preparation: ch. 9) has called his two-dimensional logic, quantifying over times as well as objects. Hodges and Johnston (2017: 15) claim that the resulting logic contains a part that is equivalent to his modal logic, and also to Buridan's logic of divided modals, and they construct a possible-worlds semantics that shows this equivalence.[22] This might suggest that Avicenna accepted the statistical model of modality, equating worlds with times. But Hodges and Johnston (2017: 15) cite Avicenna as denying this identification:

> Being permanent is not the same as being necessary … But it is not for the logician as logician to know the truth about this.[23]

That is, the logical laws of times and worlds are the same, as shown by the model theory, but their metaphysics is different.

Williamson (2013: 45), following Movahed (2016: 34), claims that Avicenna anticipated and endorsed the Barcan formula and its converse, that "Something is possibly *F*" is equivalent to "Possibly, something is *F*".[24] This would force a constant domain across worlds. Movahed (2016: 34) cites Avicenna as saying:

> But to say that *some people are possibly not writers* is modally the same as saying that *possibly some people are not writers*, and although one implies the other the meaning of the one may be opposite to the other.

At first glance, this is puzzling: Avicenna seems to be saying of the two expressions that they are at the same time equivalent and opposite and that one follows from the other. But Bäck's translation makes things a little clearer:

> As for our saying (that) some man can not-be a scribe [Bäck adds in a footnote: "literally, some man it is not possible that he is not a scribe"] it may be modally equivalent to our saying (that) it is possible that some man is not a scribe, while it might differ from it even though it is a consequence of it, so that the intention in one of them is that some man is being characterized by the possibility of the denial of scribehood of him, and in the other that it is possible for the speaker's saying (that) some man is a scribe to hold.
>
> (Avicenna 2013: 142)

Avicenna is saying that "Some man is possibly not a writer" (or scribe) is ambiguous, having two different senses, the composite or *de dicto* ("It is possible that some man is not a writer") and the divided reading, *de re* ("Some man can fail to be a writer"), the former entailing the latter. This is exactly what Ruth Barcan Marcus (1961: 307) intended by her eponymous formula, suggesting that Quine's *bête noire*, the third grade of modal involvement (Quine 1953: 66) was in fact a consequence of the second, acceptable if construed as equivalent to the first. So Avicenna does endorse the Barcan formula, though the passage does not endorse its converse; indeed, it says that the two are different and not equivalent.

The Barcan claim is orthogonal to the question of the statistical interpretation, though one might suspect them to be in tension: its being possible, for example, that someone fly to Mars, even if realized, might not be realized by anyone now alive. But whether Avicenna anticipated the idea of possible worlds, as well as the Barcan formula, is an open question. To paraphrase Williamson (2013: 45):

> We leave to [the future] the hard work of making sense of this tantalizing glimpse in its historical and systematic context.

Notes

1 See, e.g., von Wright (1984: 74, 96).
2 Hintikka took the expression from Bekker (1952: 16), for whom possibilities are cases (Fälle), like rows in a truth-table. Cf. von Wright (1984: 73 n. 1).
3 Porphyry (2003: 12 [§5]).
4 See, e.g., van Rijen (1989: ch. 4), and Aristotle's *Posterior Analytics* I 6 and *Rhetoric* III 17, 1418a3–5.
5 See also Knuuttila (1993: 9–12).
6 "Potuit sed noluit" (Knuuttila 1993: 69; 2008: 518).
7 Scotus, *De Primi Principio* IV.4, cited in Knuuttila (1993: 143; 2008: 550).
8 See Adamson (2016: 147); Averroes (1952: xiii).
9 But it is a fallacy many have attributed to Aristotle in connection with his modal logic—see the following.
10 See also Scotus (1994: 132–7).
11 For more on the game of obligations, see Dutilh Novaes and Uckelman (2016).
12 The rule is found in, e.g., William of Sherwood's *Obligations* (Green 1963, vol. 2: 8; cf. vol. 1: 62)—though the authorship is in doubt—and in Walter Burley's *Obligations* (Green 1963, vol. 2: 59) of 1302: "When a false contingent proposition concerning the present has been posited, one must deny that it is the present instant" (Kretzmann and Stump 1988: 394).
13 See Knuuttila (2016).
14 See, e.g., Buridan (2015: 48, 104–5).
15 See, e.g., Alexander of Aphrodisias (2013, 124, 3–17: 59 and 118–19).
16 And still is: see, e.g., Malink 2013.
17 One case of the compounded/divided distinction is the *de dicto/de re* distinction. The terminology is medieval: see Knuuttila (2008: 533–4) and Lambert of Auxerre (2015: 353 n. 99).
18 *Prior Analytics* I 13.
19 See Buridan (2001: 44–5; 2005: 100)
20 See *Prior Analytics* I 10, 30b20, and Malink 2013: 275.
21 See also the texts from Avicenna cited in Hodges and Johnston (2017: 25–31).
22 This proof is in Hodges (in preparation: §12.3).
23 They refer to *Qiyas* 48.14–16.
24 Buridan, being committed, as we saw, to pure possibilia, rejected the Barcan formula. See Buridan (2015: 35–36).

References

Adamson, P. (2016) *Philosophy in the Islamic World*, Oxford: Oxford University Press.
Alexander of Aphrodisias (2013) *On Aristotle Prior Analytics: I.8–I.13*, trans. Mueller and J. Gould, London: Bloomsbury.
Aristotle (1928) *De Sophisticis Elenchis*, trans. W. Pickard-Cambridge, in W.D. Ross (ed.), *The Works of Aristotle*, vol. I, Oxford: Oxford University Press.
——— (1939) *On the Heavens*, trans. W.K.C. Guthrie, London: Heinemann.
——— (1963) *Categories and De Interpretatione*, trans. J.L. Ackrill, Oxford: Clarendon Press.
——— (2006) *Metaphysics Book Θ*, trans. S. Makin, Oxford: Clarendon Press.
——— (2009) *Prior Analytics Book One*, trans. G. Striker, Oxford: Clarendon Press.
Averroes (1952) *Tahafut al Tahafut* (The Incoherence of the Incoherence), trans. S. van den Bergh, 2 vols., Oxford: Oxford University Press.
Avicenna (2013) *al-ʿIbārā: Avicenna's Commentary on Aristotle's De interpretatione: Part One and Part Two*, trans. A. Bäck, Munich: Philosophia.
Barcan Marcus, R. (1961) "Modalities and Intensional Languages", *Synthese* 13(4): 303–322.
Bekker, O. (1952) *Untersuchungen über den Modalkalkül*, Meisenheim/Glan: Westkulturverlag.
Boethius, A.M.S. (1998) *On Aristotle on Interpretation 9: First and Second Commentaries*, trans. N. Kretzmann, in *On Determinism: Ammonius, on Aristotle on Interpretation 9*, trans. D. Blank, London: Duckworth.
Buridan (n.d.) *Quaestiones in Analytica priora*, ed. H. Hubien, unpublished. URL: http://www.logicmuseum.com/wiki/Authors/Buridan/Quaestiones_in_analytica_priora.
Buridan, J. (2001) *Summulae de Dialectica*, trans. G. Klima, New Haven: Yale University Press.
——— (2005) *Summulae de Propositionibus*, ed. R. van der Lecq, Nijmegen: Ingenium.

——— (2015) *Treatise on Consequences*, trans. S. Read, New York: Fordham University Press.
Dutilh Novaes, C. and Uckelman, S. (2016) "Obligationes", in C. Dutilh Novaes and S. Read (eds.), *The Cambridge Companion to Medieval Logic*, Cambridge: Cambridge University Press, pp. 370–95.
El-Rouayheb, K. (2016) "Arabic Logic after Avicenna", in C. Dutilh Novaes and S. Read (eds.), *The Cambridge Companion to Medieval Logic*, Cambridge: Cambridge University Press, pp. 67–93.
Green, R. (1963) *An Introduction to the Logical Treatise "De Obligationibus", with Critical Texts of William of Sherwood (?) and Walter Burley*, PhD thesis, Université Catholique de Louvain.
Hintikka, J. (1973) *Time and Necessity: Studies in Aristotle's Theory of Modality*, Oxford: Clarendon Press.
Hodges, W. (in preparation) *Mathematical Background to the Logic of Ibn Sina*, URL: http://wilfridhodges.co.uk/arabic44.pdf.
Hodges, W. and Johnston, S. (2017) "Medieval Modalities and Modern Methods: Avicenna and Buridan", *IFCoLog Journal of Logic and Its Applications*.
Hughes, G.E. (1989) "The Modal Logic of John Buridan", in G. Corsi et al. (eds.), *Atti del Convegno Internazionale di Storia della Logica: Le Teorie delle Modalità (San Gimignano, 5-8 dicembre 1987)*, Bologna: Editrice CLUEB, 93–111.
Johnston, S. (2015) "A Formal Reconstruction of Buridan's Modal Syllogism", *History and Philosophy of Logic* 36(1), pp. 2–17.
Knuuttila, S. (1993) *Modalities in Medieval Philosophy*, London: Routledge.
——— (2008) "Medieval Modal Theories and Modal Logic", in D. M. Gabbay and J. Woods (eds.), *Handbook of the History of Logic*, vol. 2: *Mediaeval and Renaissance Logic*, Amsterdam: North-Holland, pp. 505–78.
——— (2016) "Medieval Theories of Modality", in E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2016 edition), URL: https://plato.stanford.edu/archives/win2016/entries/modality-medieval/.
Kretzmann, N. and Stump, E. (1985) "The Anonymous *De arte obligatoria* in Merton College m.s. 306", in E. P. Bos (ed.), *Mediaeval Semantics and Metaphysics*, Nijmegen: Ingenium, pp. 239–80.
——— (1988) "Walter Burley, *Obligations* (Selections)", in N. Kretzmann and E. Stump (eds.), *The Cambridge Translations of Medieval Philosophical Texts*, vol. 1: *Logic and the Philosophy of Language*, Cambridge: Cambridge University Press, pp. 369–412.
Kukkonen, T. (2000) "Possible Worlds in the *Tahâfut al-Falasifa*: Al-Ghazâlî on Creation and Contingency", *Journal of the History of Philosophy* 38(4): 479-502.
Lambert of Auxerre (2015) *Logica or Summa Lamberti*, trans. T. S. Maloney, Notre Dame: University of Notre Dame Press.
Malink, M. (2013) *Aristotle's Modal Syllogistic*, Cambridge, MA: Harvard University Press.
Movahed, Z. (2016) "Ibn Sina's anticipation of the formulas of Buridan and Barcan", in Z. Movahed (ed.) *Reflections of the Logic of Ibn Sina and Suhrawardi*, Tehran: Hermes, pp. 25–38.
Ockham, Guillelmi de (1974) *Summa Logicae*, in P. Boehner et al. (eds.), (Opera Philosophica I), St Bonaventure, NY: Franciscan Institute Publications.
Porphyry (2003) *Introduction*, trans. J. Barnes, Oxford: Clarendon Press.
Quine, W.V. (1953) "*Three Grades of Modal Involvement*", *Proceedings of the 11th International Congress of Philosophy*, vol. 14, North-Holland, pp. 65–81.
Scotus, J. Duns (1963) *Ordinatio: Liber Primus, a distinctione vicesima sexta ad quadragesimam octavam*, in P. Carolo Bali (ed.), Opera Omnia, vol. VI, Vatican City: Typis Polyglottis Vaticanis.
Scotus, J.Duns (1994) *Contingency and Freedom*, trans. A. Vos Jaczn et al., Dordrecht: Kluwer.
Thom, P. (2003) *Medieval Modal Systems*, Aldershot: Ashgate.
van Rijen, J. (1989) *Aspects of Aristotle's Logic of Modalities*, Dordrecht: Kluwer.
Williamson, T. (2013) *Modal Logic as Metaphysics*, Oxford: Oxford University Press.
von Wright, G. (1984) *Truth, Knowledge and Modality*, Oxford: Blackwell.
Wyatt, N. (2000) "Did Duns Scotus Invent Possible Worlds Semantics?", *Australasian Journal of Philosophy* 78(2): 196–212.

Further readings

Knuuttila, S. (2008) "Medieval Modal Theories and Modal Logic", in D. M. Gabbay and J. Woods (eds.), *Handbook of the History of Logic*, vol. 2: *Mediaeval and Renaissance Logic*, Amsterdam: North-Holland, pp. 505–78. (Detailed survey article by the major authority on the subject.)

Scotus, J. Duns (1994) *Contingency and Freedom*, trans. A. Vos Jaczn et al., Dordrecht: Kluwer. (Contains a translation, with introduction and commentary, of dist. 39 of Book I of Scotus' *Lectura*, the original version of his discussion of contingency, later elaborated by Scotus, or his followers, in the *Ordinatio*.)

Thom, P. (2003) *Medieval Modal Systems*, Aldershot: Ashgate. (A detailed presentation and comparison of the modal systems of Abelard, Avicenna, Averroes, Robert Kilwardby, Richard Campsall, William Ockham and Buridan.)

Chapter 31

MODALITY IN DESCARTES'S PHILOSOPHY

Alan Nelson

31.1 Introduction

Descartes's distinctive use of the concepts of necessity and possibility drew both vigorous support and denunciation in the 17th and 18th centuries.[1] To this day, it remains a topic of interest among historians of philosophy; in the last fifty years more than a hundred scholarly articles have been published on the topic. The main focus has been on what has come to be called Descartes's Doctrine of the Creation of the Eternal Truths (Creation Doctrine for short). Here is a brief preliminary sketch. What Descartes calls "eternal truths" (he gives examples of things naturally taken as necessary truths) are said to be freely created by God. Otherwise, they would have some kind of existence that did not depend on God's creative power, which is unacceptable to Descartes. But if the eternal truths are freely created, then by making use of omnipotence, God could have made them false and other ones true. But then it seems absolutely everything must be possible. Even the eternal truth, "contradictories cannot be true together" seems to be rendered possibly false. Further unacceptable consequences obviously ensue.

In what follows, an interpretation is proposed that builds on recent scholarship and includes no consequences that Descartes would have considered unacceptable. This will require a little background in Cartesian metaphysics, which is provided in the next section. Some of the strangeness of Descartes's treatment of modal notions (only some of it) is removed by seeing how they fit into his system of philosophy. Section 31.3 is about the contrast between necessary and contingent existence which is, in some ways, a special case. In Section 31.4 the question of the ontology of eternal truths is taken up in preparation for considering the sense in which they are necessary in Section 31.5. Section 31.6 then takes up the general interpretation of modality in Descartes's philosophy.

31.2 Some Cartesian metaphysics

Descartes's ontology fundamental ontology is characterized in a passage of utmost importance in his *Principles of Philosophy*, a book which he hoped would be widely adopted as a textbook (CSM1: 209–16).[2] This ontology includes primarily substances: the infinite substance, God, and finite thinking and extended substance. To be a substance is to be independent from other

substances. Because finite substances depend on God for their existence, only God is a substance in a very strict sense of the term (CSM1: 210). Substances have *attributes*, and finite substances also have *modes*. The term 'mode' is used to refer to a way in which the substance it depends on exists at some time. The term 'attribute' is used "when we are simply thinking in a more general way of what is in a substance" (CSM1: 211). Modes are similar to what philosophers now think of as properties of objects. An important consequence of this is that God's having the attribute of immutability means that God has no modes. But what are attributes?

To get at Descartes's concept of an attribute, his version of the theory of distinctions must be brought in. The famous *real distinction* between thinking and extended substance is one of these. Important for the matter at hand is the *rational distinction*, sometimes instead less literally translated as 'conceptual distinction.'[3] Attributes and the conceptual distinction are intimately linked. Descartes writes that,

> a *conceptual distinction* is a distinction between a substance and some attribute of that substance without which the substance is unintelligible; alternatively, it is a distinction between two attributes of a single substance.
>
> (CSM1: 214)

The most salient example of an attribute for finite substances is its essence. For minds it is thinking, and for matter it is extension. Descartes calls these the principal attributes of finite substances. Other attributes include existence, duration, and unity (CSM1: 45, 214). God has the familiar infinite attributes of omnipotence, omniscience, omni-etc., simplicity, and so on. These infinite attributes are most relevant to the topic of modality in Descartes's philosophy. Regarding the linked notion of conceptual distinction, Descartes scholars have come up with two basic, competing interpretations. One can be called the identity interpretation. On this reading, the distinction is a product of the concepts used by the human mind in thinking about substances. So there is no distinction in substances themselves, between substances and their attributes, or among the attributes themselves. This is the reading required to make the best sense of Descartes on modality, and it will be assumed in what follows.[4] There is an obvious, albeit incomplete, analogy between the conceptual distinction among attributes and ordinary names of substances. The names 'Cicero' and 'Tully' are distinct, but they name one thing. Cicero and Tully are identical. Moreover, the distinction between the names is no distinction at all in Cicero, the man. God's attributes are analogously distinguished only conceptually with their being no real differentiation in the infinite substance.

31.3 Necessary and contingency existence

Like the other attributes, existence is equivocally predicated in finite and infinite cases. It is not that the Eiffel Tower and God both *exist*, albeit one contingently and the other necessarily. Regarding the term 'substance,' Descartes says that "there is no distinctly intelligible meaning of the term which is common to God and his creatures" (CSM1: 210). Attributes are only conceptually distinct from substances, and we take this to mean that the attribute in the substance is identical to it (CSM1: 214; Nolan 1997). So the equivocality of 'substance' entails the equivocality of 'existence' between the infinite and finite cases. This is textually confirmed by Descartes's regarding the following as an axiom.

> Existence is contained in the idea or concept of every single thing, since we cannot conceive of anything except as existing. Possible or contingent existence is contained

> in the concept of a limited thing, whereas necessary and perfect existence is contained in the concept of a supremely perfect being.
>
> (CSM2: 117; see also CSM1: 306; CSM2: 83)

Let's mark this by distinguishing (finite) existence from (infinite) necessary-existence. So, for example, in Descartes's version of the ontological argument is that necessary-existence is proven to be an attribute of God, not the kind of existence we enjoy. Of course Descartes does sometimes write things that can be translated into English as, "it is necessary that God exists." This should be read as "it is necessary that God necessarily-exists." We shall consider the meaning of 'it is necessary that' in what follows. For now, we can register the special modal contrast between God's necessary-existence and the existence (or contingent-existence) enjoyed by finite things.

31.4 The ontological status of the eternal truths

The application of Descartes's theory of attributes to God is even more explicit in the following statements which do not directly bear on the necessity of God's existence.

> In God willing and knowing are a single thing in such a way that by the very fact of willing something he knows it and it is only for this reason that such a thing is true.
>
> (CSMK: 24)

> In God, willing, understanding, and creating are all the same thing without one being prior to the other even conceptually.
>
> (CSMK: 25–6)

Quotes like the first have encouraged commentators to use the term 'voluntarism' for Descartes's Creation Doctrine. In virtue of God's attribute of infinitely free will, the truths are freely chosen. One might be tempted to draw the conclusion that had God's free choices gone differently, the created truths would have been different. And, as we have seen, God is perfectly independent, so the choice of eternal truths cannot be constrained by their being true prior to God's creating them. Nor can they be true in virtue of being divine thoughts because of the conceptual distinction among the infinite attributes. As Descartes writes in these two passages, willing, knowing, and creating are identical in God. If God knows a truth, it is willed to exist. Here the attribute is omnipotence, because creation is an exercise of power. These quotes also show that the label 'voluntarist' does not accurately describe Descartes's treatment if that term is understood to involve some kind of priority of the will over the intellect. In the infinite substance, what we call omnipotence, omniscience, perfect freedom, independence, are all identically the substance itself. In fact, Descartes singles out simplicity as an attribute that is especially important for understanding God (CSM1: 201; CSM2: 34, 98; CSMK: 235).

The eternal truths must be created because they cannot be known by God without creative power being exercised on them. This follows from the identity of knowing and creating. Scholars have wondered about the manner in which eternal truths exist as created beings. A natural, but unsatisfactory, suggestion is that they are created as something like Fregean third realm entities. This reading is encouraged by the Fifth Meditation's famous introduction of "true and immutable natures." These seem at first to qualify as third realm entities for at least two reasons.

> When, for example, I imagine a triangle, even if perhaps no such figure exists, or has ever existed, anywhere outside my thought, there is still a determinate nature, or

> essence, or form of the triangle which is immutable and eternal, and not invented by me or dependent on my mind. This is clear from the fact that various properties can be demonstrated of the triangle, for example that its three angles equal two right angles.
>
> (CSM2: 45)

This text seems to say that immutable and eternal truths about triangles follow from a created essence that does not depend for its existence on anything triangular (presumably bodies with triangular faces) or on the mind. A closer look, however, presents a different picture. Another text from the crucial passage in the *Principles* locates eternal truths in human minds.

> All the objects of our perception we regard either as things, or affections of things, or else as *eternal truths, which have no existence outside our thought*.
>
> (CSM1: 208, emphasis added)

The conflict between this and the previous quotation is merely apparent. Eternal truths reside in the mind but are not dependent on the mind because minds themselves are created by God supplied with ideas enabling them to think these truths. Moreover, these truths correspond to ordinary substances (triangles are extended and extension is the essence of matter) and do not correspond to third realm items. How then could they be true "even if perhaps no such figure exists"? The answer is that the 'perhaps' in the quote is epistemic and not metaphysical. In the context of the Fifth Meditation, one is still applying methodological doubt to the hypothesis that extended things exist. One can nevertheless think eternal truths about them.[5] Descartes's point is that even at this stage of the *Meditation*'s method of doubt, one can think these truths.[6] Some have suggested that eternal truths could reside entirely in God's mind (Rozemond 2008), but as we have seen, this is incompatible with Descartes's very explicit commitment to the strict identity of the divine attributes. A final worry about the metaphysical status of eternal truths might come from Descartes's using examples of such axiomatic (for him) truths such as "nothing has no properties" and "he who thinks cannot but exist while he thinks." This worry is alleviated by remembering that these truths have a nominalist basis in truths about particulars (Nolan 1998). So, for example, the particular truth "I think, therefore I am" is prior to the general truth "he who thinks, exists" (Wilson 1978: ch. II.2)

The conclusion is that God creates eternal truths by creating our minds such that we are able to think them. This has led some commentators to think that the way necessity and possibility work in connection with the Creation Doctrine must have something to do with concepts as opposed to being intrinsic modal features of the world. We now turn to this line of thought.

31.5 The necessity of the eternal truths

For Descartes, necessary truths and eternal truths are co-extensive. Is that a verbal point, or are there reasons for the eternal truths being necessary? This section attempts to give reasons. One thing that is often noted is that there is some connection between immutability and necessity. Curley (1984) and Osler (1995) find the connection between God's immutability and the immutability of created eternal truths.[7] This makes sense especially in light of Descartes's famously deriving his conservation principle for quantity of motion using God's immutability as a premise (CSM1: 240). Then, given that God is immutable, we can be sure that the items we understand as necessary are understood to have always been thus and to remain so eternally.[8] Otherwise, God's operation in the world would change (inconsistent with immutability) and

God would be a deceiver (inconsistent with omni-benevolence). This suggests another way to characterize truths that are necessary. They are the truths that are perceived entirely "clearly and distinctly." The famous Fourth Meditation result that whatever is clearly and distinctly perceived is true applies to purely intellectual matters that are understood with maximal certainty. This 'truth rule' applies strictly only in metaphysical contexts in which we are prescinding from sensory input. This means that clarity and distinctness are scaled, so that some very certain contingent items can be considered true and certain for philosophical convenience. Contingent propositions do not have the maximal truth and certainty enjoyed by necessary truths, once the latter are clearly and distinctly perceived. It seems, then, that while Descartes never gives an outright definition of 'necessity,' he does manage helpfully to make some thick conceptual connections.

One might object to the preceding paragraph on the grounds that it has Descartes failing to distinguish our perceptions of things as necessary with *necessity itself*. That might be pursued as an global objection to Descartes's system of philosophy, but there are strong reasons for thinking that what makes necessary truths about finite things count as necessary is our being able to clearly and distinctly perceive them. An important text for interpreting the Creation Doctrine and the nature of the necessity obtaining in necessary truths is in the second set of replies to objections to the *Meditations*.

> If by 'possible' you mean what everyone commonly means, namely '*whatever does not conflict with our human concepts*,' then it is manifest that the nature of God, as I have described it, is possible in this sense, since I supposed it to contain only what, *according to our clear and distinct perceptions*, must belong to it; and hence it cannot conflict with any of our concepts. Alternatively, you may well be imagining some other kind of possibility which relates to the object itself; but unless this matches the first sort of possibility it can never be known by the human intellect, and so it … [will] undermine the whole of human knowledge.
>
> (CSM2: 107, emphasis added)

We don't need to bother here with Descartes's assertion that this is what is commonly meant by 'possible'; what matters here is that it likely is how he usually uses it himself. As J. Bennett (1994) points out in his conceptualist treatment of Descartes on modality, it is *human* concepts in terms of which we understand claims about possibilities.[9] A possibility that one hypothesizes to be attached to things outside thought is otiose if it cannot actually be thought. Or even worse, the general strategy of hypothesizing inconceivable things will be destructive of "the whole of human knowledge"—a skeptical worry that Descartes thinks he has discharged in the *Meditations*.

Two things should be added here to what Bennett finds in the text. First, in his example of the clearly and distinctly perceived nature of God, Descartes would say that nature is not only possible, but necessary. Possibility understood as opposed to necessity—as in the preceding quotation—pertains more naturally to perceptions that are not clear and distinct.[10] Lacking clear and distinct perception, we might not perceive a conflict actually present in our concepts. So, for example, we might imagine that a chimera is a possible being, but if we knew more, we might find a conflict in how the lion's head and goat's body are understood to link up (CSM2: 83–4; also see CSMK: 343). Second, because of the focus on concepts, Descartes always takes propositions to be perceptions of relations among conceptual items. When Descartes writes in the quote "conflicts with human concepts," the verb in Latin is *repugnat*. The same word is sometime translated as 'contradict.' Descartes uses *repugno*-words and *contradico*-words interchangeably. We shall return to this point.

31.6 Modality and the Creation Doctrine

A common way of stating a problem for the Creation Doctrine comes from the familiar reasoning that a literally omnipotent God can do *anything*. That might seem to include making contradictions true, which is bad enough. But it might also be taken to imply that such a God could alter its own nature, or self-annihilate, etc. This has produced much discussion about whether Descartes can justify an exception for eternal truths about God that is not objectionably *ad hoc* (Curley 1984 and the sources cited there). This worry misconstrues the context of Descartes's Creation Doctrine. As we saw in Section 31.2 of this chapter, there is a fundamental ontological division between God and creation such that no term or concept applies univocally to both. The Creation Doctrine in no way entails that God is freely self-created. God's immutable, eternal nature includes what we clearly and distinctly understand it to be along with aspects beyond our nature to comprehend. It does not follow from 'God necessarily-exists' being understood by us as a created eternal truth, that God's necessary-existence is subject to omnipotence. In God, the attributes of omnipotence and necessary-existence are identical. Recall the analogy with being Cicero and being Tully. When, therefore, we say that God can do "anything," the scope of the quantifier is restricted to items that are actually created. Another point to note is that if we mean to instantiate "anything" and say something certainly true, we must clearly and distinctly perceive it. If we say with certainty that it is possible for God to make a chimera, we would have to clearly and distinctly perceive a chimera, including the detailed way in which the head and body are connected. On the other hand, because God's infinity exceeds our conception of it,

> I would not dare to say that God cannot make a mountain without a valley, or that one and two should not be three. I merely say that he has given me such a mind that I cannot conceive a mountain without a valley, or an aggregate of one and two which is not three, and that such things involve a contradiction in my conception.
>
> (CSMK: 358–59)

In this letter, Descartes does not go so far as to say it is possible that God made 1 + 2 equal some particular number other than three. The point is put in a negative way. He does not say that God can do things; he says that he *dare not say* that God can *not* do things. The double negative point is theologically prudent and based on the clear and distinct perception of what God did do in making 1 + 2 = 3. The "dare not" prevents the double negative from collapsing into a positive. Note that 1 + 2 = 17, e.g., cannot be understood at all, so there is in that case no particular item in question.

We "dare not" or ought not to say that it is impossible for God to do things that exceed our grasp. That is just to acknowledge God's infinity insofar as it is incomprehensible to a finite mind. But, to put the final point of the last paragraph in another way, it does not follow that it is possible for God to do something fitting a very specific description.[11] We are certain we have a hold of a specific thing in God's power only when we clearly and distinctly perceive it. Here is another example of an initially puzzling passage that can be interpreted this way.

> You ask also what necessitated God to create these truths; and I reply that he was free to make it not true that all radii of the circle are equal—just as free as he was not to create the world. And it is certain that these truths are no more necessarily attached to his essence than are other created things.
>
> (CSMK: 25)

Again, Descartes does not say God could make it true that circles have some radii that are L in length and others that are 2L. He says that God could make it *not* true that they are the same.

It is interesting that he backs this up with the possibility that God not create at all. Remembering that for God, being free to do X is identical to having the power to do X, this quote says only that God could do something or other that we do not understand. There are a few places in Descartes's philosophy where he writes that there are some exceptional, particular things we know that God does that are beyond our comprehension or even conflict with what we do understand. These special cases fit his general stricture that we must not doubt what we do clearly and distinctly perceive because of things we do not so perceive.

> I also understand that it would have been easy for God to ordain certain things such that we men cannot understand the possibility of their being otherwise than they are. And therefore it would be irrational for us to doubt what we do understand correctly just because there is something which we do not understand and which, so far as we can see, there is no reason why we should understand.
>
> (CSM2: 294)

Examples of these special cases include God's unifying really distinct minds and bodies in human beings (CSM2: 294), the compatibility of divine providence and our free will (CSM1: 206) and the fact that matter is actually divided without limit into ever smaller bits (CSM1: 239). One might think this is bad philosophy, but it fits comfortably with Descartes's holding that certain judgments of possibility and necessity derive from the innate capacities for thought with which we have been created. Let's consider an example in which this reasoning is applied directly to the Creation Doctrine.

Bennett writes that there is only one text that poses a hard problem for a conceptualist understanding of Descartes on modality. He calls it a "hot potato" (1994: 657), and it does seem subtly different from the ones considered and cited earlier.

> I turn to the difficulty of conceiving how God would have been acting freely and indifferently if he had made it false that the three angles of a triangle were equal to two right angles, or in general that contradictories could not be true together. It is easy to dispel this difficulty by considering that the power of God cannot have any limits, and that our mind is finite and so created as to be able to conceive as possible the things which God has wished to be in fact possible, but not be able to conceive as possible things which God could have made possible, but which he has nevertheless wished to make impossible. The first consideration shows us that God cannot have been determined to make it true that contradictories cannot be true together, and *therefore that he could have done the opposite.*
>
> (CSM3: 235, emphasis added)

This passage is smooth sailing on a conceptualist reading up until the last clause of the last sentence. God's freedom can in no way be constrained, his will is infinite, his being is independent, his power is limitless, and so on, as the theory of attributes tells us. So God cannot be determined to make it true that contradictories cannot be true together, and we dare not say that he might not have. The problem is that in saying "he could have done the opposite," it seems Descartes is making a positive claim about some specific inconceivable thing that God could have done.

Bennett tries to downplay this text on the grounds that it is an outlier written in a long letter on many topics which, Descartes seems to suggest, he wrote hurriedly. I think, however, that this text can be handled by bringing in some of the background Cartesian philosophy that Bennett does not consider in his article. In the first place, it is wrong to think of the truth, "contradictories

cannot be true together" as a specific entity created by God. As noted in Section 31.4, Descartes thinks that general principles are abstractions from specific instances. A particular thing cannot be both round and square. The example of a round square also recalls the fact from Section 31.5 that contradiction itself is to be understood in terms of the conflict between concepts. So when Descartes writes in the quoted passage that God could do the opposite of what he has actually done, i.e. make contradictories not true together, he is *not* saying there is some specific, clearly and distinctly perceivable necessary truth that is possibly false. One cannot get at such a clear and distinct perception by appealing in general and indirectly to what is its "opposite." A specific opposite would need positively to be clearly and distinctly perceived, and that cannot be done.

31.7 Conclusion

So far I have been using tools of Cartesian metaphysics and epistemology to buttress and extend what Bennett calls a conceptualist theory of modality. It is, perhaps, not quite right to call it a *theory*. Many of our contemporaries who work on theories of modality would not regard it as a theory because of its extremely reductive character. Metaphysical modalities are reduced to facts about how we operate with the specific, actual concepts we have. It might be better to say that Descartes's philosophy does not have any use for modal concepts that apply to "objects themselves," as he says in CSM2 (107), quoted in Section 31.5. It might be appropriately characterized as a no-theory theory (Nelson and Cunning 1999).

This chapter has connected Descartes's thinking on this topic to theses developed elsewhere in this philosophy—the extraordinary reliance on clear and distinct perceptions, for example. But what about cases in which philosophers want to consider the modality of items that, unlike eternal truths, are not completely certain or not clearly and distinctly perceived in Descartes's special sense? Consider "Trump might have lost the election." For Descartes, these cases are also to be handled by taking things that do not conflict with our concepts as possibilities. Taken just as concepts, *Trump* and *losing the election* do not seem to conflict. That is all we need to understand "Trump might have lost the election." Even in these cases in which our concepts are not clear and distinct, Descartes does not need to appeal to any metaphysics of modality.

Notes

1 For the positive reaction, see Schmaltz (2002); for a Leibnizian critique, see Nelson (1993); and for a general overview of its reception, see Lennon (1998).
2 CSM(1/2/K), abbreviate the three volumes of the standard English translation in Cottingham et al. (1984–91).
3 In this chapter, I shall stick with the latter, CSM translation.
4 A great deal of textual and philosophical evidence for the identity interpretation of attributes is in the seminal Nolan (1997). The other main interpretation is that the attributes of a substance are not identical to it or to one another, but merely inseparable from one another (see, e.g., Hoffman 2002). It is much harder to give a coherent account of Descartes's views on modality if attributes are treated as non-identical. This comes out in Rozemond (2008). For an overview of the debate about attributes and conceptual distinction, see Sowaal (2016).
5 Descartes gives arithmetical examples of eternal truths, but he regards algebraic truths as being abstracted from truths about extension (Nolan 1998).
6 For a full treatment and a review of the literature, see Cunning (2010: ch. 6).
7 Kaufman (2005) argues that a strong reading of immutability comes with a necessitarian destruction of contingent truth. It will emerge in the following text that the conception of necessity that drives the argument is not Descartes's. On this topic also see Frankfurt (1977), and Normore (1986, 1991).
8 It is awkward on any interpretation that the eternal truths are created by divine fiat. That makes them "less eternal" than God. That is no surprise, however, given that nothing is attributed univocally to God and creatures. God and the eternal truths are *eternal* in different senses.

9 Bennett also notes that aspects of a conceptualist reading are found in Wilson (1978), Ishiguro (1986), and L. Alanen (2008). Those authors want to hold on to the difference between conceptual modalities and "real" necessity and possibility, so their suggestions are not as thoroughgoing as Bennett's.
10 For more discussion of how Descartes deploys his principle that clearly and distinctly perceived things are true, see Nelson and Cunning (1999).
11 Bennett has a good account here (1994: 653ff).

References

Alanen, L. (2008) "Omnipotence, Modality, and Conceivability," in J. Broughton and J. Carriero (eds.) *A Companion to Descartes*, Oxford: Blackwell.
Bennett, J. (1994) "Descartes Theory of Modality," *Philosophical Review* 103: 639–67.
Cottingham, J., Stoothoff, R., Murdoch, D., and Kenny, A. (eds. and translators) (1984–91) *The Philosophical Writings of Descartes*, vol. 3, Cambridge: Cambridge University Press.
Cunning, D. (2010) *Argument and Persuasion in Descartes' Meditations*, New York: Oxford University Press.
Curley, E. (1984) "Descartes on the Creation of the Eternal Truths," *Philosophical Review* 93: 569–97.
Frankfurt, H. (1977) "Descartes on the Creation of the Eternal Truths," *Philosophical Review* 86: 36–57.
Hoffman, P. (2002) "Descartes's Theory of Distinction," *Philosophy and Phenomenological Research* 64: 57–78.
Ishiguro, H. (1986) "The Status of Necessity and Possibility in Descartes," in A. Rorty (ed.) *Essays on Descartes's Meditations*, Berkeley: University of California Press.
Kaufman, D. (2005) "God's Immutability and the Necessity of Descartes's Eternal Truths," *Journal of the History of Philosophy* 43: 1–19.
Lennon, T. (1998) "The Cartesian Dialectic of Creation," in M. Ayers and D. Garber (eds.) *The Cambridge History of Seventeenth-Century Philosophy*, Cambridge: Cambridge University Press.
Nelson, A. (1993) "Cartesian Actualism in the Leibniz-Arnauld Correspondence," *Canadian Journal of Philosophy* 23: 675–94.
Nelson, A. and D. Cunning (1999) "Cognition and Modality in Descartes," *Acta Philosophica Fennica* 64: 137–53.
Nolan, L. (1997) "Reductionism and Nominalism in Descartes's Theory of Attributes," *Topoi* 16: 129–40.
Nolan, L. (1998) "Descartes's Theory of Universals," *Philosophical Studies* 89: 161–80.
Normore, C. (1986) "Meaning and Objective Being: Descartes and His Sources," in A. Rorty (ed.) *Essays on Descartes's Meditations*, Berkeley: University of California Press.
Normore, C. (1991) "Descartes's Possibilities," in G. Moyal (ed.) *Rene Descartes*, vol. 3, London: Routledge.
Osler, M. (1995) "Divine Will and Mathematical Truth: Gassendi and Descartes on the Status of the Eternal Truths," in R. Ariew and M. Grene (eds.) *Descartes and His Contemporaries*, Chicago: University of Chicago Press.
Rozemond, M. (2008) "Descartes's Ontology of the Eternal Truths," in P. Hoffman, D. Owen & G. Yaffe (eds.) *Contemporary Perspectives on Early Modern Philosophy: Essays in Honor of Vere Chappell*, Peterborough, Ontario: Broadview Press.
Schmaltz, T. (2002) *Radical Cartesianism*, Cambridge: Cambridge University Press.
Sowaal, A. (2016) "Attribute," in L. Nolan (ed.) *The Cambridge Descartes Lexicon*, Cambridge: Cambridge University Press.
Wilson, M. (1978) *Descartes*, London: Routledge.

Chapter 32
HUME ON MODALITY

Peter Millican

David Hume is famous for his treatment of causation, which has inspired so many later philosophers towards "regularity" accounts of causal modality. It features prominently in his first and most comprehensive philosophical work, the *Treatise of Human Nature* of 1739–40, whose longest part (Book 1 Part 3) is largely framed around his analysis of our idea of causal necessity. Moreover, there is evidence that Hume's early philosophical interests were particularly sparked by this topic, which had wider resonance given its association with contentious religious issues such as the Cosmological Argument, Hobbist materialism, determinism, free-will, and the Problem of Evil. It also seems likely that the specific inspiration for Hume's approach to causation came from the prospect of applying Lockean conceptual empiricism to the crucial idea of causal necessity: to identify the empirical source of that idea, and thus to shed light on its nature and its application to these important wider issues.[1]

Hume also recognises a form of *conceptual* modality, quite distinct from *causal* modality, and closely associated with his oft-used *Conceivability Principle* (that *whatever is conceivable, is possible*). But unlike causal modality, Hume never investigates the origin of these other modal ideas or applies his conceptual empiricism to them, and he says relatively little on the relationship between the two forms of modality. So putting together a unified Humean account of modality requires some reading between the lines and charitable reconstruction.

32.1 Empiricism and modality

Hume's conceptual empiricism is trumpeted very early in the *Treatise*, in the form of what he calls his "first principle" (*T* 1.1.1.12), now generally known as his Copy Principle:

> *That all our simple ideas in their first appearance are deriv'd from simple impressions, which are correspondent to them, and which they exactly represent.*
>
> (*T* 1.1.1.7)

Here an "idea" is the content of a thought (and a "simple idea" is an atomic component of such content), while an "impression" is a sensation or feeling. Hume thus takes our thinking to consist, quite literally, of material copied from our external sensations or inner feelings (and accordingly derived from our faculties of *sensation* or *reflection*). Our thought is therefore *imagistic*, in a

broad sense that embraces all perceptual modalities, not just vision. Hume gives two empirically based arguments for the principle. First, he claims that as a matter of observation, "the rule … holds without any exception, and that every simple idea has a simple impression, which resembles it; and every simple impression a correspondent idea" (*T* 1.1.1.5; cf. *T* 1.1.1.8, *E* 2.6). Secondly, he argues that anyone who lacks a particular sense modality – for example, sight or hearing – and is accordingly deprived of the relevant impressions, is thereby also unable to acquire the relevant ideas; likewise, unless we have some specific simple impression, we will lack also the corresponding idea (e.g. the taste of a pineapple, *T* 1.1.1.9; cf. *E* 2.7).

Neither of these arguments seems particularly compelling. The first gives the appearance of simply begging the question against anyone who denies the principle (by claiming that they do in fact have a non-imagistic idea of, say, an indeterminate triangle, or the self, or property inhesion in a substance; cf. *T* 1.3.1.7, 1.4.6.2, 1.4.5.6). The second argument seems to conflate the question of whether someone *could have some idea in their mind* (e.g. a colour or a taste) with the question of whether they *could express or identify that idea to others*: how would we know if a blind man had coloured thoughts, and how can we be so sure that he does not? Hume's own confidence on the issue suggests that he rather took conceptual empiricism for granted, persuaded by the dominant Lockean orthodoxy.[2] But he was perceptive enough to see the inadequacy of John Locke's own account of the idea of causal *power* in particular, criticising this explicitly both in the *Treatise* (*T* 1.3.14.5) and in the 1748 *Enquiry concerning Human Understanding*:

> Mr. Locke, in his chapter of power, says, that, finding from experience, that there are several new productions in matter, and concluding that there must somewhere be a power capable of producing them, we arrive at last by this reasoning at the idea of power. But no reasoning can ever give us a new, original, simple idea; as this philosopher himself confesses. This, therefore, can never be the origin of that idea.
>
> (*E* 7.8 n. 12)

It was Hume's own attempt to replace this account that led him through the rich range of topics covered in *Treatise* Book 1 Part 3, from the basis of the Causal Maxim (*T* 1.3.3), to inductive inference in humans and animals (*T* 1.3.6, 16), the psychology of belief (*T* 1.3.7–10), probability (*T* 1.3.11–13), and finally his famous analysis of the idea of causal power or necessary connexion (*T* 1.3.14–15).[3]

32.2 Seeking the "impression" of causal necessity

Hume's analysis of the idea of causal necessity – unlike Locke's – evinces clear appreciation of the serious problem that modal ideas pose for conceptual empiricism. For such ideas concern not *what is in fact the case* – and might thus plausibly be apprehended by empirical perception – but rather *what must be* or *what might be the case*. How can any sequence of actual sensory input, whether external or internal, possibly give conceptual access to such modal thoughts? Hume's celebrated answer, at least in the case of causation, is that we acquire the idea of power or necessary connexion from our own inferential behaviour in response to observed regularities.

Hume's argument to this conclusion proceeds by elimination, and the overall structure is essentially the same in both *Treatise* 1.3.14 and *Enquiry* 7, though the latter is more streamlined. In the *Treatise* he starts by rejecting Locke's account of the origin of the idea of power (*T* 1.3.14.5), inferring that "since reason can never give rise to the idea", it "must be deriv'd from experience, and from some particular instances … which make their passage into the mind by

the common channels of sensation or reflection" (*T* 1.3.14.6). He then briefly attacks scholastic and Cartesian views, before getting to the crucial point: that no impression of power – whether sensory or reflective – can be found in single instances of causal interactions, either material (*T* 1.3.14.10–11) or mental (*T* 1.3.14.12). The *Enquiry* discussion covers similar ground, but is more clearly structured around a systematic search for the elusive impression (*E* 7.5), starting from "external objects … in any single, particular instance" (*E* 7.6) and moving on to "reflection on the operations of our own minds" (*E* 7.9). Here the critique of Locke, which does not fit neatly into this structure, is relegated to a footnote (as quoted earlier). And the primary critical emphasis is now against a quite different theory, that our idea of power is derived from internal reflection on the voluntary power that we exercise over our own bodies (*E* 7.10–15) or minds (*E* 7.16–19).[4] Hume then goes on to attack Nicolas Malebranche's occasionalism (*E* 7.21–5), culminating in the point that bringing God into the discussion is of no help, since any idea of power in the divine mind could only be derived from an idea of power in our own mind.

Having eliminated the hypothesis that we might discover an impression of power or necessary connexion in *single instances* of causal interactions (whether physical or mental), Hume quickly finds the source of the crucial impression when he turns his attention to *repeated instances* (*T* 1.3.14.15–20; *E* 7.26–8). Here he appeals to his earlier discussion of inductive inference, which concluded that such inference is not a matter of rational insight but an instinctive response to observation of "the constant conjunction of any objects", whereby "we always draw an inference from one object to another" (*T* 1.3.6.3). Having seen *A*s followed by *B*s repeatedly, and then seeing another *A*, we find ourselves irresistibly expecting a *B*, not because of any reasoning or other cognitive achievement, but simply because nature has given us this mental tendency, which Hume calls *custom* or *habit* (*T* 1.3.7.6, 1.3.8.10–14, *E* 5.5–6).[5] It is this "determination of the mind" or "customary transition", from the impression of *A* to the expectation or "lively idea" of *B*, that somehow constitutes the impression of causal power or necessary connexion.[6]

Hume then goes on to give what he (later) calls two "definitions of *cause*" (*T* 2.3.2.4, *E* 8.27), informed both by this identification of the relevant impression and by the circumstances under which it characteristically arises. The first definition focuses on these circumstances, defining a cause as "*an object, followed by another, and where all the objects, similar to the first, are followed by objects similar to the second*". The second definition focuses on the inferential "impression", defining a cause as "*an object followed by another, and whose appearance always conveys the thought to that other*" (*E* 7.29; cf. *T* 1.3.14.31).[7] The first definition is the basis for seeing Hume as a *regularity theorist* about causation, *reducing* causal relations to objective regularities between kinds of event, and this has been his most influential legacy in the philosophy of causation. The second definition has proved far less popular and harder to interpret in any straightforward way. Indeed it seems to threaten Hume's own theory with incoherence, because not only is it different in detail from the first definition (so the two can diverge), but also – in contrast to a regularity theory – it seems to make causation observer-dependent.

Related threats of incoherence and subjectivism derive more directly from Hume's account of the "impression" of necessary connexion, which seems problematic both psychologically and philosophically. One obvious point is that a "determination of the mind" or "customary transition" – apparently a causal process or a change in mental state – seems very unlike what Hume supposes an "impression" to be: namely a sensation or feeling. Barry Stroud (1977, pp. 85–6) suggests that he must be postulating some distinctive simple feeling which accompanies inductive inference and provides the impression in question. But against this, Hume never states explicitly that there is any such "third perception" between the impression of *A* and the enlivened idea of *B*. Moreover, such a claim would seem hard to reconcile with what he does say about the immediacy and insensibility of inductive inference (*T* 1.3.8.2, 13; 1.3.12.7), and it is

not clear why any such simple feeling – even if it did happen to *accompany* inductive inference – would thereby be uniquely qualified to provide the required impression. This would, apparently, constitute just another constant conjunction, involving a feeling whose connexion with the causal process is no more intimate than that of the various other candidate impressions that Hume himself dismisses (*T* 1.3.14.12; *E* 7.9–19, 7.15 n. 13).

Philosophically, a far more attractive resolution is to see Hume's "impression" as involving reflective *monitoring* of the mental transition, rather than a distinct *feeling*. This would contrast with his usual view of reflection,[8] but has a clear precedent in John Locke's treatment of reflection in his *Essay Concerning Human Understanding* (Locke 1690, II i 4). In the *Enquiry*, moreover, Hume's language does seem to move in a Lockean direction, with talk of "reflection on the operations of our own minds" (*E* 7.9) and "reflection on our own faculties" (*E* 7.25; cf. also *E* 1.13–14), but no mention of the cruder conception – raw *feeling* of passions – which had dominated the *Treatise*. If Hume's account of the crucial impression is indeed informed by this Lockean perspective, then it becomes relatively easy to understand why he so often writes as though it is, literally, a "determination" of the mind or thought (*T* 1.3.14.20, 22, 29; 1.4.7.5), or a customary "transition of the imagination" (*E* 7.28, 30).[9] For thus interpreted, the "impression" is not simply some *feeling* that happens to accompany inductive inference; rather, it is *reflective awareness* of such inference taking place: of the very transition itself. Whether this sort of mental monitoring can plausibly qualify as an "impression" in Hume's sense is debatable, but if he was thinking along these lines, this would neatly explain why he takes such "inference of the understanding" to be a "connexion, that we can have … comprehension of" (*E* 8.25), since this form of reflection would enable us to grasp the inference *as a movement of the mind from A to B*, rather than just as a succession of distinct perceptions.[10] This in turn could explain why Hume sees a crucial insight here, potentially solving the empiricist conundrum of how causal concepts can be acquired by experience: we understand causal connexion – one event's being inferable from another – by reference to our own activity of drawing such inferences. This also potentially alleviates the threat of subjectivism, because ascribing a relation of *inferability* between events *A* and *B* looks far more plausible than projecting some subjective *feeling* onto the "objects" concerned. Many readers have understood Hume's theory as involving some such extreme subjectivism, but thus interpreted, it looks farfetched and even ridiculous. We shall return to this matter shortly, but first, it will be helpful to introduce a quite different aspect of Hume's theory of modality.

32.3 Conceptual modality

Alongside *causal* necessity, Hume also recognises what he sometimes calls *metaphysical* or *absolute* necessity, though he makes no attempt to analyse this in terms of his empiricism (by seeking an impression from which the idea could be derived), and gives far less explicit attention to it. This form of modality nevertheless plays a major role in his philosophy through his well-known *Conceivability Principle*, which he expresses or uses around 30 times. For example:

> 'Tis an establish'd maxim in metaphysics, *That whatever the mind clearly conceives includes the idea of possible existence*, or in other words, *that nothing we imagine is absolutely impossible*.
>
> (*T* 1.2.2.8)

> whatever we *conceive* is possible, at least in a metaphysical sense …
>
> (*A* 11)

Here we see Hume referring explicitly to *absolute* and *metaphysical* possibility (while *T* 1.3.7.3 and 1.3.14.35 refer to the corresponding *necessity*), but most often he omits any such qualification:

> Whatever can be conceiv'd by a clear and distinct idea necessarily implies the possibility of existence …
>
> (*T* 1.2.4.11)

> To form a clear idea of any thing, is an undeniable argument for its possibility …
>
> (*T* 1.3.6.5)

Since Hume's terminology is not consistent (as we shall see further below), it seems appropriate to use the non-Humean term *conceptual* possibility for this, given its intimate relationship with the Conceivability Principle. Hume's most systematic treatment of it is at the beginning of Section 4 of the first *Enquiry*, where he coins the term "matters of fact" for propositions whose truth and falsehood are both possible in this sense, and he opposes these to "relations of ideas" which can be known to be necessarily true a priori because their falsehood is inconceivable. This explicit distinction – commonly known as *Hume's Fork* – brings a welcome clarity to the logical theory of the *Enquiry*, hugely improving upon the rather confused logic of the *Treatise*, which had been based on a sketchy and ultimately incoherent theory of relations.[11]

Even in the *Enquiry*, however, there is scope for confusion about Humean modality, because most of his references to *necessity* are to *causal* necessity, while most of his references to *possibility* are to *conceptual* possibility. So when reading Hume, "necessity" should normally be interpreted *causally*, and "possibility" *conceptually*, generating a potential tension with the standard equivalence between *necessarily*(*P*) and *not*(*possibly*(*not*(*P*))). But the tension is only in Hume's language, not his thought, and he recognises this equivalence both as regards conceptual modality:[12]

> We can never demonstrate the necessity of a cause to every new existence, or new modification of existence, without shewing at the same time the impossibility there is, that any thing can ever begin to exist without some productive principle …
>
> (*T* 1.3.3.3; see also 1.3.7.3)

and as regards causal modality:

> It is universally allowed, that matter, in all its operations, is actuated by a necessary force, and that every natural effect is so precisely determined by the energy of its cause, that no other effect, in such particular circumstances, could possibly have resulted from it.
>
> (*E* 8.4; see also *T* 2.2.2.27)

Another potential source of confusion is that the two types of modality can be mixed, as when we pursue the conceptual implications of what we take to be (causally) necessary laws. One example of this would be in applied mathematics (cf. *E* 4.13). More prominently, Hume repeatedly denies the possibility of forming certain kinds of idea, based on the logical implications of the Copy Principle, which, as we saw in Section 32.1, he takes to be an empirical, causal truth about the human mind:

> Now since nothing is ever present to the mind but perceptions, and since all ideas are deriv'd from something antecedently present to the mind; it follows, that'tis impossible for us so much as to conceive or form an idea of any thing specifically different from ideas and impressions.
>
> (*T* 1.2.6.8, cf. 1.4.5.19)

> It is impossible, therefore, that the idea of power can be derived from the contemplation of bodies, in single instances of their operation; because no bodies ever discover any power, which can be the original of this idea.
>
> (*E* 7.8)

Finally, we need to be wary of Hume's sometimes loose or unphilosophical language, for example, using "absolute necessity" to mean either *causal determination* (*T* 1.3.14.33, 1.4.1.7, 1.4.5.28) or human *need* (*T* 2.3.7.7), and talking of the "absolute impossibility" of miracles (*E* 10.27). The last of these, in particular, has provided scope for critics to allege inconsistency, but even if Hume's language might be considered inappropriately dogmatic here, it is clear that he has in mind *causal* rather than *conceptual* impossibility in such contexts.[13]

The most fundamental Humean theme relating the two types of modality is the inability of *conceptual* necessities to tell us anything at all about *causal* necessities, from which it follows that only *experience* can inform us of causal relations:

> Whatever is absurd is unintelligible; nor is it possible for the imagination to conceive any thing contrary to a demonstration. But … in reasonings from causation, … this absolute necessity cannot take place, and the imagination is free to conceive both sides of the question …
>
> (*T* 1.3.7.3)

> there are no objects, which … without consulting experience, we can determine to be the causes of any other; and no objects, which we can certainly determine in the same manner not to be the causes. Any thing may produce any thing. Creation, annihilation, motion, reason, volition; all these may arise from one another, or from any other object we can imagine. … Since therefore 'tis possible for all objects to become causes or effects to each other, it may be proper to fix some general rules, by which we may know when they really are so.
>
> (*T* 1.3.15.1–2, cf. 1.4.5.30, 1.4.5.32)

> The mind can always *conceive* any effect to follow from any cause, and indeed any event to follow upon another: whatever we *conceive* is possible, at least in a metaphysical sense … It follows, then, that all reasonings concerning cause and effect, are founded on experience …
>
> (*A* 11, 13)

> matter[s] of fact … are … incapable of demonstration. … The proposition, which affirms it not to be, however false, is no less conceivable and intelligible, than that which affirms it to be. … The existence, therefore, of any being can only be proved by arguments from its cause or its effect; and these … are founded entirely on experience. If we reason *a priori*, any thing may appear able to produce any thing. … It is only experience, which teaches us the nature and bounds of cause and effect …
>
> (*E* 12.28–9)

Thus, the distinction between the two types of modality is crucial to Hume's philosophical system: it is the absence of any *conceptual* necessity between distinct events that leaves the field entirely open for empirical constant conjunctions to determine what *causal* necessities actually obtain in the world.[14]

32.4 Hume's apparent modal subjectivism

We have seen in Section 32.2 how Hume's identification of the "impression" of necessary connexion, and his corresponding second definition of cause, raise a threat of subjectivism, apparently making causal relations and causal necessity dependent on the feelings or inferential reactions of an observer.[15] This seems both philosophically unattractive and also dubiously consistent with the objectivist spirit of his first definition (not to mention various other aspects of his philosophy, to be discussed in Section 32.5). But nevertheless, in what we might call the "purple paragraphs" of the *Treatise* between his identification of the crucial impression and his presentation of the two definitions, Hume seems wholeheartedly to endorse modal subjectivism through a series of notoriously extravagant pronouncements. Some of these state explicitly *that the necessity and power of causes are in the mind, not in objects*:[16]

> necessity is something, that exists in the mind, not in objects
>
> (*T* 1.3.14.22)

> power and necessity … are … qualities of perceptions, not of objects, and are internally felt by the soul, and not perceiv'd externally in bodies
>
> (*T* 1.3.14.24)

Others insist *that we cannot even form an idea of necessity and power as qualities of objects*:[17]

> nor is it possible for us ever to form the most distant idea of [necessity], consider'd as a quality in bodies.
>
> (*T* 1.3.14.22)

> it is not possible for us to form the most distant idea of [necessity and power], when it is not taken for the determination of the mind, to pass from the idea of an object to that of its usual attendant.
>
> (*T* 1.3.14.25)

Hume apparently takes these claims to follow straightforwardly from his identification of the impression of power or necessity as a "determination of the mind", on the principle that an idea can only properly represent the impression from which it is copied (stated at *T* 1.2.3.11 and echoed at *T* 1.3.14.6 and 11).

In the midst of these extravagant claims about *causal* necessity, Hume also offers, with seeming nonchalance, an equally subjectivist claim about *mathematical* necessity:

> Thus as the necessity, which makes two times two equal to four, or three angles of a triangle equal to two right ones, lies only in the act of the understanding, by which we consider and compare these ideas; in like manner the necessity or power, which unites causes and effects, lies in the determination of the mind to pass from the one to the other.
>
> (*T* 1.3.14.23)

There is nothing like this in Hume's other writings: elsewhere, the necessity of arithmetic appears to be entirely objective, rather than lying "only in the act of the understanding".[18] Accordingly, most scholars have dismissed this apparently subjectivist claim as either an aberration or a careless gloss on the far more plausible view that *our awareness* of mathematical necessity is derived from an "act of the understanding" (e.g. recognition that the falsehood of the relevant proposition is inconceivable).[19] As remarked in Section 32.3, Hume never explicitly discusses the empiricist origin of our idea of *conceptual* necessity, but it is interesting to note that the potential development in his view of reflection discussed in Section 32.2 – acknowledging mental monitoring as well as feeling – might relatively easily accommodate reflective awareness of *attempting to conceive P without Q*, and presumably also of *failing in this attempt* (and thus being unable to resist inferring *Q* from *P*). As before, accepting such awareness as an "impression" involves significant stretching of the notion, but arguably the Humean theory must anyway allow such stretching if it is to provide a plausible account of our ideas of other mental acts, few of which can be reduced to simple *feeling*. So this could be a charitable interpretative development of his modal theory, providing a unified Humean explanation of the origin of our ideas of causal and modal necessity in terms of reflective mental activity and irresistible inference. But it still leaves us with the potential discomfort of modal ideas that have a merely subjective source, and might for that reason be thought to commit Hume – in respect of both causal and conceptual modality – to the same implausible subjectivism that he expresses in those notorious paragraphs of the *Treatise*.

32.5 Securing causal objectivity

The extreme causal subjectivism of Hume's purple paragraphs is both implausible in itself, and in serious tension with the regularity view that is implicit in the first definition of cause presented only a few paragraphs later. Moreover, the very next section of the *Treatise* is devoted to setting out eight "rules by which to judge of causes and effects" which are clearly elaborations of the first definition (since the first three rules repeat its content, stipulating in turn *contiguity*, *priority*, and *constant union*).[20] Indeed the text introducing these rules could hardly be more objectivist, describing them as "general rules, by which we may know when [objects] really are [causes or effects to each other]" after having explicitly emphasised the principle "*that the constant conjunction of objects determines their causation*" (*T* 1.3.15.2, 1). Later in the *Treatise* Hume refers back to these opening paragraphs (at *T* 1.4.5.30 n. 48), re-emphasising their message and arguing at greater length "that all objects, which are found to be constantly conjoin'd, are upon that account only to be regarded as causes and effects" (*T* 1.4.5.32).

Causal objectivism also seems to lie at the heart of Hume's overall philosophical project. The subtitle of the *Treatise* declares it to be "an Attempt to introduce the experimental Method of Reasoning into Moral Subjects", and Hume takes such reasoning to depend crucially on causal relations (*T* 1.3.2.1–3, *E* 4.4). His researches involve looking for hidden causes behind inconstant phenomena, something he explicitly advocates (*T* 1.3.12.5, *E* 8.13) and for which, as we have seen, he prescribes detailed rules. Moreover, he apparently believes that there is always some correct causal story to be found (whether or not we can find it), since he argues at length that causality applies as much to the moral as the physical world (*T* 2.3.1.5–15 and *E* 8.7–20), denies genuine chance or indifference (e.g. *T* 1.3.12.1, 2.3.1.18; *E* 6.1, 8.25), and is accordingly a convinced determinist.[21] Most of this cited argumentation occurs in the context of Hume's discussions "of liberty and necessity", where his two definitions play a crucial role in delimiting our possible conception of causal necessity (and thus undermining the orthodox view – most notably championed by Samuel Clarke – that the "physical necessity" of mechanical interactions is

quite different from the "moral necessity" that operates in the sphere of human action). Indeed it seems likely that the definition of causation for the purpose of resolving this issue provided much of the motivation for Hume's entire analysis of causation.[22]

The natural interpretation of all this is to see Hume as fundamentally a regularity theorist, whose first definition is unambiguously intended to dominate the second in the ascription of causal relations. Where the first applies but not the second – i.e. uniformities exist without any "inference of the mind" – he urges us to discover the genuine but unknown causes so that we can extend our inferential powers accordingly. Where the second definition applies but not the first – for example, where we exhibit prejudices that fail to reflect genuine uniformities – he urges us to correct our inferential behaviour (*T* 1.3.13.7, 11). The priority of the first definition is also evident in Hume's numerous statements to the effect that "the very essence" of cause and effect, power, or necessity is constituted by "constant conjunction of objects" (*T* 1.4.5.33), "multiplicity of resembling instances" (*T* 1.3.14.16), "constancy" (*E* 8.25 n. 19) or "uniformity" (*T* 2.3.1.10). Only in one passage does he talk of mental inference as constituting such an essence (*T* 2.3.2.2, copied at *E* 8.22 n. 18), but as we shall see, there he means inference by an *idealised* mind which is fully informed of all the relevant facts and uniformities.

We thus have compelling reason for questioning the apparently extreme subjectivist implications of Hume's "impression" of necessary connexion and the associated second definition. And if we ignore the strident subjectivism of the *Treatise* purple paragraphs for the moment, the obvious resolution is to see the impression and second definition as aiming merely to clarify our subjective understanding of causal relations – and in particular the *genesis* of our thought about causation – rather than to specify the objective circumstances in which such relations obtain. Hume's investigation, right from the start, is focused on the *origin* of our idea of causal necessity, and he never suggests that our *application* of that idea should be confined to the circumstances that initially generate it. So a definition that encapsulates the results of such a genetic investigation should not be expected to deliver necessary and sufficient conditions for the relevant term's correct application.[23] Those conditions are sketched instead in the first definition and spelled out more fully in the rules of *Treatise* 1.3.15.

All very well, but doubts may still remain about those purple paragraphs, which seem too conspicuous and emphatic to ignore.[24] It is fortunate, therefore, that Hume himself provided a resolution here, by removing them from his mature treatment of the issue, presumably because he recognised how badly they misrepresented the spirit of his causal theory. Thus in the *Enquiry*, the extravagant subjectivist statements disappear, Hume no longer insists that ideas can only represent impressions, and only once does he suggest that necessity is in the mind rather than in objects:

> The necessity of any action, whether of matter or of mind, is not, properly speaking, a quality in the agent, but in any thinking or intelligent being, who may consider the action; and it consists chiefly in the determination of his thoughts to infer the existence of that action from some preceding objects …
>
> (*E* 8.22 n. 18)

But even this occurs only in a footnote whose topic is peripheral to the theory of causation (namely, "The prevalence of the doctrine of liberty"), and which is copied largely verbatim from *Treatise* 2.3.2.2 with no signs of careful updating. Moreover the note's final sentence clarifies that "the very essence of necessity" is not so much *actual* inference by some particular human mind, but rather *potential* inference by an *idealised* mind which is "perfectly acquainted

with every circumstance of our situation and temper, and the most secret springs of our complexion and disposition". Thus both its context and content rob the passage of virtually all its subjectivist force.

The only other passage in the *Enquiry* that might be thought to imply extreme subjectivism is placed far more prominently, in the paragraph immediately prior to the two definitions of cause:

> When we say, therefore, that one object is connected with another, we mean only, that they have acquired a connexion in our thought, and give rise to this inference, by which they become proofs of each other's existence: A conclusion, which is somewhat extraordinary; but which seems founded on sufficient evidence.
>
> (*E* 7.28)

This is far less explicit than the bold statements in the *Treatise*, and Hume's usage elsewhere gives reason to doubt that his talk here about what "we mean" is to be interpreted as a literal semantic claim.[25] But more crucially, just one paragraph later – after giving his two definitions – Hume states explicitly that we can indeed "mean" something more by our attributions of causal connexion:

> We say, for instance, that the vibration of this string is the cause of this particular sound. But what do we mean by that affirmation? We either mean, *that this vibration is followed by this sound, and that all similar vibrations have been followed by similar sounds:* Or, *that this vibration is followed by this sound, and that upon the appearance of one, the mind anticipates the senses, and forms immediately an idea of the other.* We may consider the relation of cause and effect in either of these two lights; but beyond these, we have no idea of it.
>
> (*E* 7.29, original emphasis)

This seems to be a deliberate *extension* or *correction* of the previous passage.[26] When we first infer from observed *A* to anticipated *B* by custom, this naturally leads us to assert a connexion between them, and at that stage, we may only "mean" that they are connected *in our thought*. Having gone on to analyse the objective circumstances that generate this connexion, however, and correspondingly framing the first definition, we can then clearly mean something more: namely, the obtaining of those objective circumstances. The two passages thus constitute an illustration of the thinking behind Hume's two definitions, with no trace of the strident subjectivism of the *Treatise*. So by the time he came to write the *Enquiry*, at least, Hume evidently considered the obtaining of causal relations – and of the causal necessity that is essential to those relations – to be a thoroughly objective matter. Although an element of subjectivity remains in regard to the connection of objects in our thought and our tendency to draw corresponding inferences (with considerable significance for our own understanding of causation and of causal necessity), this is ultimately irrelevant to the objective truth or falsehood of causal claims.

32.6 Conclusion

Both causal and conceptual modality play a major role in Hume's philosophy, and the central material of his early *Treatise* develops directly from his attempt to accommodate causal necessity within an empiricist framework. The resulting account flirts notoriously with extreme subjectivism – though this is wildly at odds with his overall philosophical perspective – while also

hinting at a unified treatment of the two kinds of modality. By contrast, Hume's mature account in the first *Enquiry* clearly distinguishes between conceptual and causal modality, with the former falling within the realm of "relations of ideas" (answering to the Conceivability Principle), while the latter falls within the realm of "matters of fact" (and is unambiguously determined by objective regularities). Somewhat disappointingly, Hume does not now even gesture towards any unifying treatment of the two types of modality, but we have seen that plausible refinements of his view of *reflection*, along with some loosening of his atomistic theory of meaning and representation, could perhaps provide the resources for a more coherent account that embraces both types. This would undoubtedly go beyond Hume's own theory, but it would do so in ways that are at least foreshadowed by the development of his texts.[27]

Notes

1 For detailed discussion of the suggestions in these last two sentences, see Millican (2016, §II and §III respectively).

2 At *A* 6, Hume observes that his Copy Principle "seems to be equivalent to that which Mr. *Locke* has taken such pains to establish, *viz. that no ideas are innate*". At *T* 1.3.14.6, he remarks that "the principle of innate ideas … has been already refuted, and is now almost universally rejected in the learned world". Hume's correspondence with Thomas Reid in 1763 confirms that he considered the theory of ideas as copied from impressions to be generally accepted. Reid's *Abstract* of his *Inquiry into the Human Mind* takes this to be Hume's founding principle (1764/1997, pp. 257–8), and Hume writes on 25 February, "my Principles … were the common ones" (p. 263). In response, Reid writes on 18 March, "Your System appears to me … justly deduced from principles commonly received among Philosophers" (p. 264).

3 The *Enquiry* frequently refers to the crucial idea as that of "power or necessary connexion" (*E* 7.5, 6, 9, 26, 28, 30), and I shall often do the same, since this makes clear that it is supposed to embrace both causal power and necessity. At the beginning of the parallel *Treatise* discussion, Hume says "I begin with observing that the terms of *efficacy, agency, power, force, energy, necessity, connexion*, and *productive quality*, are all nearly synonimous" (*T* 1.3.14.4), and his discussion in both works tends to switch between the terms, often apparently for the sake of mere stylistic variety.

4 This contrast is particularly striking, given that the *Treatise* paragraph dealing with mental causation (*T* 1.3.14.12) was added only in the *Appendix* of 1740, suggesting that Hume's initial focus was almost entirely on *physical* causation.

5 Hume in the *Treatise* also presents a neat theory of how custom operates, building on his definition of belief as "a lively idea related to or associated with a present impression" (*T* 1.3.7.5). He sees inductive inference as a sort of hydraulic process in which the liveliness (or "force and vivacity") of the impression of *A* is conveyed through the associational channel to enliven the inductively associated idea of *B* (*T* 1.3.8.2, cf. 1.3.10.7); this enlivened idea then itself constitutes the expectation of *B*. In the *Enquiry*, by contrast, Hume suggests that belief is indefinable (*E* 5.12) and abandons the quasi-hydraulic model (e.g. compare the treatment of probability at *T* 1.3.11.12–13 with *E* 6.3).

6 Note that Hume's quest for the impression of necessary connexion accordingly *succeeds*, and the corresponding idea is thereby *vindicated*. It is a fundamental (but common) misunderstanding to see Hume as *debunking* causal necessity.

7 These two definitions are of *cause*, not *necessary connexion*: they aim to identify when a specific "object" can properly be identified as a cause of another. Hume later provides corresponding definitions of the relation of necessary connexion when he discusses "liberty and necessity" (*T* 2.3.2.4, *E* 8.27; cf. *T* 2.3.1.4, *E* 8.5), clarifying his repeated insistence that necessary connexion is essential to causation (*T* 1.3.2.11, 1.3.6.3, 2.3.1.18, 2.3.2.4; *E* 8.25, 8.27).

8 Though a contrast with Hume's other reflective impressions is inevitable, given that he usually restricts these to "passions, desires, and emotions" (*T* 1.1.2.1, cf. 1.1.1.1, 1.1.6.1, 1.2.3.3), none of which can plausibly serve as an impression of power or necessary connexion.

9 In the *Enquiry* Hume consistently prefers "transition" to "determination", perhaps because of the latter's causal overtones, which can seem viciously circular when he is trying to account for the *origin* of our causal concepts. He makes a parallel observation at *E* 8.25 n. 19, preferring a term expressing *temporal succession* to any hint of *causation*.

10 Recognition of such a reflective "impression" would, however, put serious pressure on the conceptual atomism that Hume seems to presuppose in the *Treatise*, for example in the strict simple/complex distinction introduced at *T* 1.1.1.2, the Separability Principle stated at *T* 1.1.7.3, and the view of internal impressions as "original facts and realities, compleat in themselves" (*T* 3.1.1.9, cf. 2.3.3.5). It is therefore notable that all of these disappear after the *Treatise* (though a vaguer – and perhaps less absolute – simple/complex distinction remains implicit at *E* 2.6, 2.8, 3.1, 7.4, and 7.8 n.12). These changes fit well with, and may perhaps be related to, the apparent development in Hume's view of "reflection".

11 For detailed discussion of Hume's logic in both works (including his appeals to conceivability), see Millican (2017).

12 *T* 1.3.3.3 concerns whether the Causal Maxim is "intuitively [or] demonstratively certain", appealing to conceivability, so the modality in question is unambiguously *conceptual*, even though the maxim itself concerns causation. *E* 8.4 introduces Hume's application of his analysis of *causal* necessity to the question of "liberty and necessity".

13 As also at *T* 1.1.1.10 and *E* 2.8, on the possibility of the missing shade of blue, and *T* 2.1.5.3, on the self-focus of the passions of pride and humility (though Hume's view on the latter changed in the 1757 *Dissertation on the Passions*, where *P* 2.1 apparently treats self-focus as *definitive* of pride and humility, in contrast to *T* 2.1.2.1).

14 Some scholars, however, have interpreted Hume as believing that genuine causation involves underlying *conceptual* necessities which, though epistemologically inaccessible to us, would in principle enable the effect to be inferred a priori from the cause. Galen Strawson (1989, p. 111) calls this the "a-priori-inference-licensing property", while Peter Kail (2007, p. 256) calls it the "reference-fixer for 'power'". This *sceptical realist* (or "New Hume") reading, which originated with John Wright (1983), is in serious tension with Hume's frequent appeals to the Conceivability Principle, because if such necessities do exist between cause *A* and effect *B*, then it will in fact be *impossible* for *A* to occur without *B* following, even though – as Hume repeatedly insists – we are perfectly able to *conceive* of such an outcome. Sceptical realist interpreters have accordingly played down his commitment to the Conceivability Principle, but this seems implausible given how often he appeals to it. For detailed discussion of this objection, see Millican (2009, §6).

15 The words "objective" and "subjective" are notoriously slippery, so to be clear, a causal relation is here understood to be *objective* if it is ascribable truly or falsely, irrespective of the subjective observer's point of view.

16 *T* 1.3.14.23 is similar, encompassing also the "efficacy or energy of causes".

17 Again, see *T* 1.3.14.27 for a similar comment about "efficacy".

18 The case of geometry is not so clear, since the *Treatise* treats it as less than certain, founded on imprecise ideas that are copied from visual impressions (*T* 1.2.4.17–29; 1.3.1.4). In the *Enquiry*, however, geometry is explicitly embraced within the realm of "relations of ideas", alongside algebra and arithmetic (*E* 4.1).

19 Thomas Holden, however, has recently argued that *T* 1.3.14.23 manifests "an expressivist account of absolute necessity" (2014, p. 379) which *equates* necessity with (idealised) human inconceivability of the contrary. Against this, Hume appears to accept that some "matters of fact", though possible, are inconceivable owing to human conceptual limitations, and he never asserts that possibility implies conceivability (i.e. the converse of the Conceivability Principle). Moreover, even Holden's account acknowledges an objective correlate to inconceivability of the contrary, namely objective relations between the "semantic content" of ideas (p. 398), and Hume's clear recognition of "relations of ideas" in the *Enquiry* suggests that he sees his metaphysics of conceptual necessity as founded directly on those relations, rather than on limited human conceivability. For more detailed discussion, see Millican (2017, §5).

20 These rules do not occur in the *Enquiry*, perhaps because Hume by then had a more sophisticated view of causation in terms of the continuous influence of quantitative *forces*, rather than specific patterns of discrete events. This more sophisticated view is apparent at *E* 4.13, 4.16, 7.25 n. 16 and 7.29 n. 17 (for discussion, see Millican 2007, pp. 232–3).

21 For a detailed case justifying the standard assumption that Hume is a determinist, see Millican (2010).

22 Another important application is Hume's defence of materialism (*T* 1.4.5.29–33), including a passage cited in the previous paragraph. These uses of his definitions of cause seem to require that he takes them to be genuinely definitive of what causal necessity is, thus providing a major objection against interpretations such as the "New Hume" (cf. note 14) which deny this. For detailed consideration of this objection, see Millican (2009, §§7–8; 2011).

23 For more on this approach to the two definitions, see Millican (2009, §4). Don Garrett (1997, pp. 97–101) gives a useful overview of the many alternative interpretations that have been proposed, and some of the issues that face them.

24 Indeed, the supposed subjectivity of causal necessity has generally been considered the main objection to interpreting Hume as a regularity theorist about causation: see, for example, Helen Beebee (2012, pp. 137–8), Francis Dauer (2008, p. 95), and Garrett (1997, pp. 99–100). One popular response, the *projectivist* interpretation, aims to reconcile the objectivist and subjectivist strands in Hume's texts by building on parallels with his moral theory (Beebee 2006, p. 150; Simon Blackburn 2008, pp. 28–9; Angela Coventry 2006, pp. 133–7), and in particular his famous statement that moral taste operates by "gilding or staining … natural objects with the colours, borrowed from internal sentiment" (*MAppx* 1.21). Such interpretations make a plausible case regarding the attitudinal aspect of Hume's "impression of necessary connexion". But they cannot, I believe, seriously challenge the objectivist nature of his own causal theory, given that in the paragraph containing that famous statement (as elsewhere), he clearly puts causal relations in the province of ungilded *reason* rather than *taste*.

25 See for example *T* 2.3.1.2, 2.3.8.13, 3.1.1.26, *App*. 2; *E* 7.25 n. 16.

26 As is strongly suggested by the otherwise coincidental wording. The phrase "we mean" occurs in only 12 paragraphs of Hume's philosophical works, and this is the only case of its occurring in adjacent paragraphs.

27 I am grateful to Amyas Merivale for comments on an earlier version of this chapter.

References

Beebee, Helen (2006), *Hume on Causation*, London: Routledge.

Beebee, Helen (2012), "Causation and Necessary Connection", in Alan Bailey and Dan O'Brien (eds.), *The Continuum Companion to Hume* (London: Continuum), pp. 131–145.

Blackburn, Simon (2008), *How to Read Hume*, London: Granta Books.

Coventry, Angela (2006), *Hume's Theory of Causation: A Quasi-Realist Interpretation*, London: Continuum.

Dauer, Francis Watanabe (2008), "Hume on the Relation of Cause and Effect", in Elizabeth S. Radcliffe (ed.), *A Companion to Hume* (Oxford: Blackwell), pp. 89–105.

Garrett, Don (1997), *Cognition and Commitment in Hume's Philosophy*, Oxford: Oxford University Press.

Holden, Thomas (2014), "Hume's Absolute Necessity", *Mind* 123, pp. 377–413.

Hume, David (all texts), the website www.davidhume.org contains searchable electronic versions of all Hume's philosophical publications and has been used extensively in the research for this chapter. The codes for individual works as cited here ("*T*", "*A*", "*E*", "*M*", "*P*") follow the practice of this website.

Hume, David (1739–40), *A Treatise of Human Nature: A Critical Edition*, vol. 1, ed. David Fate Norton and Mary J. Norton, Oxford: Clarendon Press, 2007 (references indicated by "*T*").

Hume, David (1740), *An Abstract of … A Treatise of Human Nature*, included in both Hume (1739–40) and Hume (1748); (references indicated by "*A*" and given to paragraph number).

Hume, David (1748), *An Enquiry Concerning Human Understanding*, ed. Peter Millican, Oxford: Oxford University Press, 2007 (references indicated by "*E*").

Hume, David (1751), *An Enquiry Concerning the Principles of Morals*, ed. Tom L. Beauchamp, Oxford: Oxford University Press, 1998 (references indicated by "*M*").

Hume, David (1757–72), *A Dissertation on the Passions and The Natural History of Religion: A Critical Edition*, ed. Tom L. Beauchamp, Oxford: Clarendon Press, 2007 (references to *Dissertation* indicated by "*P*" and given to section and paragraph number).

Kail, P. J. E. (2007), "How to Understand Hume's Realism", in *Read and Richman* (2007), pp. 253–269.

Locke, John (1690), *An Essay Concerning Human Understanding*, ed. P. H. Nidditch, Oxford: Clarendon Press, 1975.

Millican, Peter (2007), "Against the 'New Hume'", in *Read and Richman* (2007), pp. 211–252.

Millican, Peter (2009), "Hume, Causal Realism, and Causal Science", *Mind* 118, pp. 647–712.

Millican, Peter (2010), "Hume's Determinism", *Canadian Journal of Philosophy* 40, pp. 611–642.

Millican, Peter (2011), "Hume, Causal Realism, and Free Will", in Keith Allen and Tom Stoneham (eds.), *Causation and Modern Philosophy* (New York: Routledge), pp. 123–165.

Millican, Peter (2016), "Hume's Chief Argument", in Paul Russell (ed.), *The Oxford Handbook of Hume* (Oxford: Oxford University Press), pp. 82–108.

Millican, Peter (2017), "Hume's Fork, and His Theory of Relations", *Philosophy and Phenomenological Research* 95, pp. 3–65.
Read, Rupert and Richman, Kenneth A. (eds.) (2007) *The New Hume Debate: Revised Edition*, London: Routledge (1st edition, 2000).
Reid, Thomas (1764/1997), *An Inquiry into the Human Mind on the Principles of Common Sense*, edited by Derek R. Brookes, Edinburgh: Edinburgh University Press.
Strawson, Galen (1989), *The Secret Connexion: Causation, Realism, and David Hume*, Oxford: Clarendon Press.
Stroud, Barry (1977), *Hume*, London: Routledge & Kegan Paul.
Wright, John P. (1983), *The Sceptical Realism of David Hume*, Manchester: Manchester University Press.

Chapter 33
KANT ON REAL POSSIBILITY

Nicholas Stang

33.1 Introduction

Kant's Critical philosophical system is replete with modal notions. Limiting ourselves to his theoretical philosophy, consider several of the core doctrines of the *Critique of Pure Reason*:

(i) Space and time are conditions of the possibility of objects as appearances; they are not conditions of the possibility of those objects as things in themselves.[1]
(ii) The pure concepts of the understanding, the categories (e.g. <*cause-effect*>), as well as the principles of understanding (e.g. that all alterations have a cause), are conditions of the possibility of experience of objects.[2]
(iii) The conditions of the possibility of experience of objects are also conditions of the possibility of the objects of experience. We cannot know whether they are also conditions of the possibility of things in themselves.[3]

It is thus no exaggeration to say that Kant's theoretical philosophy simply could not be articulated without modal concepts and, more specifically, the concept of possibility. Despite this, modality has until recently received relatively little attention from scholars. In the past decade or so, there has been a burst of interest in Kant's views on modality.[4] In this chapter, I present an opinionated introduction to the role of modality in Kant's theoretical philosophy—in contemporary terms, his metaphysics, epistemology, and theory of mind—focusing on his views about possibility.[5] Although Kant developed sophisticated views on modality in his writings of the 1750s and 1760s—most notably in his 1763 book *The Only Possible Ground of Proof for a Demonstration of the Existence of God*—I will focus mainly on his views on possibility, actuality, and necessity in the *Critique of Pure Reason* (1st ed. 1781; 2nd ed. 1787) and other texts from his so-called 'Critical system,' including the transcripts of his lectures on metaphysics and his unpublished handwritten notes, known as *Reflexionen*.

In Section 33.1, I explore the most basic distinction in Kant's modal theory, that between 'logical' and 'real' possibility. In Section 33.2, I introduce key elements for understanding Kant's Critical theory of experience and the role of modality within it: the distinction between intuition and concept, and the idea that intuitions and concepts have a form and a matter. In Section

33.3, I explain Kant's official definition of possibility in the *CPR* and why it is a definition of real, rather than logical, possibility. In Section 33.4, I argue that this notion of possibility, which I call *formal* real possibility, is not the only kind of real possibility Kant accepts in the Critical system.

33.2 Logical vs. real possibility

Kant was educated in a philosophical tradition now referred to as 'German rationalism,' and throughout his philosophical career, up to and including the *Critique of Pure Reason*, he was engaged in critical debates with other representatives of this tradition.[6] With respect to modality, Kant's pre-Critical writings are marked by a confrontation with a view about possibility shared by G.W. Leibniz (1646–1716), Christian Wolff (1679–1754), and A.G. Baumgarten (1714–1762), which I call 'logicism,' namely, that the possible is that which is internally logically consistent.

Here are some representative statements of the logicist view of possibility:

> *Leibniz*: 'Provided also that it is understood that necessity and possibility, taken metaphysically and strictly, depend solely upon this question, whether the object in itself or that which is opposed to it implies contradiction or not.'[7]
>
> *Wolff*: 'One can further see that whatever contains [*enthält*] nothing contradictory in itself is possible.'[8]
>
> *Baumgarten*: 'Every thing that is representable, that does not contain a contradiction, which is not A and not-A, is possible.'[9]

As the quote from Leibniz indicates, in addition to what he here calls 'strict' or 'metaphysical' possibility, Leibniz also accepts several other kinds of possibility, such as 'moral' and 'physical' possibility. Wolff and Baumgarten also distinguish the notion of possibility defined here (which they refer to as *per se* or *an sich* possibility) from others (e.g. what is possible *in nexu*), but in this chapter I will focus on their theories of what Leibniz would call strict metaphysical possibility. In contemporary terms, the logicist view is that something is possible just in case its concept is logically consistent, that is, it does not contain both a given mark (subconcept) and its negation. I will refer to this as the *Coextensiveness thesis*.[10] To use a mathematical example, a right-angled triangle is possible (because <*right-angled*> does not contradict any of the marks contained in <*triangle*>), but a triangle that does not have three sides is impossible (because <~*three-sided*> contradicts <*three-sided*>, which is one of the marks of <*triangle*>).[11]

For modal issues, Kant's most important pre-Critical work is his 1763 book *The Only Possible Ground of Proof for a Demonstration of the Existence of God* (*OPG*), in which he distinguishes between the 'logical,' or 'formal' element in possibility and the 'real,' or 'material' element:

> Likewise, in every possibility that which is thought must be distinguished from the agreement of what is thought with the principle of contradiction. A triangle with a right angle is in itself possible. The triangle, as well as the right angle, are the data or the material [element] of this possibility, while the agreement of the one with the other according to the principle of contradiction is the formal [element] of possibility. I will also call the latter the logical [element] of possibility, for the comparison of the predicates with their subject according to the rule of truth is nothing other than a logical relation; the something, or what stands in this agreement, will sometimes be called the real [element] of possibility.
>
> (*OPG*, 2: 77–8)[12]

The 'logical' element in possibility is a requirement of logical consistency: the marks that constitute the concept of a thing must be logically consistent. But being internally logically consistent is not sufficient for possibility: to be possible, a thing must also satisfy the 'real' requirement on possibility. Its matter, the marks that stand in a logically consistent relation, must themselves be possible:

> Since you must eventually arrive at something whose possibility cannot be further analyzed, suppose that it is the concept of extension that you cannot further analyze into simpler data in order to show that there is nothing contradictory in it. The question then arises, whether 'space' or 'extension' are empty words or whether they signify something. The lack of a contradiction does not settle it; an empty word never designates something contradictory. […] As long as you prove possibility through the principle of contradiction, you are depending on the thinkable that is given in things and you are considering only connections according to this logical rule. But when you finally consider how this [thinkable] is given to you, you can base it on nothing other than an existence [*ein Dasein*].
>
> (*OPG*, 2:80–1)

Even of a logically simple concept (e.g. the ultimate products of an analysis of <*extension*>), which has no non-trivial marks, we can ask whether it is possibly instantiated. The 'real' requirement on possibility is thus not a formal requirement on the internal consistency of the marks of a concept. Kant here identifies it with the requirement that the concept 'signify' something thinkable, which I will take to mean that it refers to a possible property. (In *OPG* Kant takes the possible and the thinkable to be coextensive.)[13] The real requirement is thus a metaphysical requirement on possibility: there must be possible properties in the world referred to by that concept.[14] Kant argues that possible properties (properties that satisfy the real requirement on possibility) must be possible in virtue of some existing thing; that is, their possibility must have an existing ground. I will call this the *Grounding thesis*. A significant part of *OPG* consists in an extended argument that there is a unique such ground of all possibility, and that this unique ground of all possibility is God.[15]

This pre-Critical view of possibility already involves a break with the 'logicist' view of possibility, albeit a subtle one. The logicists accept the Coextensiveness thesis but they deny that logical consistency alone *grounds* possibility. In fact, they accept the Grounding thesis: the possibility of a thing requires an extralogical ground. In particular, they agree with the pre-Critical Kant that God is the ground of the possibility of the simple predicates that compose logically consistent complex concepts.[16] Thus, Kant's assertion that possibility has a real requirement may appear to be anti-logicist, but it is not.

Instead, the anti-logicist element of Kant's view is his denial of the Coextensiveness thesis. He departs from the logicists in thinking that a thing might satisfy the logical requirement on possibility without satisfying the real requirement.[17] His example in *OPG* is thinking and extension: 'The impenetrability of bodies, extension and such like, cannot be attributes of that which has understanding and will' (*OPG*, 2:85). This is not because <*understanding*> contains some hidden mark that contradicts a mark contained in <*extension*> (or vice versa); these concepts are logically compatible, but not possibly co-instantiated. Logical consistency and grounding (by God) are not merely different requirement on possibility; they can come apart. Some things are logically consistent (e.g. an extended thing that thinks) but are impossible because they lack an appropriate ground.

Kant continues to reject the Coextensiveness thesis throughout the Critical period, during which he introduces a helpful piece of terminology. That which is internally self-consistent is

said to be *logically possible*. That which additionally meets the real requirement on possibility (i.e. its possibility has an extralogical ground) is said to be *really possible*.[18] Kant's denial of the Coextensiveness thesis can now be concisely stated: not all logical possibilities are real possibilities.

33.3 Concept and intuition

Before considering Kant's Critical view of real possibility, we must first pause to investigate the distinction between concepts and intuitions, which structures Kant's whole Critical theoretical philosophy, including his modal theory. Concepts and intuitions are representations, that is, mental states that represent objects, but they represent their objects in fundamentally different ways. Concepts are general and mediate representations.[19] They are general because they represent a general class of objects, namely, the objects in their domain. For instance, the concept <*square*> does not represent any individual square, but all square objects as such. Concepts are mediate representations because they represent objects as falling within their extension *by means of* (mediated by) representing them as falling under their constituent concepts, or marks. For instance, if <*square*> is composed of the concepts <*equilateral*>, <*equiangular*>, and <*four-sided plane figure*>, then one thinks of an object that it is square by means of (mediated by) thinking that it is an equilateral, equiangular, and four-sided plane figure. It represents the extension it does because that extension is the intersection of the extensions of these three concepts.

Intuitions, by contrast, are singular and immediate representations.[20] There is an ongoing scholarly controversy about the nature of Kantian intuitions and their relation to concepts, but for this chapter I will abstract from that controversy and simply state my own interpretation.[21] The singularity of intuition means that an intuition is of one and only one object. Of any intuition we can ask, 'What is the object of that intuition?' The answer cannot be a generic class of objects, nor can the answer be 'none'; it must be a *single* object. The immediacy of intuitions means that intuitions do not represent their object by means of (mediated by) any other representation of that object. Intuitions do not, for instance, represent their objects through mediating images or descriptions of their objects. Intuition is an immediately mental relation to a single object; the object is said to be 'given' to the subject. It is in virtue of this relation that the object becomes available to be *cognized* by the subject, assuming the subject has appropriate concepts to apply to it. A cognition is an act of subsuming a single object under a concept; its basic logical form is F*a*. To think that F*a* is to think of *a* that it falls under F, and this requires two fundamentally different capacities: (i) the capacity to think of objects that they fall under F; and (ii) the capacity for *a* itself to figure in the contents of one's thought, as opposed to another object *b*, or a description of a class of objects (e.g. thinking F(λxGx)). I identify (i) with a Kantian concept (hence its generality), and (ii) with a Kantian intuition of *a* (hence its singularity).

Either capacity can be present without the other, but in isolation they do not produce cognition. As Kant famously writes: 'Thoughts without content [*Inhalt*] are empty; intuitions without concepts are blind' (A51/B75). A concept has content in the relevant sense if it is a concept of objects that can be given to us in intuition.[22] I possess concepts that lack content (e.g. <*things in themselves*>) because I cannot intuit any instances. Consequently, I can think general thoughts involving these concepts (e.g. that things in themselves are non-spatiotemporal), but these thoughts are 'empty': because I cannot think *de re* of any object that it instantiates these concepts, I cannot use this thought to cognize any object.[23]

Kant calls our capacity (*Vermögen*) to be given objects (i.e. to intuit them) *sensibility*, because our capacity for intuiting objects is *sensible*, that is, receptive: we intuit the particular objects we

do because they causally affect our sensory organs.[24] By contrast, *understanding*, that is, the capacity to think sensibly given objects under concepts, is *spontaneous*.[25] We are active in subsuming an object under a concept because nothing in the sensorially received data forces us to subsume that object under that concept (though it makes certain cognitions correct and others false). Both capacities have a *form*, that is, an inner structuring principle that is invariant across particular objects of cognition and is independent of the particular course of our experience, as well as a *matter*, which varies from object to object and depends upon the sensory data we receive.[26] The experience-independent (*a priori*) form of our intuition, for instance, is space and time. Regardless of the course of our experience, if we intuit anything at all, we intuit objects outside of us (as opposed to our own mental states) in space, and we intuit object inside us (our own mental states) in time. Space and time are the form of our intuition, while the matter is the particular objects given to us in space and time.

Likewise, the form of our understanding is a set of concepts determined by the very nature of our minds, which Kant calls the 'pure concepts of the understanding,' or *categories* for short. Regardless of which particular objects affect us and what particular properties they have, in thinking about them (i.e. cognizing them by subsuming them under concepts) we must think about them using the categories. In particular, we must think of them as having *quantity* (i.e. being unities, pluralities, and totalities) as having *quality* (i.e. having properties that are real, as opposed to the mere negation or absence of such properties, where these real properties can come in degrees), and standing in *relation* (i.e. the relation of substance and accident, of cause and effect, and the relation of mutual causal interaction or 'reciprocal action').[27] Just as the matter of intuition is the particular objects we intuit, the matter of understanding is the particular empirical concepts we find instantiated by these objects. We can specify *a priori* the intuitional and conceptual form of objects of cognition; for anything more determinate than that, we must consult experience.

33.4 Formal real possibility

Kant's Critical conception of logical possibility remains essentially what it had been in his pre-Critical works and in his logicist predecessors: the lack of internal contradiction within a concept.[28] But the distinction between concepts and intuitions, and the discovery of the *a priori* form of sensibility and cognition, transform Kant's conception of real possibility. In the 'Postulates of Empirical Thinking in General,' the section of the *CPR* on the modal categories, Kant gives the following principle for the category of <*possibility*>: 'Whatever agrees with the formal conditions of experience (in accordance with intuition and concepts) is possible' (A217/B264).

I want to note seven things about this principle:

1. It is a *definition*. As Kant writes: 'The principles of modality are also nothing further than definitions of the concepts of possibility, actuality, and necessity in their empirical use' (A219/B266). This means that it is not merely an epistemic criterion of how we know whether something is possible; it tells us what possibility is, at least within the realm of experience (i.e. objects that can be cognized by us).
2. It is a definition of *real* possibility, for not all logically consistent concepts satisfy it. There are logically consistent concepts for which an appropriate object can never be given to us in intuition. To use one of Kant's favorite examples, <*triangle whose internal angles sum to fewer than 180 degrees*> is logically consistent, but given the spatial form of our intuition, no such object can be given to us. Consequently, it fails to meet the quoted definition of possibility. This means that it is a definition specifically of real possibility.[29] This also confirms Kant's

consistent rejection of the co-extensiveness thesis (that all logical possibilities are real possibilities) affirmed by the logicists.[30]

3. The concept <*possibility*> does not apply directly to objects. Kant writes:

> The categories of modality have this peculiarity: they do not in the least augment the concept to which they are ascribed as a determination of the object, but rather express only the relation to the faculty of cognition.
>
> (A219/B266)

Modal categories are not first-order concepts of objects that could 'augment' the concept of an object (i.e. enrich its content by further specifying the range of objects to which it applies). For instance, the concept <*equilateral*> augments the concept <*four-sided plane-figure*> to form the more determinate concept <*rectangle*>; because the new concept has a richer content (intension) than the former concept, it has a narrower scope (extension). <*Possibility*> is not like this. The role of <*possibility*> is not to augment a concept of objects and thus further restrict its extension; it is to express the relation of that concept to our faculties of cognition, that is, to express whether an object instantiating that concept can be given to us (agrees with our form of intuition) and be subsumed under that concept by us (agrees with our form of understanding).[31] In the very next sentence, Kant writes: 'If the concept of a thing is already entirely complete, I can still ask about this object whether it is merely possible, or also actual, or, if it is the latter, whether it is also necessary' (A219/B266). <*Possibility*> thus fundamentally applies to concepts of objects, not to objects themselves. It applies when that concept is the concept of an object that is cognizable by us.[32]

4. Real possibility is here defined in terms of form: what 'agrees' with the sensible form of our intuitions and the intellectual form of our concepts is really possible. It is unclear precisely what 'agreement' means here, but we can say this: the form of our intuition (space and time) and the form of understanding (the categories) determine certain things to be the case, and the really possible is what is consistent or compatible with them.[33] For instance, the form of space determines a geometrical structure; whatever is really possible must agree with that geometrical structure. Likewise, the form of the understanding determines that objects fall under the categories (e.g. <*cause-effect*> and <*substance-accident*>); whatever is really possible must agree with that conceptual structure (i.e. it must consist of substances that cause one another's accidents to change). And the really possible is whatever agrees with both forms: the geometrical structure of space and time and the conceptual structure of the categories. The matter of experience (which objects are in space and time, which empirical concepts they instantiate) agrees with the form but is not required by it; other matter would also agree with the form.[34] This means that the matter of experience is really *contingent* (i.e. its opposite is really possible).[35] Because real possibility, as Kant defines it in the Postulates, is concerned with the agreement of an object with the conceptual and intuitional form of our experience, I refer to it as *formal* real possibility.
5. Formal real possibility has two kinds of ground: the forms of intuition (space and time) and the forms of understanding. This means that possible objects in space and time are possible because of—their possibility is grounded in—space and time themselves, the forms of intuition.[36] With respect to the other ground of the formal real possibility of objects, Kant's considered view is that this ground is what he calls the 'unity of apperception,' that is, the unity of self-conscious subjectivity, which is itself the ground of the categories.[37] This means that Kant continues to maintain his pre-Critical endorsement of the Grounding

thesis, but with an important modification: the ground of (formal) real possibility is no longer an existing substance, but the forms of our cognition (space, time, the unity of apperception). Kant maintains at least the structure of his pre-Critical theory of real possibility such that real possibilities must have a ground, but generalizes his conception of what that ground must be to include cognitive forms.[38]

6. While this may appear to be a definition of 'epistemic' possibility (i.e. real possibility is what is cognizable by us), it is also a definition of an *alethic* modality, because Kant is a transcendental idealist about objects in space and time (objects of our cognition). What exactly his transcendental idealism amounts to is a matter of significant controversy.[39] It is relatively uncontroversial that the realm of what is possible for objects in space and time (metaphysical possibility) is determined by what it is possible for us to cognize (epistemic possibility). As Kant writes: 'The conditions of the possibility of experience in general are conditions of the possibility of the objects of experience' (A158/B197). I take this to mean that what is (alethically) possible for objects in space and time is what is epistemically possible for us to cognize, and the latter explains or 'grounds' the former: if something is possible for objects in space and time, this is because (in a non-causal sense) it is cognizable by us.[40]
7. This is a definition of the category <*possibility*> in its application specifically to empirical objects, that is, objects that can be given in intuition: 'The principles of modality are also nothing further than definitions of the concepts of possibility, actuality, and necessity *in their empirical use*' (A219/B267; my emphasis). Whether another definition of <*possibility*> might be required for thinking about the possibility of objects that we cannot intuit (what Kant calls *noumena*) is not specified here. That is, this definition leaves unspecified whether there is another concept of real possibility that could be used to think about the possibility of such objects.

33.5 Beyond real formal possibility

An exclusive focus on the 'Postulates of Empirical Thinking in General,' the section of the *CPR* about modal categories and principles, might suggest that formal real possibility is the only kind of real possibility Kant accepts—in other words, that he maintains that formal real possibility exhausts real possibility *überhaupt*. In this section I will argue that this is not the case.[41]

In later sections of the *CPR*, and at greater length in the *Critique of Practical Reason* (*CPrR*), Kant introduces a new notion of real possibility when he tries to reconcile the applicability of the moral law to our actions with the thoroughgoing determinism that governs all events in space and time.[42] Kant accepts an 'ought implies can' principle: if I ought not to have φ'ed, then it was possible for me not to φ.[43] But this presents a problem, for Kant takes himself to have proved in the Second Analogy of Experience that all events, and hence all of our actions in space and time, are necessitated by antecedent causes (i.e. that there is causal determinism).[44] To resolve this problem, Kant needs to show that there is a sense in which the omission of some action I actually commit (e.g. make a lying promise) is *possible* (otherwise, by the 'ought implies can' principle, it would not be the case that I ought to have omitted it and the moral law would lose its hold on us). But Kant regards neither the *logical* possibility of my omission of the act nor, crucially, the *formal* real possibility of my omission as sufficient for satisfying the 'ought implies can' principle and thus saving the applicability of the moral law to my action.[45] The fact that the omission of my lying promise agrees with the formal conditions of experience just as much as my actual commission does nothing to vindicate the possibility of my having done otherwise. This means that, although the omission of an act I actually commit may be formally really possible, its formal real possibility is not the kind of possibility relevant to satisfying the 'ought implies can' principle.

Instead, to satisfy the 'ought implies can' principle, Kant appeals to his transcendental idealism and invokes the distinction between appearances and things in themselves:

> Every action [of a subject] […] is to be regarded in the consciousness of his intelligible existence as nothing but the consequence and never as the determining ground of his causality as a noumenon. So considered, a rational being can now rightly say of every unlawful action he performed that he could have omitted it even though as appearance it is sufficiently determined in the past, and so far, is inevitably necessary; for this action, with all the past which determines it, belongs to a single phenomenon of his character, which he gives to himself and in accordance with which he imputes to himself, as a cause independent of all sensibility, the causality of those appearances.
>
> (*CPrR*, 5: 97–8)

Considered as an object in space and time, my actions are governed by deterministic causal laws. However, I can also be considered as an 'intelligible' existence, that is, a thing in itself outside of space and time, and as such I have a causality, that is, a power to act. In particular, I have the power to act so that my action in space and time, such as the lying promise, would have been omitted. Thus, the possibility of my omitting the lying promise is explained by my intelligible causality, that is to say, the causal powers I possess considered 'in myself,' not as I appear to others and myself in space and time.[46]

In conclusion, I want to make three observations about this notion of possibility:

1. It is a notion of *real* possibility. The mere logical consistency of my omitting the act does not entail that, as an 'intelligible' cause, I have the power to act so that it is omitted. Since it is defined in terms of the powers I possess as an intelligible being rather than as a sensible being—in Kantian terms, as a noumenon rather than as a phenomenon—it is not formal real possibility. I will refer to it as *noumenal* real possibility.
2. Noumenal real possibility has a ground: my 'intelligible' causality, that is, my causal powers as a noumenon. It is not a brute fact that it is really possible that I omit my lying promise. It is really possible that I omit my lying promise in virtue of the fact that I possess the power to act so that it would be omitted if I were to so act. Noumenal real possibility, in other words, satisfies the Grounding thesis from Section 33.1. Unlike formal real possibility, it is not grounded in subjective cognitive forms, but in the causal powers of noumena.
3. This suggests that Kant maintains the basic structure of his pre-Critical theory of real possibility (rejecting the Coextensiveness thesis while accepting the Grounding thesis) within the Critical system, but generalizes it by abandoning the assumption that the ground of real possibility must be God. However, this also raises the intriguing question of what role, if any, Kant sees for God in grounding real possibility in the Critical system. However, answering that question will have to await another occasion.[47]

Notes

1 A27/B43, A46/B64, A46/B64, A89/B122, B148, 20:390. The *Critique of Pure Reason* is cited in the customary fashion: the page number in the 1781 edition (A) followed by the page number in the 1787 edition (B). Kant's other works are cited by volume and page number in the *Akademie* edition, Kant 1902–. 'Refl.' refers to *Reflexionen*, which can be found in volumes 14–19 of the *Akademie* edition. The *CPR* is quoted from the Guyer and Wood translation (Kant 1998), with my modifications.

2 A89/B122, A94/B127, B161, B188, 20:210, 20:225, 20:274, 28:187, 28:1520, 28:1525, 29:804, Refl. 5165.

3 A111, A139/B178, A158/B197, 4:318, 5:167, 10:428, 29:798, Refl. 4469, Refl. 5184, Refl. 6317.
4 Abaci (2008, 2013, 2014); Chignell (2009, 2010, 2011, 2012, 2014a, 2014b); Kannisto (2013, 2016, 2017); Leech (2012, 2014, 2017); Rosenkoetter (2013); Stang (2010, 2011, 2015, 2016a, 2016b).
5 Modal notions are also central to Kant's moral philosophy, as well as his aesthetics and teleology, though I do not have the space to discuss that point here. Cf. the discussion of 'conditions' and 'grounds' of possibility in the *Critique of Practical Reason* (5:34, 134, 143) and the *Critique of Judgment* (5:369–70, 388, 406).
6 For background on German rationalism, see Wundt (1939, 1945) and Beck (1969); more recent commentators who have emphasized Kant's dialogue with the German rationalists include Watkins (2005) and Hogan (2009).
7 Leibniz (1985, §367).
8 Wolff (1965, §12).
9 Baumgarten (2013, §8).
10 While, strictly speaking, the quoted passages from Wolff and Baumgarten state only a sufficient condition on possibility (logical consistency), I think that, in context, it is clear that they are intended to be definitions of possibility, so I attribute the biconditional to them.
11 For more on the logicist conception of possibility see Stang (2016a, ch. 1.2).
12 Translation from Kant (1992).
13 See *OPG*, 2:77, 78, 80, 82, 83, 85, 87, and 162.
14 Kant does not explicitly distinguish concepts (representations) and properties (worldly features) in *OPG*, but speaks of 'predicates.' I have introduced this distinction to mark two roles that 'predicate' is playing: as logically complex representational items (concepts) and worldly features that can be instantiated (properties). See Stang (2016a, ch. 4.5) for more discussion of this issue.
15 In recent years a number of different reconstructions of this argument have been offered by scholars. See Abaci (2014), Boehm (2012), Chignell (2009), Stang (2010, 2016a), and Yong (2014).
16 This is clearest in the case of Leibniz: 'It is also true that God is not only the source of existences, but also that of essences insofar as they are real or of that which is real in possibility. This is because God's understanding is the realm of eternal truths or that of the ideas on which they depend; without him there would be nothing real in possibles, and not only would nothing exist, but also nothing would be possible (Leibniz 1989, 218; cf. Leibniz 1985, §184). Wolff (1965, §974) and Baumgarten (2013, §864) also express commitment to such views. This is a point I made, but did not emphasize, in Stang (2016a, 144n40).
17 This claim is controversial. Yong (2014) and Abaci (2014) point out, contra Chignell (2009), that the passages I quote do not by themselves show that there can be real incompatibilities among logically consistent predicates (cf. Chignell's response in his 2014a). While I agree with Yong and Abaci in their reading of the passage in which Kant introduces the material requirement on possibility, I agree with Chignell that already in *OPG* Kant accepts real metaphysical incompatibilities (not to be confused with Chignell's 'real conflicts') between logically consistent predicates. See Stang (2016a, ch. 3.5) for my complete argument for this claim.
18 The first instance of this terminology of which I am aware is Refl. 4196 (17: 452), which Erich Adickes dates to 1769–70.
19 A320/B377, 9:91, 10:266, Refl. 1705.
20 On the singularity of intuition see A320/B377, A713/B741, A734/B762, 28:545–6, 29:566, 29:568, 29:616, 29:805, 806, 29: 904, 29:977, 29:981; on the immediacy of intuition see B41, A32/B48, A68/B93, A109, A156/B195, A226/B273, 4:281.
21 The interpretation that is closest to mine is Allais (2015), though I would differentiate intuition more sharply from the contemporary notion of a 'perception' or 'perceptual experience' than Allais does.
22 Kant defines the content of a concept as its 'relation to an object' (A55/B79, A58/B83, A62–3/B87), which I take to mean a relation to an object that can be given to us in intuition. Concepts of objects that cannot be given in intuition are not contentless in the contemporary sense; that is, they are not meaningless or senseless. See Tolley (2011) for critical discussion.
23 Kant claims that a concept for which no object can be given in intuition 'has no sense and is entirely empty of content' (A156/B195), but makes clear shortly thereafter what this means: that they cannot be used to cognize objects because no object that instantiates them can be presented to us in intuition (A245–6).
24 B33.
25 A51/B75, A68/B93, A97, A126, B150, B430.

26 For the form and matter of sensibility see A20/B34, A42/B59–60, A50/B74; for the form and matter of understanding see A54–6/B78–80.
27 Modal categories (possibility-impossibility, existence-nonexistence, and necessity-contingency) 'do not augment the concept to which they are ascribed in the least, but rather express only the relation to the faculty of cognition' (A219/B266). I take this to mean that they do not characterize properties of objects of cognition, but the relation of those objects to the faculty of cognition (i.e. agreement with the formal conditions of cognition). See Stang (2016a, ch. 6.5) for more on this issue.
28 A244/B302, 9:51, 28:554, 28:557, 28:796, 28:1015, 28:1157, 29:811.
29 Kant makes this point explicitly in the body of the Postulates: 'That in such a concept no contradiction must be contained is, to be sure, a necessary logical condition; but it is far from sufficient for the objective reality of the concept, i.e., for the possibility of such an object as is thought through the concept' (A220/B227–8).
30 Cf. Bxxvi, A244/B302, A596/B624.
31 See Stang (2016a, ch. 1.6) for more on the idea of 'real predicates' as adding to the intensional content of a concept. See also the criticisms in Bader (2018) and my reply in Stang (2018a).
32 Similarly, logical possibility is not a feature of objects directly. It makes no sense to say that an object is logically consistent: it is a feature of the *concept* of that object.
33 See Stang (2016a, ch. 7.3) for a more precise account of 'agreement.'
34 This reading is not uncontested. Kant's apparent agnosticism as to whether matter that lacks repulsive force, a capacity to see the future, and a faculty of telepathy are really possible (A222/B269–70) poses a problem for my reading, for it is hard to see why he would be agnostic as to their agreement with the forms of experience. See Abaci (2013) and Chignell (2014b) for critical discussion on this point, as well as my reply in Stang (2016a, 204n22).
35 This identification of modal concepts with hylomorphic concepts (matter/form) is confirmed in the 'Amphiboly' section of the *CPR* (A266/B322).
36 See Refl. 4119 and Refl. 6290, where Kant explicitly compares space as the ground of the real possibility of objects in space to his pre-Critical view that God is the ground of real possibility.
37 A111, B136. In both of these passages Kant also confirms that space and time are grounds of the real possibility of objects in space and time.
38 See Stang (2016a, ch. 9) for more discussion of this point.
39 The best overall guide to the controversy remains Ameriks (1982). For more contemporary accounts, from quite different perspectives, see Allais (2015) and Stang (2018b).
40 It is also an alethic modality because its dual—namely, formal real necessity—is alethic: if it is incompatible with the forms of experience that $\sim p$ then p (where p is restricted to claims about objects in space and time). This is what allows Kant to infer from claims about what is, or is not, compatible with the forms of experience (what is formally really possible or not possible) to what is the case about objects in space and time. Whether we count formal real possibility as a kind of 'metaphysical' possibility is a complicated question, in part because it is not entirely clear what 'metaphysical' possibility in contemporary metaphysics refers to. See Stang (2016a, ch. 9) for more on what notion in Kant's modal theory most closely resembles the contemporary notion of metaphysical modality.
41 In Stang (2016a), I argue at length for even more distinctions among kinds of real ossibility in the Critical system, e.g. 'empirical-causal' real possibility and 'nomic' real possibility. See McDaniel (2018) for some criticisms of this 'modal pluralist' reading, as well as my reply (Stang 2018c).
42 In the *CPR* I mainly have in mind the section 'The Possibility of Causality through Freedom Unified with the Universal Law of Natural Necessity' (A538/B566–A541/B569); and in the *CPrR,* 'On the Warrant of Pure Practical Reason to an Extension Which Is Not Possible to It in Its Speculative Use' (5:50–7) and the 'Critical Elucidation' that concludes the Analytic of Practical Reason (5:89–103).
43 A548/B576.
44 Fully explaining the conflict would require analyzing Kant's definition of (real) necessity in the *CPR*: 'that whose connection with the actual is determined in accordance with general conditions of experience is (exists) necessarily' (A217/B266). In this chapter, however, I have focused on Kant's conception of possibility. See Stang (2016a, ch. 7.4–7.5) for more on this notion of necessity. The most complete working out of the modal aspects of Kant's theory of freedom of which I am aware is Rosefeldt (2012).
45 The case for this is negative: in none of his discussions of reconciling empirical determinism with the possibility of having done otherwise does Kant consider the mere logical or formal real possibility of doing otherwise to be sufficient.

46 Cf. A538/B566–A541/B569
47 This theme is taken up in Stang (2016a, ch. 9). For an alternative view, see Chignell (2009).

References

Abaci, U. 2008. Kant's Theses on Existence. *British Journal for the History of Philosophy*, 16(3): 559–593.

Abaci, U. 2013. The Coextensiveness Thesis and Kant's Modal Agnosticism in the 'Postulates.' *European Journal of Philosophy*, 23(2): 129–158.

Abaci, U. 2014. Kant's Only Possible Argument and Chignell's Real Harmony. *Kantian Review*, 19(1): 1–25.

Allais, L. 2015. *Manifest Reality: Kant's Idealism and His Realism*. Oxford: Oxford University Press.

Ameriks, K. 1982. Recent Work on Kant's Theoretical Philosophy. *American Philosophical Quarterly*, 19(1): 1–24.

Bader, R. 2018. Real Predicates and Existential Judgements. *European Journal of Philosophy* 26(3): 1153–1158.

Baumgarten, A.G. 2013. *Metaphysics: A Critical Translation with Kant's Elucidations, Selected Notes and Related Materials*. C. Fugate and J. Hymers (trans. and eds.). London: Bloomsbury.

Beck, L.W. 1969. *German Philosophy before Kant*. Cambridge, MA: Harvard University Press.

Boehm, O. 2012. Kant's Regulative Spinozism. *Kant-Studien*, 103(3): 292–317.

Chignell, A. 2009. Kant, Modality, and the Most Real Being. *Archiv für Geschichte der Philosophie*, 91(2): 157–192.

Chignell, A. 2010. Real Repugnance and Belief about Things-in-Themselves: A Problem and Kant's Three Solutions. In J. Krueger and B. B. Lipscomb (eds.), *Kant's Moral Metaphysics*. Berlin: Walter de Gruyter, 177–210.

Chignell, A. 2011. Real Repugnance and Our Ignorance of Things-in-Themselves: A Lockean Problem in Kant and Hegel. *Internationales Jahrbuch des Deutschen Idealismus*, 7: 135–159.

Chignell, A. 2012. Kant, Possibility, and the Threat of Spinoza. *Mind*, 121(483): 635–675.

Chignell, A. 2014a. Kant and the 'Monstrous' Ground of Possibility: A Reply to Abaci and Yong. *Kantian Review*, 19(1): 53–69.

Chignell, A. 2014b. Modal Motivations for Noumenal Ignorance: Knowledge, Cognition, and Coherence. *Kant-Studien*, 105(4): 573–597.

Heimsoeth, H. 1956. *Metaphysik und Kritik bei. Chr. A. Crusius. Kant-Studien*, Ergänzungsheft 71. Cologne: Kölner Üniversitats-Verlag.

Hogan, D. 2009. Three Kinds of Rationalism and the Non-Spatiality of Things in Themselves. *Journal of the History of Philosophy*, 47(3): 355–382.

Kannisto, T. 2013. Modality and Metaphysics in Kant. In Stefano Bacin, Alfredo Ferrarin, Claudio La Rocca & Margit Ruffing (eds.), *Kant und die Philosophie in weltbürgerlicher Absicht. Akten des XI. Kant-Kongresses 2010*. Berlin: Walter de Gruyter.

Kannisto, T. 2016. Positio contra complementum possibilitatis – Kant and Baumgarten on Existence. *Kant-Studien*, 107(2): 291–313.

Kannisto, T. 2017. Kant on the Necessity of Causal Relations. *Kant-Studien*, 108 (4):495–516.

Kant, I. 1902. *Kants gesammelte Schriften*. Berlin-Brandenburg (formerly: Royal Prussian) Academy of Sciences (eds.). Berlin: Walter de Gruyter.

Kant, I. 1992. *Theoretical Philosophy, 1775–1770*. D. Walford and R. Meerbote (trans. and eds.). Cambridge: Cambridge University Press.

Kant, I. 1998. *Critique of Pure Reason*. P. Guyer and A. Wood (trans. and eds.). Cambridge: Cambridge University Press.

Leech, J. 2012. Kant's Modalities of Judgment. *European Journal of Philosophy*, 20(2): 260–284.

Leech, J. 2014. Making Modal Distinctions: Kant on the Possible, the Actual, and the Intuitive Understanding. *Kantian Review*, 19(3): 339–365.

Leech, J. 2017. Kant's Material Condition of Real Possibility. In Mark Sinclair (ed.), *The Actual and the Possible: Modality and Metaphysics in Modern Philosophy*. Oxford: Oxford University Press.

Leibniz, G.W. 1985. *Theodicy*. E. M. Huggard and A. Farrer (trans. and eds.). Chicago and La Salle, IL: Open Court.

Leibniz, G.W. 1989. *Philosophical Essays*. R. Ariew and D. Garber (trans. and eds.). Indianapolis: Hackett.

McDaniel, K. 2018. Nicholas Stang's *Kant's Modal Metaphysics*. *Kantian Review*, 23(3): 461–472.

Rosefeldt, T. 2012. Kants Kompatibilismus. In M. Brandhorst, M. Hahmann, and B. Ludwig (eds.), *Sind wir Bürger zweier Welten? Freiheit und moralische Verantwortung im transzendentalen Idealismus*. Frankfurt: Meiner.

Rosenkoetter, T. 2013. Towards a Non-Embarrassing Account of the Modal Functions of Judging. In S. Bacin, A. Ferrarin, C. LaRocca, and M. Ruffing (eds.), *Akten des XI. Internationalen Kant-Kongresses*. Berlin: Walter de Gruyter.

Stang, N. 2010. Kant's Possibility Proof. *History of Philosophy Quarterly*, 27(3): 275–299.

Stang, N. 2011. Did Kant Conflate the Necessary and the A Priori? *Noûs*, 45(3): 443–471.

Stang, N. 2015. Kant's Argument that Existence Is Not a Determination. *Philosophy and Phenomenological Research*, 91(3), 583–626.

Stang, N. 2016a. *Kant's Modal Metaphysics*. Oxford: Oxford University Press.

Stang, N. 2016b. The Unity of Kant's Critical Theory of Real Possibility. In M. Cresswell, E. Mares, and A. Rini (eds.), *Logical Modalities from Aristotle to Carnap: The Story of Necessity*. Cambridge: Cambridge University Press.

Stang, N. 2018a. *Kant's Modal Metaphysics*: A Reply to My Critics. *European Journal of Philosophy*, 26(3): 1159–1167.

Stang, N. 2018b. Kant's Transcendental Idealism. In E. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2016 edition), URL: http://plato.stanford.edu/archives/spr2016/entries/kant-transcendental-idealism/.

Stang, N. 2018c. Replies to Critics. *Kantian Review*, 23(3): 473–487.

Tolley, C. 2011. Kant on the Content of Cognition. *European Journal of Philosophy*, 20(4): 200–228.

Watkins, E. 2005. *Kant and the Metaphysics of Causality*. Cambridge: Cambridge University Press.

Wolff, C. 1965. *Verunünftige Gedancken von Gott, der Welt und der Seele des Menschen, auch allen Dingen überhaupt*. In J. École, H.W. Arndt, Ch. A. Corr, J. E. Hoffmann, and M. Thomann (eds.), *Gesammelte Werke*, Abt. I, Bd. 2. Hildesheim, Zürich, and New York: Georg Olms.

Wundt, M. 1939. *Die deutsche Schulmetaphysik des 17. Jahrhunderts*. Tübingen: J. B. C. Mohr.

Wundt, M. 1945. *Die deutsche Schulphilosophie im Zeitalter der Aufklärung*. Tübingen: J. B. C. Mohr.

Yong, P. 2014. God, Totality and Possibility in Kant's Only Possible Argument. *Kantian Review*, 19(1): 27–51.

Chapter 34
QUINE ON MODALITY

Roberta Ballarin

Willard van Orman Quine (1908–2000) is one of the most prominent philosophers of the 20th century. His contributions to logic, philosophy of science, epistemology, and ontology have been groundbreaking and remain extremely influential. Yet, when it comes to modality, Quine's work seems to fall sorely short on both fronts. It is fair to say that Quine's criticism of the modalities, at times voiced as a rebuke of modal logic itself, was poorly received exactly because it seemed to reveal an inability to keep up with the logical progress of his times. This logical progress was multifaceted. It included the technical development of systems of quantified modal logic and a model theory thereof, but also various logical reinterpretations of the modalities. In this chapter, we argue that Quine's opposition was aimed not so much at the technical developments, but rather at the reinterpretations of necessity that the new logic brought in its wake.

Quine's stand on the modalities is best understood against the background of logical empiricism, especially Carnap's systematic work in *Meaning and Necessity*. Quine shared the logical empiricists' strong distaste for traditional metaphysics, and, at least to start with, his aversion to the modalities, i.e., necessity and possibility, embodied this common anti-metaphysical stance. But the agreement ended there. Surely, the best-known feature of the Carnap-Quine philosophical controversy is Quine's critique of the epistemological distinction between analytic and synthetic truths. But even in the case of the modalities, Quine was critical of Carnap's logical-empiricist stand. Far from rejecting only old-fashioned modal metaphysics, Quine also censured any progressive attempt to reform necessity in terms of analyticity or logical truth. This censure was not well received. Thus, if there is an aspect of Quine's philosophical work whose influence has been negative rather than positive, it is on the topic of the modalities. It didn't help that Quine expressed his concerns as logical. Soon enough, technically savvy philosophers rebutted Quine's closed-mindedness towards modal logic, and his qualms were fast dismissed (Barcan Marcus 1967; Parsons 1967, 1969; Kaplan 1986). However, the baby was thrown out with the bathwater: Quine's most valuable interpretative insights remained long unnoticed (Burgess 1997; Neale 2000; and Ballarin 2004 are recent defenses of Quine). In what follows, we examine what Ruth Barcan Marcus famously called "Quine's animadversions on the modalities" thematically rather than historically (see Ballarin 2012 for a more detailed historical reconstruction). After a brief historical excursus, we introduce Quine's concerns about quantified modal logic. This will lead us to examine Quine's take on analyticity, essentialism, and their connections. We conclude the entry with a short presentation of Quine's own positive interpretation of necessity.

Historical background

The first half of the last century has witnessed the fast-paced development of different systems of modal logic. Starting with C.I. Lewis's propositional systems and ending with Carnap's and Barcan Marcus's independent construction of quantified modal logics (Lewis 1914, 1918; Lewis and Langford 1932; Barcan Marcus 1946; Carnap 1946). Lewis's intention in devising a series of modal propositional systems was to find out the best calculus for strict, logical implication, as opposed to the material implication of ordinary propositional logic. Consequently, the idea was to interpret the operator of necessity as reflecting at the object level the metalinguistic predicate "is logically true."

In 1946 both Barcan Marcus and Carnap devised modal systems combining Lewis's propositional modal logic with quantification. At least for Carnap, it is clear that the interpretive intention remained at heart the same as Lewis's. The necessity operator of the object language was read as logical necessity and linked to the semantic notion of logical truth:

> [T]he guiding idea in our constructions of systems of modal logic is this: a proposition *p* is logically necessary if and only if a sentence expressing *p* is logically true. That is to say, the modal concept of the logical necessity of a proposition and the semantical concept of the logical truth or analyticity of a sentence correspond to each other.
>
> (Carnap 1946: 34)

According to Carnap, however, there is no independent notion of logical truth. Logical truths are just a subclass of analytic truths, namely, those true in virtue of the meaning of the logical terms alone. So, via logical truth, necessity is ultimately understood as analyticity. Additionally, Carnap introduces a formal notion of L-Truth as follows:

> Our concept of L-truth is … intended as an explicatum for the familiar but vague concept of logical or necessary or analytic truth as explicandum. This explicandum has sometimes been characterized as truth based on purely logical reasons, on meaning alone, independent of the contingency of facts.
>
> (1956: 10)

In a modal system, like *S5*, the operator of necessity is meant to capture at the object level this formal counterpart of analyticity; thus, "□ *p*" is true iff "*p*" is L-True (see Carnap 1956, ch. 5).

As a formal semantic notion intended to explicate the intuitive notion of analyticity as truth in virtue of meaning alone, L-Truth is given a rigorous model theoretic analysis as truth across all state-descriptions. For the purposes of this chapter, we may take state-descriptions as models of first-order logic, which makes it clear that L-Truth is a form of model-theoretic validity, that is, truth in all models (see Ballarin 2005 on the relation between state-descriptions and models). The essential point is that Carnap interprets the necessity operator as symbolizing at the object level (the model-theoretic counterpart of) the semantic notion of logical truth.

Quine, who had been a student of C.I. Lewis at Harvard, had been in contact with Carnap since the early 1930s. He certainly endorsed some of the fundamental tenets of Carnap's logical empiricism, particularly the rejection of obscure metaphysical notions, such as what we nowadays call metaphysical necessity, which he labeled Aristotelian essentialism. Thus, at least in principle, the general spirit of the Lewis-Carnap enterprise should not have displeased him. Yet, immediately after the publication of Barcan Marcus's and Carnap's 1946 papers, Quine wrote

"The Problem of Interpreting Modal Logic" whose main thesis seems to be that the attempt to mix a clear semantic interpretation of the modalities with quantification is bound to fail. In Quine's own (later) words:

> Evidently this reversion to Aristotelian essentialism … is required if quantification into modal contexts is to be insisted on. An object, of itself and by whatever name or none, must be seen as having some of its traits necessarily and other contingently, despite the fact that the latter traits follow just as analytically from some ways of specifying the object as the former traits do from other ways of specifying it … Essentialism is abruptly at variance with the idea, favored by Carnap, Lewis, and others, of explaining necessity by analyticity … For the appeal to analyticity can pretend to distinguish essential and accidental traits of an object only relative to how the object is specified, not absolutely. Yet the champion of quantified modal logic must settle for essentialism.
>
> (Quine 1953: 155; all quotes from "Reference and Modality" come from the 1961 revised version)

What did Quine mean?

Quantified modal logic

Quine's 1947 paper starts by saying: "There are logicians, myself among them, to whom the ideas of modal logic … are not intuitively clear until explained in non-modal terms" (1947: 43). We need then a non-modal reinterpretation of the obscure modal idioms. We have seen Carnap's proposal: reinterpret necessity as analyticity. However, charges Quine, this reinterpretation of necessity cannot be combined with quantification. At the very least, there is no obvious interpretation of quantified-in modal sentences, i.e., formulas like "$\exists x \Box Fx$" where variables in the scope of a modal operator are bound to an external quantifier – what Quine labeled the third grade of modal involvement (at the first grade we allow only semantic predicates like "is logically true," at the second grade we introduce the modal operators but no quantifiers; see Quine 1953b). In Quine's own words: "[T]he important point to observe is that granted an understanding of the modalities … and given an understanding of quantification ordinarily so called, we do not come out automatically with any meaning for quantified modal sentences" (Quine 1953: 150). And even more forcefully:

> It is logically possible to like modal logic without confusing use and mention. You could like it because, apparently at least, you can quantify into a modal context by a quantifier outside the modal context, whereas you obviously cannot coherently quantify into a mentioned sentence from outside the mention of it. Still, man is a sensemaking animal, and as such he derives little comfort from quantifying into modal contexts that he does not think he understands.
>
> (Quine 1961: 323)

This is Quine's original problem with the third grade of modal involvement: no clear understanding of quantified-in modal sentences is forthcoming. To make sense of Quine's concern, we need to keep in mind that analyticity (truth in virtue of meaning alone) and logical truth are semantic properties. They are *de dicto* properties, where a *de dicto* property is a property of a *dictum*, i.e., of a verbal expression. Moreover, the verbal expressions in question are (closed) sentences. This is the case because analyticity and logicality are modes of truth, and truth is a

property of sentences. In other words, if analyticity is a mode of truth, and truth is a property of sentences, analytic truth is also a property of sentences. Recall now Carnap's proposal that "a proposition *p* is logically necessary if and only if a sentence expressing *p* is logically true." This means that "□*p*," where "□" represents logical necessity, is true iff "*p*" is logically or analytically true. Quine's reply is that semantic interpretations of the modal operators are not immediately extendable once quantification is added. But why so? Where does the obstacle arise?

To understand Quine, one must keep in mind that his concern is not with the uninterpreted modal formulas. Nor is he speaking of the proof theory or the model theory of formal systems of modal logic. Throughout his career, Quine has emphasized that his focus is on the intended interpretation of quantified-in modal sentences. Here is a representative statement:

> I've never said or, I'm sure, written that essentialism could be proven in any system of modal logic whatever. I have never even meant to suggest that any modal logician even was aware of the essentialism he was committing himself to, even implicitly in the sense of putting it into his axioms. I'm talking about quite another thing – I'm not talking about theorems, I'm talking about truth, I'm talking about true interpretation.
>
> (1962: 140)

In other words, the question is: can we make sense of the intended English translation of such formulas? The answer to this question, claims Quine, is no, if the modalities are interpreted Carnap's way and the quantifiers are also given their ordinary meaning. For example, if "*F*" symbolizes "is human," the intended translation of a formula like "$\exists x \Box Fx$" is something along the following lines: "There is an object such that it is logically/analytically true that it is human" or even "There is an object which is logically/analytically human." However, given that analyticity is a mode of truth, and truth is a property of sentences, it makes no sense to say of an object, as opposed to a sentence, that it is analytically anything. To paraphrase Quine, if "analytically true" is a predicate of sentences, then being-analytically-human, or even being-analytically-self-identical, are "quotation breaking pseudo-properties" (see Quine 1961: 325, where the point is made for "tautological").

There are ways of bypassing this interpretative impasse, primarily by denying Quine his preferred, object-oriented reading of the quantifiers. So Marcus proposed a substitutional interpretation of the same. We can read Marcus's suggestion as an attempt to align the interpretation of the quantifiers with the semantic interpretation of the modal idioms. Substitutionally, "$\exists x \Box Fx$" is true if and only if "$\Box Fa$" is true for some constant "*a*" in place of "*x*," thus, ultimately when "*Fa*" is analytically true. As Marcus notices, this requires no mysterious property to be conferred on an object (Quine 1962: 132). Quine acknowledges the point:

> Professor Marcus's proposed reinterpretation of existential quantification is this: the quantification is to be true if and only if the open sentence after the quantifier is true for some substitute for the variable of quantification. Now this is, I grant, an intelligible reinterpretation, and one that does not require objects, in any sense, as values, in the real sense, of the variables of quantification.
>
> (1961: 328)

Quine gives in, but he reiterates that his concerns are ontological, and as such they arise only under the intended interpretation of the quantifiers:

> I grant further that essentialism does not come in if we interpret quantification in your new way. By quantification I mean, quantification in the ordinary sense … If on the

> other hand we do not have quantification in the ordinary sense then I have nothing to suggest at this point about the ontological implications or difficulties of modal logic. The question of ontology wouldn't arise if there were no quantification of the ordinary sort.
>
> (Quine 1962: 134)

So, the problem is not to supply a – that is, any – coherent interpretation of quantified modal logic. The problem concerns the coherence of the intended interpretation. Quine's blunt conclusion is the following:

> I think there are many ways you can interpret modal logic. I think it's been done. Prior has tried it in terms of time and one thing and another. I think any consistent system can be found an intelligible interpretation. What I have been talking about is quantifying, in the quantificational sense of quantification, into modal contexts in a modal sense of modality.
>
> (Quine 1962: 143)

Quine thus admits that coherent, non-essentialist interpretations of modal logic are possible. However, he insists that the intended objectual interpretation of the quantifiers induces an unintended essentialist interpretation of the modal operators.

Analyticity and essentialism

To better understand Quine's contention, let us restate that the problem strikes at the third grade of modal involvement. Quine challenges the feasibility of upholding the analytic interpretation of the modal operators in the face of the objectual interpretation of the quantifiers. Quine's better-known qualms against analyticity *in and of itself* are irrelevant to this question: "We cannot in conscience blame these varied sorrows of modality on analyticity" (Quine 1960: 200). In fact, at least in the 1940s, Quine regarded analyticity as clearer than any other interpretation of necessity: "As for other notions of necessity, possibility, etc., for example, notions of physical necessity or possibility, the first problem would be to formulate the notions clearly and exactly" (Quine 1943: 124). And:

> The notion of analyticity thus appears … to lack a satisfactory foundation. Even so, the notion is clearer to many of us, and obscurer surely to none, than the notions of modal logic; so we are still well advised to explain the latter notions in terms of it. This can be done … so long as modal logic stops short of quantification theory.
>
> (Quine 1947: 45)

Quine's hope at the time was to naturalize meanings by identifying them with classes of synonymous expressions, and to define synonymy in psychological/behavioral terms:

> Synonymy, like other linguistic concepts, involves parameters of times and persons, suppressible in idealized treatment: the expression x is synonymous with the expression y for person z at time t. A satisfactory definition of this tetradic relation would no doubt be couched, like those of other general concepts of general linguistics, in behavioral terms.
>
> (Quine 1947: 44)

As we will see, Quine would soon change his mind on the prospects of an extensional, naturalized reinterpretation of analyticity. Nonetheless, his concerns about quantified modal discourse are best understood under the assumption that the analytic interpretation of necessity is to be maintained if at all possible.

To make sense of Quine's contention, we must also understand what Quine means by essentialism. In particular, we need to figure out the ways in which the essentialist interpretation of necessity is taken to differ from the analyticity-based reading. We have already met a typical Quinean characterization of essentialism: "An object, of itself and by whatever name or none, must be seen as having some of its traits necessarily and other contingently." At the end of "Three Grades," Aristotelian essentialism is so defined:

> This is the doctrine that some of the attributes of a thing (quite independently of the language in which the thing is referred to, if at all) may be essential to the thing, and others accidental. E.g., a man, or talking animal, or featherless biped (for they are in fact all the same *things*), is essentially rational and accidentally two-legged and talkative, not merely qua man but qua itself. More formally, what Aristotelian essentialism says is that you can have open sentences—which I shall represent here as "*Fx*" and "*Gx*"—such that $(\exists x)$ (nec Fx . Gx . ~nec Gx).
>
> (Quine 1953b: 176)

We need to distinguish three features of this characterization. Essentialism is taken to be an (i) invidious, (ii) *de re*, (iii) non-representational view of necessity. First, essentialism is invidious in the following sense: an object bears *some and only some* of its properties essentially. Invidiousness is not unique to essentialism. Analyticity is also an invidious notion, insofar as it is part and parcel of the doctrine of analyticity that *some and only some* truths are analytic. The first dogma of empiricism does not consist simply in the endorsement of the meaningfulness of the notion of analyticity; it is rather "a belief in some fundamental cleavage between truths which are *analytic* … and truths which are *synthetic*" (Quine 1951: 20). Invidiousness is certainly not congenial to Quine's own philosophical outlook. Indeed, Quine is well known for his opposition to the cleavage between science and philosophy, based as it is on an alleged invidious distinction between two different sorts of truths. Quine finds all such separations uncongenial: a priori versus a posteriori, analytic versus synthetic, necessary versus contingent, and of course essential versus accidental. To all such distinctions Quine is firmly opposed. Nonetheless, in Quine's eyes, *necessity* is itself a fundamentally invidious notion: "A striking divergence between Kaplan's intuitions and mine is keynoted midway in his section XVIII, where he writes that 'some metaphysician may assert that all truths are metaphysically necessary … [a]nd it wouldn't be an abandonment of modality.' I would not see what point the metaphysician was trying to make, for want of invidious distinctions" (Quine 1986: 293); therefore, any adequate interpretation of necessity is bound to be invidious, too.

What separates essentialism from analyticity, then, are its two other key features: it is a non-representational, and *de re* interpretation of necessity. Essentialist necessity is *de re* in the sense that the fundamental cleavage it envisions is not so much, as in the case of analyticity, between two different sorts of truths, but rather between two different sorts of properties: those that apply essentially versus the mere accidental ones, at least relative to an object. In this sense, essentialism is fundamentally predicative: it envisions two distinct modes of predication, not of truth. Moreover, an object bears some of its properties essentially and some accidentally *in and of itself*, that is, independently of any (linguistic) representation of the object. So, the fundamental opposition between the analytic and the essentialist interpretations of necessity can be cashed out as the *de dicto*/*de re* distinction. This is fine, as long as we keep in mind that *de dicto* necessity is to

be understood as both representational and sentential, while *de re* necessity is predicative and non-representational. Logically, these two aspects of the opposition may be disentangled, yet Quine never envisioned non-representational yet non-predicative views of necessity (for example, necessity as a mode of holding for facts, independently of how such facts are represented, but also independently of the ways in which the objects and properties, if any, that make up such facts are related). This may or may not be a philosophical shortcoming of Quine's, yet it must be borne in mind that Quine's concern is with real, historical doctrines, not merely consistent positions in logical space. Given these characterizations, we can understand Quine's contention that the quantifiers cannot be combined with the analytic interpretation of necessity as an explicit rejection of predicative forms of representational necessity. This combination, Quine seems to suggest, is incoherent.

Let us now return to our main question, the alleged incoherence of quantification into the logical, analytic modalities. Recall Quine's reaction to Marcus's proposal in the early 1960s: yes, we can combine these modalities with substitutional quantification, but not with an objectual interpretation of the quantifiers. This, however, had not always been Quine's view. When Church (1943) proposed to interpret the quantifiers as objectual but ranging over a domain of intensions, Quine endorsed the suggestion as an adequate interpretation. This is how Quine put the matter when commenting on Carnap's project in *Meaning and Necessity*: "I agree that such an adherence to an intensional ontology … is indeed an effective way of reconciling quantification and modality. The cases of conflict between quantification and modality depend on extensions as values of variables" (Quine's comments are published in Carnap 1956 [1947]: 197).

Quine's main idea seems to be that letting the quantifiers range over intensional entities (e.g., Church's and Carnap's meanings, intensions or individual concepts, but words would also do) preserves their intended object-oriented, as opposed to substitutional, interpretation. Still the semantic nature of the entities we are quantifying over well combines with the semantic reading of the modal operators. Let us try to reconstruct the details.

Recall that according to the substitutional interpretation of the quantifiers, "$\exists x \Box Fx$" is true if and only if "$\Box Fa$" is true for some constant "a" substituted for the variable "x" in "$\Box Fx$." In the substitutional interpretation, the quantifiers are not taken to range over a domain of entities; instead they are associated to a substitution class of constants. This is why Quine regards the substitutional interpretation as inadequate: it does not capture the meaning of the existential quantifier as "There exists something such that." In the Church-Carnap proposal, instead, we maintain a classical domain (not a substitution class) for the quantifiers, despite the fact that the elements in the domain are semantic entities. This way, the quantifiers can be read, so to speak, as about those entities. Moreover, Church and Carnap quantify over intensions, as opposed to words. (This, at least, is Quine's interpretation of Carnap.) Thus, according to the intensional interpretation, "$\exists x \Box Fx$" is true if and only if there is some intension a which satisfies the open formula "$\Box Fx$." Intuitively, meanings can be analytically F, namely, F by definition, thus supplying a predicative form of representational necessity. The puzzling question we need to face is the following: Why did Quine become dissatisfied even with this proposal?

There is no easy answer to this question. Whenever he explains his change of mind, Quine mentions – as if crucial to such a change – his coming to realize that intensional entities can be specified in non-analytically equivalent ways. Essentialism, he says, consists in "adopting an invidious attitude toward certain ways of uniquely specifying x … and favoring other ways … as somehow better revealing the 'essence' of the object" and:

> we can see pretty directly that any quantified modal logic is bound to show such favoritism among the traits of an object; for surely it will be held, for each thing x, on the

> one hand that necessarily ($x = x$) and on the other hand that ~ necessarily [$p \wedge (x = x)$], where "p" stands for an arbitrary contingent truth.
>
> (Quine 1953: 155)

What is puzzling about all this is that, as we have seen, analyticity is itself an invidious notion, so favoritism is to be expected. It is analytically true that bachelors are unmarried and only factually true, say, that they are all happy. So, let *a* be the individual concept *the youngest living bachelor*. Then "□ *Fa*" holds when "*F*" is interpreted as "unmarried," while "*Ga*" does hold, but "□ *Ga*" does not, when "*G*" is interpreted as "happy" (the predicates must be semantically reinterpreted too, so a concept satisfies "*F*" iff all its members are *F*); hence, we can derive ($\exists x$) (nec *Fx* . *Gx* . ~nec *Gx*): Quine's definition of essentialism is satisfied. Yet, this surely ought not to suffice to make the necessity in question of the essentialist sort: in this case the reason why "□ *Fa*" holds is that "*Fa*" is analytically true given the definition or analysis of *a*. If essentialism can be so easily derived, it may very well be the case that Quine's formal definition of essentialism, not quantified modal logic, is to be blamed. More could be said on Quine's behalf, yet this is no place to attempt a, perhaps impossible, full defense of Quine's third-grade challenge. Let us then grant that, despite Quine's protests to the contrary, the intended interpretation of the quantifiers can indeed be mixed with analytic necessity, that is to say, Quine's third-grade interpretative concerns must be dismissed.

Nonetheless, in what follows we must return to the question of analyticity *per se*, independently of quantification. We have seen that in the 1940s, Quine endorsed the analytic interpretation of necessity over all others. Analyticity is surely preferable to the "metaphysical jungle of Aristotelian essentialism," especially so if a scientifically respectable behavioristic analysis of synonymy is available. Yet, as is well known, in "Two Dogmas of Empiricism" Quine gives up the hope of a clear explication of analyticity and finally rejects the analytic-synthetic distinction. Curiously enough, it is in this paper about analyticity that Aristotelian essentialism makes its initial appearance:

> The Aristotelian notion of essence was the forerunner, no doubt, of the modern notion of intension or meaning. For Aristotle it was essential in men to be rational, accidental to be two-legged. But there is an important difference between this attitude and the doctrine of meaning. From the latter point of view it may indeed be conceded (if only for the sake of argument) that rationality is involved in the meaning of the word "man" while two-leggedness is not; but two-leggedness may at the same time be viewed as involved in the meaning of "biped" while rationality is not. Thus, from the point of view of the doctrine of meaning it makes no sense to say of the actual individual, who is at once a man and a biped, that his rationality is essential and his two-leggedness accidental or vice versa. Things had essences, for Aristotle, but only linguistic forms have meanings. Meaning is what essence becomes when it is divorced from the object of reference and wedded to the word.
>
> (Quine 1951: 22)

It is natural to read this passage as claiming that meanings are an improvement over essences. Whereas Aristotle attributed essences to objects directly, thus rendering necessity an absolute and absolutely objective feature of reality, Quine's contemporaries employ meanings as filters between objects and necessity. Necessity is thus both relativized and detached from the real concrete world. This reading emphasizes the "important difference" between the Aristotelian and the contemporary attitude towards necessity, and it underscores the philosophical progress which has taken place from Aristotle to Carnap. Nonetheless, the last sentence of this passage

("meaning is what essence becomes") forcefully suggests a less benevolent interpretation: meanings are nothing but the essences of words. Meanings, Quine seems to say, are the last refuge of the invidious attitude philosophers would do well to relinquish. If this is the case, we start to make sense of Quine's claim that the only result achieved by the logician who quantifies over intensions is to limit essentialism to these entities. This may not be the inevitable result of *de re* modal speech, yet it is the conclusion that Quine reaches about analyticity itself. If analyticity is to be grounded in the meanings or definitions of words, then it is itself a form of essentialism. Quine is reversing the link between analyticity and necessity: it now looks like the reason why "*Fa*" is analytically true is that "*a*" is essentially and so necessarily defined as "*F*."

Necessity

Is there any place at all for necessity in Quine's own world view? In 1963, Quine recognizes that in common parlance the adverb "necessarily" often attaches to direct predications, yet urges us to reconsider this form of speech:

> Someone says, "Surely it has spots," "Necessarily it will have spots," when told that he is about to be shown a leopard. Now instead of viewing the adverb "necessarily" as governing outright the simple prediction of spots, we can understand it as governing, implicitly, the conditional sentence as a whole: "If it is a leopard, it has spots."
>
> (Quine 1963: 69)

This rejection of *de re* attributions of necessity is not surprising. It is also to be expected that Quine's preference for this *de dicto* reformulation has nothing to do with analyticity. There are no truths by definition, and no invidious distinctions between truths to be upheld; thus, claims Quine, the highest form of necessity is natural necessity, itself understood in deflationary, Humean terms:

> The doctrine that necessity is no more than regularity was expounded by David Hume more than two centuries ago. His, indeed, was the battle cry, "There are no necessary connections of matters of fact." He meant, of course, that there are none if 'necessary' is taken as claiming more than generality, or regularity. I am not taking it as claiming more.
>
> (Quine 1963: 71)

> In principle, therefore, I see no higher or more austere necessity than natural necessity; and in natural necessity, or our attributions of it, I see only Hume's regularities.
>
> (Quine 1963: 76)

It is also not surprising that Quine never renounced his conviction that the intended interpretation of *de re* modal talk is essentialist. What is definitely unexpected, however, is Quine's claim, as late as 1986, that ordinary discourse, with its commitment to essentialism, is more useful than a benign reinterpretation of necessity as plain logical truth:

> [T]his course is committed to an invidious essentialism … in which an essential trait of one object can be an accident of another. Moreover it is this pattern, rather than the benign one, that I see as useful and customary in daily discourse, though only meaningful in context.
>
> (Quine 1986: 292)

We can understand this late recourse to context-relative meaningfulness as Quine's attempt to make sense of *de re* modal talk, despite his rejection of the essentialist world view. As Humean regularities lend meaning to *de dicto* modal talk, while eschewing analyticity, the contextualization of *de re* attributions of necessity lends them meaning while still managing to avoid essentialism. Quine seems to have found a way of reconciling himself with modal discourse, after all.

References

Ballarin, R. (2004) "The Interpretation of Necessity and the Necessity of Interpretation," *Journal of Philosophy* 101*:* 608–38.

———. (2005) "Validity and Necessity," *Journal of Philosophical Logic* 34*:* 275–303.

———. (2012) "Quine on Intensional Entities: Modality and Quantification, Truth and Satisfaction," *Journal of Applied Logic* 10: 238–49.

Barcan Marcus, R. (1946) "A Functional Calculus of First Order Based on Strict Implication," *Journal of Symbolic Logic* 11: 1–16.

Barcan Marcus, R. (1967) "Essentialism in Modal Logic," *Noûs* 1: 91–6.

Burgess, J. P. (1997) "Quinus ab Omni Naevo Vindicatus," in A. A. Kazmi (ed.) *Meaning and Reference, Canadian Journal of Philosophy*, Supp. Vol. 23: 25–65.

Carnap, R. (1946) "Modalities and Quantification," *Journal of Symbolic Logic* 11: 33–64.

Carnap, R. (1956 [1947]) *Meaning and Necessity*. Chicago: University of Chicago Press. 2nd edition with supplements.

Church, A. (1943) "Review of Quine (1943)," *Journal of Symbolic Logic* 8: 45–7.

Kaplan, D. (1986) "Opacity," in L. E. Hahn and P. A. Schilpp (eds.) *The Philosophy of W.V. Quine*, La Salle, IL: Open Court, pp. 229–89.

Lewis, C. I. (1914) "The Calculus of Strict Implication," *Mind* 23: 240–47.

——— (1918) *A Survey of Symbolic Logic*, Berkeley: University of California Press.

Lewis, C. I., and Langford, C. H. (1932) *Symbolic Logic*, London: Century.

Neale, S. (2000) "On a Milestone of Empiricism," in A. Orenstein and P. Kotakto (eds.) *Knowledge, Language, and Logic*, Great Britain: Kluwer, pp. 237–346.

Parsons, T. (1967) "Grades of Essentialism in Quantified Modal Logic," *Noûs* 1: 181–91.

———. (1969) "Essentialism and Quantified Modal Logic," *Philosophical Review* 78: 35–52.

Quine, W.V. (1943) "Notes on Existence and Necessity," *Journal of Philosophy* 40: 113–127.

——— (1947) "The Problem of Interpreting Modal Logic," *Journal of Symbolic Logic* 12: 43–48.

——— (1951) "Two Dogmas of Empiricism," *Philosophical Review* 60: 20–43.

——— (1953) "Reference and Modality," in *From a Logical Point of View*. Cambridge, MA: Harvard University Press. 2nd revised edition (1961).

——— (1953b) "*Three Grades of Modal Involvement*," in *Proceedings of the 11th International Congress of Philosophy*, Brussels, vol. 14, Amsterdam: North-Holland Publishing Co. Reprinted in Quine (1966a), pp. 156–74.

——— (1960) *Word and Object*, Cambridge, MA: MIT Press.

——— (1961) "Reply to Professor Marcus," *Synthese* 13: 323–30.

——— (1962) "Discussion on the Paper of Ruth B. Marcus," *Synthese* 14: 132–43.

——— (1963) "Necessary Truth," A Forum Lecture, broadcast by the Voice of America. Printed in Quine (1966), pp. 68–76.

——— (1966) *The Ways of Paradox*, New York: Random House.

——— (1986) "Reply to David Kaplan," in L. E. Hahn and P. A. Schilpp (eds.) *The Philosophy of W.V. Quine*, La Salle, IL: Open Court, pp. 290–94.

Chapter 35
KRIPKE ON MODALITY

John P. Burgess

Saul Kripke's published contributions to the theory of modality include both mathematical material (Kripke 1959a, 1963a, 1963b, and more) and philosophical materal (Kripke 1971, henceforth N&I; Kripke 1980, henceforth N&N). Here the latter must be given primary emphasis, but the former will not be entirely neglected.

Aposteriori necessity

Kripke's overarching philosophical point is that the necessary is not to be confused with the apriori, but preliminary explanations are needed before his contribution can be specified more precisely. To begin with, the linguistics literature (Palmer 1986) distinguishes three flavors of modality: deontic, epistemic/evidential, and dynamic. The pertinent notions of necessity are the *obligatory*, or *what must be if obligations are fulfilled*; the *known*, or *what must be given what is known*; and the *inevitable*, in the sense of what *cannot fail and could not have failed to be, no matter what*. This last is the only notion of necessity of interest in our present context. As Kripke emphasizes, this notion of the necessary and its correlative notion of the contingent are metaphysical concepts. They are as such at least conceptually distinct from the epistemological notions of the apriori and aposteriori and the semantical notions of the analytic and synthetic. Before Kripke this was never completely forgotten, but its importance was arguably grossly underestimated. Prior to Kripke's work it was the near-unanimous opinion that the analytic is included in the apriori, which in turn is included in the necessary (Quine 1960: 59). It was the majority opinion that these inclusions reverse, making the three classifications coextensive; but there were dissenters, conscious of being in the minority (Kneale and Kneale 1962: 637–639).

Kripke himself cites, as a putative counterexample from the previous literature to the inclusion of the necessary in the apriori, the following mathematical case. It is widely held that in mathematics the necessary coincides with the true, and the apriori coincides with the provable; but it is also widely held that the work of Kurt Gödel shows that there are unprovable truths, though there is an obvious difficulty in trying to exhibit a specific instance. Any mathematical truth not provable would be a necessity not apriori. Kripke's main philosophical contribution to the study of modality has been to give very different specific examples supposed to be necessary but aposteriori (and therefore synthetic), examples that have been very widely found to be convincing.

Before turning to these examples, however, two small points should be mentioned. First, Kripke tentatively suggests that there may also be examples that are apriori but contingent (and therefore synthetic). In a late work (Kripke 2011) he comes close to endorsing Harry Frankfurt's proposal that Descartes's *I think* is such a case. But Kripke's claims on this head are always guarded, and he never treats the matter as much more than a curiosity. Second, many distinguish the notion of interest here, the metaphysically necessary, that which could not have failed to be *no matter what*, from the physically necessary, that which could not have failed to be *without a violation of the laws of nature*, considering the latter weaker on the grounds that the laws of nature could have been other than as they are. Kripke, in the addenda to N&N, indicates he is neutral on this point.

Returning, then, to putative examples of aposteriori necessities, Kripke's discussion of them is intertwined with discussion of issues about names and denotation or reference. Among his examples of aposteriori necessities, a distinction must be made (following Salmon 1981): some emerge directly from his views on naming, while others depend on additional intuitions. Inversely, some of his views on naming directly pertain to his examples of aposteriori necessities, while others are at most indirectly relevant.

Naming and identity

Kripke's most basic examples are identity statements involving two proper names for the same item, as with *Hesperus is Phosphorus* or *Cicero is Tully*. In the former case the two names derive from two acts of naming, one in the evening, one in the morning (for etymologically the names mean, nearly enough, *evening star* and *morning star*); in the latter case they derive from two simplifications of the same original complex tripartite name (*Marcus Tullius Cicero*). That such identity statements are aposteriori, Kripke takes to be comparatively uncontroversial. Subsequently, some have argued (Soames 2005) that they are apriori, while still agreeing with Kripke that non-identities such as *Junius is not Burke* are aposteriori, on highly theoretical grounds based on a certain conception of coarse-grained propositions. That matter is a subtlety that I will not go into here.

That the identity statements, and likewise non-identity statements, are necessary is argued by Kripke to be a consequence of proper names being *rigid designators*, meaning that they denote the same thing whether speaking of how things actually are or of how things actually aren't but potentially could have been—or in jargon, whether speaking of the actual world or of some non-actual possible world. Kripke warns that what is at issue is what a name *is* used to denote when *we* are speaking *of* some counterfactual situation, and not what a phonetically and orthographically indistinguishable name *might be* used to denote by hypothetical speakers when *they* were speaking *in* that counterfactual situation. Bearing that in mind, let M and N be proper names, and suppose they denote the same thing with respect to, or when speaking of, the actual world @. Then, setting aside subtleties about possible non-existence in some worlds, they denote the same things with respect to, or when speaking of, any possible world *w*. For what M denotes with respect to *w* is by rigidity what M denotes with respect to @, which is by assumption what N denotes with respect to @, which is by rigidity what N denotes with respect to *w*.

In a rigorous mathematical treatise, whenever a constant is introduced, a uniquely identifying description is introduced along with it: the description serves as a definition of the constant, the constant as an abbreviation of the description. Kripke argues that names in natural language do not work this way. It is not enough to allow, as previous writers on the topic over the period 1957–1962, discussed in N&N, had done, that the uniquely identifying description

associated with a name may vary from speaker to speaker and occasion to occasion, or that it may be more a matter of a cluster of descriptions than of a single description, or that the description may make one speaker's reference parasitical on a previous speaker's reference, or that the description may serve to determine the reference of a name without constituting its meaning. Kripke holds, rather, in agreement with Donnellan (1970), whom he cites as having independently arrived at similar ideas, that there simply need not be any associated uniquely identifying description: that a name denotes the item it does, not because that item uniquely satisfies some associated description, but because of the history of the use of the name. This Kripke pictures as a "chain of communication" in which an "initial baptist" picks out an item by ostension or description and introduces an expression as a name for it, after which each subsequent speaker takes over the expression with the intention of using it to name whatever the speaker it was taken over from was using it to name. No uniquely identifying description need be associated with the name in order for it to continue denoting what it was introduced to denote.

According to Kripke, speakers often can associate no description with a name that is uniquely true of its bearer, and yet they do succeed in using the name to denote that bearer. In an exceptional case like *Christopher Columbus*, perhaps each of us does associate a similar uniquely identifying description with the name, some variant of *the person most responsible for large-scale contact between the eastern and western hemispheres.* In a more ordinary case like *Stephen Weinberg*, some may only be able to say *a famous physicist*, which is true of him but not uniquely so, or to say *the author of the quark hypothesis*, which is not true of him at all. And yet they are nonetheless able to say these less-than-uniquely-true things *about him* using his name. These are important and influential views, but their bearing on aposteriori necessity is less direct than that of the doctrine of rigidity.

Genealogy and composition

A second type of Kripkean example concerns pedigree. Avoiding giving royals their titles, in order to avoid issues about titles as opposed to names, consider Elizabeth Windsor, her father George Windsor, and the latter's contemporary Harry Truman. Kripke takes the following to be an aposteriori necessity: *EW is the daughter of GW and not of HT.* The example is aposteriori, it is claimed, because if a tabloid published a rumor to the contrary, the libel might not be refutable apriori by pointing out an internal contradiction, but only aposteriori by DNA evidence. The example is necessary, it is claimed, because in no counterfactual scenario, however far-fetched, would a daughter of HT be anything more than a substitute for EW. This is obviously so in a scenario in which EW is stillborn and for high reasons of state some by-blow of HT is secretly adopted and passed off as his daughter by GW, being given the name and place of EW; but it is plausibly claimed to be equally so in any other scenario.

This does not follow simply from names lacking associated uniquely identifying descriptions, or being rigid designators. The intuition behind the example seems to be something like this: if we imagine backing up the train of history to a time before EW's birth or conception, and then going forward along a different track (N&N: note 57), the only way we will have of recognizing an individual on this alternative track as EW will be by her originating from the same gametes as EW. Going beyond anything explicitly said by Kripke, the general principle would seem to be:

(1) If an individual actually has a certain ancestry, then an actual or hypothetical individual with a different ancestry does not or would not count as *that* individual.

Kripke allows that *if* we had a clear idea of the soul or mind as an independent spiritual entity, his conclusions might be challenged (N&N: note 77), but he denies that we have any such clear idea.

A third type of Kripkean example concerns constitution. Kripke suggests that so-called natural-kind terms behave in many ways like proper names. They come to denote what they do by an analogous process: an initial baptist picks out a specimen or sample and introduces an expression to denote the kind of thing or stuff of which it is an example, and then the word is transmitted with the intention of preserving its reference. In some cases we will all associate some uniquely identifying description with the term, but in many we may not: water being a much more familiar substance to most of us than is quicklime, *water* may be like *Columbus*, but *quicklime* more like *Weinberg*. Some examples involving natural kind terms, such as *gorse = furze* and *tungsten = wolfram*, are just like *Hesperus = Phosphorus*. Other examples appear to depend on further intuitions.

The case most often considered is *water* = H_2O. This is not at all like *Hesperus = Phosphorus*, because H_2O is not at all like a name: it is part of a systematic notation with a definite descriptive sense. It would be less misleading to spell out the meaning of the formula and state the example thus: *water is a compound of hydrogen and oxygen in a ratio of two to one.* For some purposes it might be less misleading to substitute an example involving a less *Columbus*-like, more *Weinberg*-like term: *quicklime is a compound of calcium and oxygen in a ratio of one to one*. The example is aposteriori because the chemical formulas for substances familiar before the rise of modern chemistry were empirical discoveries by the pioneers of atomic and molecular theory. According to Kripke, the example is necessary, and his intuitions here align with those of Putnam (1973), whom he cites as having independently arrived at overlapping ideas. A hypothetical substance in a counterfactual situation (the Kripke case), or an exotic substance in a remote part of the actual universe (the Putnam case), that had a different composition would not be or count as water or quicklime as the case may be, however much to outward appearances it behaved like water or quicklime. Going beyond anything explicitly said by Kripke, the general principle would seem to be:

(2) If a substance has a definite composition, then an actual or hypothetical substance with a different composition does not or would not count as *that* substance.

Other types of Kripkean examples concern the original material composition of artifacts (*this table was from the beginning made of wood*), the composition of atoms out of elementary particles (*gold is an element with atomic number 79*), what certain natural processes ultimately consist in (*heat is not a material caloric fluid but random molecular motion*). These can perhaps all be regarded as variations on the chemical formula examples. A yet further type concerns inclusions among biological taxa (*tigers are not reptiles but mammals*). These can perhaps be regarded as variations on the genealogical example by considering evolutionary descent, though Kripke discusses them more as if they were case of composition, as indeed they also are, since it is from the DNA in the organism's cells that its lineage is most accurately determined. None of these further examples can be examined in detail here.

Epistemology of modality

Kripke in the addenda to N&N quotes Kant's well-known remark that *experience can teach us that a thing is so, but not that it cannot be otherwise*. Kant correctly observes that while experience may teach us that a necessary truth is true, experience alone does not suffice to teach us that it is necessary. Kant goes wrong, according to Kripke, in inferring that experience is not required to teach us that it is necessary, thus effectively identifying known necessity with aprioriness. This was the first step in an historical process that led to early-to-mid-twentieth-century

philosophical schools such as logical positivism and ordinary-language philosophy taking necessity to reduce to a matter of semantic rules and linguistic conventions (Ayer 1936; Strawson 1952). Kripke's examples of aposteriori necessity, if accepted, as they widely have been, refute such reductionism. That means we must confront again the question of the epistemology of modality that puzzled Kant.

Kripke addresses the puzzle in a brief and guarded way in the addenda to N&N. There he suggests that in each of his basic examples, it is discoverable apriori by analysis that *if p, then necessarily p*, though it is only known aposteriori that *p*. (Note that he does not claim all examples of aposteriori necessity are like this; given independent basic examples that satisfy *if p, then necessarily p*, it is easy to construct simple logical compounds that do not, a construction left as an exercise to the reader here.) Behind the knowledge that *if p, then necessarily p* in the basic examples would presumably lie analytic, apriori principles that here, going beyond anything explicitly said by Kripke, have been tentatively represented by general principles (1) and (2) presented earlier.

Just as Columbus could not have discovered Atlantis, since it doesn't exist, so no Babylonian astronomer could have discovered that Hesperus and Phosphorus are distinct, since they aren't. Yet for many there remains a stubborn feeling that "it could have turned out that" Hesperus and Phosphorus were distinct. This, according to Kripke, is an illusion; but as David Papineau (2008) observes, it is one that, like some famous optical illusions, tends to persist even when it is recognized as illusory. Kripke offers an analysis of what is going on here, which tends to make the illusion a result partly of confusing epistemic with metaphysical possibility, and partly of confusing speaking *of* a counterfactual situation with speaking *in* a counterfactual situation.

What really could have been the case is this: the world could have been different from how it actually is, with creatures in it who in their epistemic situation were much like us down to some point in antiquity, who some evenings saw a heavenly body they called "Hesperus" and some mornings saw a heavenly body they called "Phosphorus," where these two bodies were not merely apparently but really distinct bodies, eventually confirmed to be such by scientific astronomy. Kripke insists, however, that what is described here is not a situation in which Hesperus and Phosphorus would have been distinct, as they actually are not, but a situation in which *Hesperus* and *Phosphorus* would have denoted something other than what they actually do denote: two distinct planets that would have existed but don't actually exist, rather than one and the same actual planet, Venus. The feeling that EW "could have turned out" to be the daughter of HT rather than GW, as the tabloid asserts, or that water "could have turned out" to have the composition HO rather than H_2O, as Dalton thought, would be explained away similarly.

Kripke's examples have convinced many of the reality and importance of the phenomenon of aposteriori necessity, but one should not take his achievement to license positing aposteriori necessities all over the place. One needs to be able to explain, in Kripke's way or some other, why there is an appearance of contingency; perhaps ultimately one needs to be able to explain as well the underlying source of the necessity in terms of appropriate analytic, apriori principles. Kripke famously argues that the illusion of contingency *cannot* be explained in his way in the case of such alleged aposteriori necessities as the identity, claimed by some materialists, between pain and C-fiber firing. What is different about this case is ultimately that while there is room for a gap between looking like two different planets and really being two different planets, or looking and feeling like water and really being water, there is no room for a gap between feeling like pain and really being pain, since pain, unlike planets or water, is a feeling. But this is not the place for any extended excursion into philosophy of mind.

Considerations of philosophy of mind have motivated the largest-scale attempted refutation of Kripke's views on necessity, represented by "two-dimensionalism" (Chalmers 1996). There can be no question of expounding this rival, anti-Kripkean view here, but it may be mentioned

that Kripke's defenders (Byrne and Pryor 2004) see two-dimensionalism as relying on a notion of "primary intension" that is too much like the old notion of uniquely identifying descriptive senses supposedly associated with names, and that is discredited by considerations of ignorance and error similar to those that arise in the *Weinberg* case.

Modal logic

As Kripke reminds us, John Stuart Mill held that while in general the signification of a word consists of a "denotation" and a "connotation," roughly what were later called *reference* and *sense*, in the case of a proper name there is no connotation, and the signification is simply the denotation. Mill famously also held that *all necessity is verbal necessity*, deriving from relations among connotations. It follows that while *philosophers are rational* can be necessary, because *rational* is arguably included in the connotation of *philosopher*, by contrast *Hegel is rational* cannot be necessary, because *Hegel* has no connotation for *rational* to be part of. He concluded that *there are no individual essences*.

Elaborating on this last doctrine, W.V. Quine held that "quantification into modal contexts," saying things like *there is something such that it necessarily Fs*, makes sense only if so-called *de re* modality makes sense, and a predicate can be said to be necessarily true of an object as it is in itself, independently of how or whether said object is named or described. If only *de dicto* modality, attributing necessity to whole sentences and not subsentential parts such as predicates, makes sense, then quantified modal logic is nonsense. And if necessity is identified with analyticity, as it was by Rudolf Carnap, a pioneer of quantified modal logic, and seems to have been by C. I. Lewis, the founder of modern modal logic as such, then only *de dicto* modality makes sense. As Quine's student Dagfinn Føllesdal puts it: "Quine saw that Carnap's and Lewis's linguistic conception of necessity was untenable if one wanted to quantify into modal contexts, and that their position was therefore incoherent" (Føllesdal 2004: xxi).

One of the most important consequences of Kripke's distinguishing necessity from analyticity the way he does is that it makes room for a rehabilitation of quantified modal logic, as a logic of metaphysical necessity rather than analyticity. Kripke's earlier mathematical work showed that a system of quantified modal logic need not be formally inconsistent, but only his later philosophical work in N&N offers hope of showing it is not intuitively incoherent, which was the real issue for Quine. Closely connected with the issue of the coherence of quantified modal logic was the debate between Quine and Ruth Barcan Marcus, the main pioneer of quantified modal logic, over the status of identity statements such as *Hesperus* = *Phosphorus*. The debate is discussed by Kripke especially in N&I, where he concludes by saying that Marcus was right that such identities are necessary, Quine was right that such identities are not analytic, and both were wrong in failing to distinguish necessity from analyticity.

It is noteworthy that when, in the winter of 1962, Føllesdal drew Kripke into the Marcus-Quine debate, he had not yet arrived at this position and did not insist on the distinction between necessity and analyticity any more than any of the other participants in the debate (Marcus et al. 1962). But indeed, Kripke in N&N dates its main ideas to a year or two later, the academic year 1963–1964. This was well after the period of his mathematical work, the bulk of which was completed by mid-August 1958, when an abstract (Kripke 1959b) announced that it was received by the editors of the *Journal of Symbolic Logic*. Thus, chronological considerations alone refute any suggestion that Kripke already had the metaphysical necessity of N&N in mind when doing his mathematical work.

That work among other things served to popularize a revival of Leibnizian talk of "possible worlds" in place of the blander terminology (such as "state descriptions") used by Kripke's

predecessors. Though he never gave up possible worlds talk entirely, Kripke by the time of N&N had come to suspect it may be potentially misleading. After all, it is one thing to say that there are states the world could have been in other than the state it actually is in, and quite another to say that what we are used to thinking of as the world is only one of many, and that for every state this world could have been in but isn't, there is some other world that is in that state. This last is a manner of speaking that leads straight to Ludovician polycosmology, the theory of a pluriverse of multiple worlds found in the work of Kripke's long-time colleague David Lewis, which for Kripke is sheer fantasy.

Model theory

As to the technical content of the mathematical work, Kripke does for modal logic something like what Alfred Tarski did for classical logic. In classical logic we have a formal syntactic *derivation procedure*—indeed, many different but equivalent ones—giving us a notion of a *provable* formula, or one that has a derivation, and of a *consistent* formula, one whose negation is not provable. We also have, thanks mainly to Tarski, a formal "semantics" or model theory, with a notion of truth in a model, giving us a notion of a *valid* formula, or one true in every model, and of a *satisfiable* formula, one whose negation is not valid, or equivalently, one true in some model. These are related by a *soundness theorem* asserting that every provable formula is valid, and a *completeness theorem*, asserting the converse, or what is equivalent, that every consistent formula is satisfiable. In modal logic there had long been many different systems, the better known among them bearing such names as T, S4, S5, and so on, each with a different class of provable formulas, and the problem was to devise a notion of model and a specification of different classes of models such that soundness and completeness would hold between T and T-models, S4 and S4-models, S5 and S5-models, and so on.

The solution to this problem presented in textbooks today incorporates the more elementary part of Kripke's mathematical work on modality. There is no need to reproduce here everything that can be found in a good textbook. Suffice it to say that a *Kripke model* for sentential model logic consists of a set X of indices, picturesquely or heuristically called "possible worlds," with one of them @ distinguished as the "actual world," and a relation R among them of "relative possibility," and a function V assigning to each world x and each sentence letter p, q, r a truth-value true (T) or false (F). The assignment is extended to compound formulas built up using negation $\sim$, conjunction $\&$, disjunction $\vee$, and the modal operators of necessity $\Box$ and possibility $\Diamond$, proceeding as in classical logic with $\sim$, $\&$, $\vee$, and taking $\Box A$ (respectively $\Diamond A$) to be true at x if and only if A is true at every (respectively some) y possible relative to x. Truth in the model as a whole means truth at @. The classes of T-models, S4-models, or S5-models are obtained by requiring R to be reflexive, reflexive and transitive, or reflexive and transitive and symmetric. The proofs of soundness are in each case tedious but routine, while the proofs of completeness are more work.

Since the papers in which Kripke presents such material cite quite a few precursors in footnotes, it is natural to ask how much of all this is distinctively his own and how much is due to his predecessors. This question can be answered only by going carefully through all the early papers, noting for each what is or isn't to be found in it. This massive task has been undertaken by Rob Goldblatt in a magisterial study, which after acknowledging that there were earlier contributions concludes that "it is only in Kripke's writings that we see such … ideas developed into an attractive model theory of sufficient power … to advance the field further" (Goldblatt 2005: 35).

Specifically due to Kripke are: (i) the notion of a model in which the elements of X are bare indices whose intrinsic nature is irrelevant, where all his predecessors had required the elements of X to be sets of formulas or something of the sort, (ii) the proofs of the more difficult completeness theorems and not just the easier soundness theorems, and (iii) the extension of the basic idea in three directions: from well-known systems like T, S4, and S5 to a range of more exotic systems; from modal logic to intuitionistic and other non-classical logics; and from sentential logic to quantification theory. He carries out this last extension in such a way as to make the so-called *Barcan formulas*, in effect asserting that everything that could possibly exist actually does exist, and that everything that actually does exist necessarily could not have failed to exist, into optional extras that one can assume or not as one pleases, whereas previous efforts had tended to build in these dubious assumptions. The importance of feature (i), emphasized by Goldblatt, will be apparent only to those who have actually worked in the field, but it is absolutely crucial; and it is simply an historical error or falsification to call models with this feature anything but *Kripke* models.

The keynote of Kripke's notion of model is its extreme flexibility, or adaptability to the widest range of modal systems. The other side of the coin is that the model-theoretic work is not tied to any particular conception of modality. If one has in mind some particular conception and wishes to show that this, that, or the other modal system is the one appropriate specifically for it, often a Kripke-style completeness theorem provides a starting point. But it must always be supplemented by an argument to the effect that the particular class of Kripke models involved is the appropriate one for the notion of modality in question. The latter task, which is not purely mathematical but largely philosophical, is generally the more difficult. For as Kripke has said in another context, *there is no mathematical substitute for philosophy*. For the cognoscenti, Robert Solovay's proof that the system known as GL is appropriate for Gödel's notion of necessity as formal provability is a case in point (see Boolos 1993).

In addition to the works of Kripke and others cited earlier, and the further readings suggested at the end of the list of references in this chapter, the following works of the present author may be cited, as enlarging on points it here has barely been feasible to allude to in passing, if that. My notes (Burgess 1999, 2003) explain the fallacies in certain objections to Kripke's model theory on grounds of its alleged inappropriateness for "logical necessity" in the sense of *truth by virtue of logical form*, and on grounds of its failing to validate $\Diamond\exists x\exists y(x \neq y)$ as a law of logical necessity. My article (Burgess 2014a) explains in some detail Kripke's views on the law of necessity of identity $\forall x\forall y(x = y \rightarrow \Box\, x = y)$ and the history of that law in Marcus and Quine. Obviously, many of the issues treated tersely in this chapter invite extended further examination. As heretofore unpublished materials (audiotapes awaiting transcription, transcripts awaiting editing, and more) in the archives of the Saul Kripke Center at the City University of New York Graduate Center gradually come forth in print in the years ahead, there will be even more to examine.

References

Ayer, A. J. (1936) *Language, Truth, and Logic*, London: Golancz.

Boolos, G. S. (1993) *The Logic of Provability*, Cambridge: Cambridge University Press.

Burgess, J. P. (1999) "Which Modal Logic Is the Right One?" *Notre Dame Journal of Formal Logic* 40: 81–93.

Burgess, J. P. (2003) "Which Modal Models Are the Right Ones (for Logical Necessity)?" *Theoria (San Sebastián)* 18: 145–158.

Burgess, J. P. (2013) *Saul Kripke: Puzzles and Mysteries*, London: Polity.

Burgess, J. P. (2014a) "Madagascar Revisited," *Analysis* 74: 195–201.

Byrne, A. and Pryor, J. (2004) "Bad Intensions," in Garcia-Carpintero, M. and Macià, J. (eds.) *Two-Dimensional Semantics*, Oxford: Oxford University Press, 38–54.
Chalmers, D. (1996) *The Conscious Mind*, Oxford: Oxford University Press.
Donnellan, K. (1970) "Proper Names and Identifying Descriptions," *Synthese* 21: 335–358.
Føllesdal, D. (2004) *Referential Opacity and Modal Logic*, London: Routledge.
Goldblatt, R. (2005) "Mathematical Modal Logic: A View of Its Evolution," in D. M. Gabbay and J. Woods (eds.) *Handbook of the History of Logic*, Amsterdam: Elsevier 6: 1–98.
Kneale, M. and Kneale, W. (1962) *The Development of Logic*, Oxford: Oxford University Press.
Kripke, S. (1959a) "A Completeness Theorem in Modal Logic," *Journal of Symbolic Logic* 24: 1–14.
Kripke, S. (1959b) "Semantical Analysis of Modal Logic" (abstract), *Journal of Symbolic Logic* 24: 323–324.
Kripke, S. (1963a) "Semantical Considerations on Modal Logic," *Acta Philosophical Fennica* 16: 83–94.
Kripke, S. (1963b) "Semantical Analysis of Modal Logic I. Normal Modal Propositional Calculi," *Zeitschrift für mathematische Logik und Grundlagen der Mathematik* 9: 67–96.
Kripke, S. (1971) "Identity and Necessity," in M. K. Munitz (ed.) *Identity and Individuation*, New York: New York University Press, 161–191.
Kripke, S. (1980) *Naming and Necessity*, Cambridge, MA: Harvard University Press.
Kripke, S. (2011) "The First Person," in S. Kripke (ed.) *Philosophical Troubles, Collected Papers Volume* 1, Oxford: Oxford University Press, 292–321.
Marcus, R. B., Quine, W.V., Kripke, S., and Others, (1962) "Discussion," *Synthese* 14: 132–143.
Palmer, F. R. (1986) *Mood and Modality*, Cambridge: Cambridge University Press.
Papineau, D. (2008) "Kripke's Proof That We Are All Intuitive Dualists," core.ac.uk/download/files/99/74068.pdf.
Putnam, H. (1973) "Meaning and Reference," *Journal of Philosophy* 19: 699–711.
Quine, W.V. (1960) *Word and Object*, Cambridge, MA: MIT Press.
Salmon, N. (1981) *Reference and Essence*, Princeton: Princeton University Press.
Soames, S. (2005) *Reference and Description: The Case Against Two-Dimensionalism*, Princeton: Princeton University Press.
Strawson, P. F. (1952) *Introduction to Logical Theory*, London: Methuen.

Further readings

Berger, A. (ed.) (2011) *Saul Kripke*, Cambridge: Cambridge University Press. (Studies based on the unpublished Kripke archives.) This work includes the following items:
Burgess, J. P. (2011) "Kripke Models," 119–140. (An attempt at a reconstruction of the route to Kripke models.)
Burgess, J. P. (2014b) "Madagascar Revisited," *Analysis* 74: 195–201. (An examination of the best-known example of how slippage can occur between one speaker and the next in an historical chain.)
Linsky, B. (2011) "Kripke on Proper and General Names," 17–48. (A fuller account of Kripke's views on naming.)
Shoemaker, S. (2011) "Kripke and Cartesianism," 327–342. (More on applications to philosophy of mind.)
Soames, S. (2011) "Kripke on Epistemic and Metaphysical Possibility: Two Routes to the Necessary A Posteriori," 78–99. (Exposition of Kripke's views with an expression of partial dissent.)
Stalnaker, R. (2011) "Possible Worlds Semantics: Philosophical Foundations," 100–118. (Exposition of Kripke's views with an expression of partial dissent.)

INDEX

a priori knowledge 5, 188–189, 198–207; defined 199–202; new developments 205–206; synthetic 204–205
a priori 199; necessary, relationship with 202–204; philosophical analysis 203; see also a priori knowledge; a priori–a posteriori distinction
a priori–a posteriori distinction 200
Abbott, B. 248
absolute fundamentality 110
abstractionists 43
accessibility 12, 41, 83, 87, 93, 95, 251
actualism: actualism/possibilism dispute 24–25; assuming 24; challenges to 22–24; domain inclusion 22, 25–26; letter of 27; nontriviality 25; proxy 26; refinements 21–22; relations 25–26; slogans 21–22; spirit of 27; truth of 25
actualism-possibilism debate 25
ad peiorem rule 349
Adams, Roberts 22, 73, 77
adiabatic accessibility 255–256
American plan 44
analytic, term 199
analyticity 132, 136, 138–139, 143, 201, 205, 390–395, 397–399, 405
Anaxagoras 332
ante rem universal 2
anti-haecceitism 71–72; arguments 73–76; counterpart theory 73
aposteriori necessities 173–174, 243, 401, 404
Aristotle 6, 125, 282, 284–286, 288–289, 332–346, 348–350, 397
Audi, P. 101
Australasian Plan 44, 47
Avicenna 345, 348, 350–351
Ayer, A 117
Ayer, A.J. 132, 137, 205

Ballarin, Roberta 390
Baumgarten, A.G. 379
Bealer, G. 210–211
Being Known 161
belief 15–16
Belot, G. 257, 259
Benacerraf, P. 117, 157–158, 165
Bennett, Karen 21
Berto, F. 41, 43–44, 129
Big Bang 51, 53, 258
bijective transformation 254
Bird, Alexander 239
Bjerring, J.C. 46–47
Blackburn, S. 104, 130, 148–149, 159–160, 164
Bose-Einstein statistics 223, 227
Brading, K. 223
Brandom, Robert 149
Brighouse, C. 257
Bueno, Otávio 1, 319
Burgess, John P. 7, 131, 400
Butterfield, J. 253

Caesar, Julius 1, 85
Cameron, Ross P. 105, 136
Cantor's paradise 286
careless consideration voicing 216
Carey, Susan 184
Carnap, R. 11, 117, 132, 169
Casullo, Albert 198
categorically inappropriate 209
causal determination 369
causation 11, 30, 214–215, 227, 268, 273, 364–346, 369, 371–373
centered world 15–16
Chalmers, D. 186
chance 75–76
Chihara, Charles 283–284

Chisholm's Paradox 88–89
Chudnoff, E. 211–212
classical logic 6, 44, 285–286, 321–323, 326, 406
classical particle mechanics (CPM) 269
classical theory 82; metaphysical challenges 86
Coextensiveness thesis 379–381
coherent modal imagination 186
Colyvan, Mark 282
common epistemic kind 213
Conceivability Principle 364, 367–368
conceivability: crucialness of 169–170; epistemic idleness of 167–179; no apt account 171–177; theories 116–117
conceptualism 51–53, 70, 73
Conditional Proof 34
conditionals: counterfactual conditions 30–39; laws 230–238; impossible worlds 17–18
conservation law 234–237
consistency 13–14, 169; see also logical consistency
Constitutive Requirement 180–182, 186
constructibility quantifier 283
Context–Index pragmatics 83, 94–95
Context–Index representation 83
contiguity 371
contingency 1–2, 120, 127, 137, 167, 189, 194, 271, 282, 293, 303, 349–350, 356–357
contingent modal profiles 78
contingentism: modal contingentism 4, 76–79; modal-operator contingentism 76–79; necessitism/contingentism dispute 24–25; necessitism-contingentism debate 24–26; relations 25–26; world-existence contingentism 76–78
convenient fiction 43
Copeland, J. 3, 42
Copy Principle 364, 368
Correia, Fabrice 100–101, 106
cotenability 32–33, 194
cotenable 32
Coulomb's law 233–4
counterfactual antecedents 30–32, 234–235
counterfactual approach 192, 194, 227
counterfactual conditionals: early theories 31–33; Lewis versus Stalnaker 36–38; overview 30–31; possible-worlds semantics 33–34; similarity between worlds 34–36
counterfactuals: conditionals 30–39; counterfactual antecedents 30–32, 234–235; defined 30; knowledge of 193–196; logic of 33, 74, 190–191, 195; theory of 41
counterpart function 15, 89, 93
counterpart theory 12, 14–15, 73
counterpossible 17–18, 41, 45, 55, 190
Craig, E. 159, 164
Cresswell, M.J. 23, 42
Cronus, Diodorus 331, 333
Crude Conventionalism 138–139, 141
cumulative hierarchy 288, 296, 297, 298, 299, 302
Curley, E. 358

de dicto ambiguity 66
de dicto essences 66
de dicto modality 62, 66, 199, 405
de dicto possibility 23
de dicto reading 66, 336, 338–339, 342
de re essence 66
de re modal knowledge 199, 206–207
de re modal speech 398
de re modality: conceptualist skepticism 70–71; de re modal knowledge 199, 206–207
de re possibility 23
de re reading 339
de se belief 11, 15–16
de se content 16
De, Michael 11
decreeing 142
Dedekind, R. 287
Dedekind-Peano axiom 283
Democritus 332
Descartes, R. 6, 167, 186, 208, 215; modality 355–363
descriptive 4, 23, 70–71, 73, 126, 148–149, 151, 315, 403, 405
determinism 55
Dialectical School 331, 333
disjunctive account of intuition (DI) 215, 217
disjunctive conception 213–214
Dispositional Identity Thesis 225–226
dispositionalism 221, 224–227
Divers, John 125
domain inclusion actualism 21, 26
double-indexing semantics 82, 92, 94, 97n19
Dummett, Michael 288–289, 314–315
Duží, M. 47
dynamics 222

Earman, J. 255
Edgington, Dorothy 30
effective cognitive organization 276
Eiffel Tower 61–64, 206, 356
elementary sets 294
El-Rouayheb, K. 350
epistemic challenge 162, 165
epistemology: intuition 211–213; modality 208–211
e-representations 148
essentialism, Kripke on 242–243
essence, modal conception of 62–65
essential properties 61–64, 66, 88, 140, 206, 209, 240, 243, 248, 320
essentialism 61–69; Chisholm's Paradox 88–89; essentialists 248–249; hyper-essentialism 62; kinds 66; and modality 61–69; natural kind essentialism, objection to 247; a posteriori truth 65; a priori truth 65; psychological essentialism 249

eternal truths: necessity of 358–359; ontological status 357–358
Etiological Question 162–163
explanation, "unificationist" face of 275–276
expressivist 133, 148

factive *vs* non-factive 101–102
factivity 101, 103
factual vs generic 101–102
Fermi-Dirac statistics 223, 227
fictionalism 43, 128
Fine, Kit 35, 62, 64, 101, 105, 108, 118–119, 128, 206, 243
Fiocco, M. Oreste 167
First Degree Entailment 44
first-order law 235
first-order modal language 310
first-order modal logic 301
first-person-phenomenologically-indistinguishable 213
Fletcher, Samuel C. 251
FLOTEs 269, 272–274
Forbes, G. 118–119, 128
Foundation 296
fragments 83
Frege, G. 152, 204
Fregean 111, 245, 247, 357
French, Steven 221
fundamentality 100, 107, 109–111

Gates, Bill 231–232
Geirsson, Heimir 175
general modal status 201, 203–205
general relativity 227, 251, 255–259
general theory of explanatory structure 268, 270
generic grounding 101–102, 111
genuine modal realism 115, 118
George III 230
German rationalism 379
God, infinity of 360
gold-cube generalization 233
Goldilocks 120
Goodman, N. 31–33, 35, 38
Grounding thesis 380, 383, 385
grounding 100–102; arity 101; fundamentality 109–111; general claims, formulating 102; generic grounding 101–102, 111; grounding problem 147; mere partial ground 103; modality, connections with 103–105; ontological dependence 109–111; patrial grounding 108–110; weak grounding 104
ground-theoretically equivalent 101–102

H20 5–6, 55, 138–140, 142–143, 152, 167, 172–174, 240, 243–249
habit 366
Haecceitism 71–72; arguments for/against 73–76
Hale, Bob 308
Hall, Ned 265
Hamiltonian framework 237
Hamilton-Jacobi theory 253
Hellie, Benj 82
Hellman, G. 283–284, 288
Hesperus 131, 173–174, 186, 202–203, 266, 401, 403–405
Hintikka, Jaakko 11, 42, 344–345
Hodges, W. 351
Horwich, Paul 148, 152
Hughes, George 23, 348
Hume, David 125–126, 147, 167–168, 174; on modality 364–377
Humean regularities 221
hyper-essentialism 62

ideal conceivability 117, 172–173
ideal rational reflection 172–173
identification question 211
ideological modal realism (IMR) 129–130, 132
if-then statements 230
immoderate essentialists 88
impossibilia 55, 121, 128
impossibility 46, 51–52, 87–88, 91, 94, 120, 133, 167, 171, 186, 255, 273, 326, 331, 335, 337–338, 342, 368–369
impossible possible worlds 42
impossible worlds 17–18, 40–41; defenders 47; nature of 42–45; objections to 45–47; uses 41–42
incompletability 286–287
indicative form 149, 151
indicative mood 149–150, 152
indicative 17, 30
individual essence 26, 64, 83, 86, 88, 114, 185, 405
inferability 367
informative 222
informativeness 222–223
initial conditions 222–223
initial segments 302
integration challenge: integration requirement 158, 160–161, 163–165; intra-domain components 161–165; mathematics, precursor to 157–158; modality 158–160; overview 157; ways to meet 160–161
intentionality 214–215
inter-domain question 160
intra-domain components (integration challenge): "integration," strength of 163–164; integration and disagreement 164–165; plausible pair 161–163
intuition: disjunctivism 213–215; epistemology of 211–213; faculty of 211, 215; learnability 215–217; Master Skeptical Argument Against Justification for Modal Beliefs Based on Intuition 212; ordinary 215–217; reliability 215–217

Jago, M. 41, 43–44, 46
Johnson, Fred 342
Johnston, S. 348, 351
judgements 30, 35, 132–133

Kant, Immanuel 198–199, 201, 203–204, 206, 241; on real possibility 378–389
Kaplan, D. 96
Kiourti, I. 43
Kitcher, P. 199–200, 202, 277
Kment, Boris 70
knowledge of necessity 4–5, 138, 167–168, 177, 181, 195, 206
Knuuttila, S. 345–348
Koslicki, K. 63
Kripke semantics 23
Kripke, Saul 23, 33, 42, 127–128, 130–131, 151, 199, 200, 202–204
Kukkonen, T. 345
Kung, Peter 175

Lange, Marc 230
Laplacian determinism 258
Law of Conditional Excluded Middle (CEM) 37
laws of nature 5, 11, 31, 35, 55, 83, 86–88, 230–238
Leftow, Brian 49
Leibniz, G.W. 379
Leuenberger, Stephan 103
Lewis, David 14, 31, 37, 41, 72, 89, 112, 125, 127–128, 132, 148, 158, 180
Lewisian framework 120
Lieb, E.H. 255–256
Lieb-Yngvason axiomatic formulation 251
linguistic conventionalism 51, 117, 136–140, 164, 404
linguistic ersatzism 22
Linnebo, Øystein 281
Linsky, Bernard 21, 25–27
logical consistency 380, 385
logical disagreement 320, 323–324, 326–327
logical omniscience 13
logical pluralism by cases 322
logical revision 89, 91, 323
logical revisionism 88, 92–93
logical space 324–325; logical pluralism 326–327; trivialism 325–326
logical space, origins 49–57; best of both worlds 52–53; case for 50–51; current "origins" views 51–52; God 54–56; natural powers 53–54; non-natural powers 53–54
logical truths 118, 161–162, 194, 205, 235, 391
Lowe, E.J. 63–65
Lycan, W. 31, 147

Mackie, Penelope 61
Malink, Marko 342
Manchak, J.B. 257
Marcus, Ruth Barcan 130, 351
mathematical knowledge 157–159, 164, 276
mathematics: modal accounts of 283–284; modal status of 281–282; potential infinity 284–286; potentials versus actualism 286–289
maximality 13
Maxwell's demon 255
McMichael problem 23
meaning, defined 243
Meinongian 49–51
Melia, Joseph 119
Menzel, Christopher 292
metalinguistic negotiation 153
metalinguistic uses 151
metaphysical accessibility 87
metaphysical challenge 82, 86, 162–163
metaphysical modal statements 148–149
metaphysical modalism 320
metaphysical modality, essence 65; knowledge of 189–196; logic of 308–318; relativized 82–99
metaphysical necessity, heterogeneous character of 273–274
metaphysical possibility 83, 94, 127, 189–190, 195, 265, 268, 270, 368, 379, 384, 404
Meyer, R.K. 18
might-conditionals 36
"might," modal expression 11
Millican, Peter 364
modal anti-realism: basics 125–126; ideological modal anti-realism proper 131–133; ideological modal realism 129–131; methodological issues 126–127; ontological modal anti-realism 127–129
modal commitment, general issues 252–253
modal contingentism 4, 76–79
modal conventionalism 136–145
modal epistemology, counterfactual-based approach 188–197
modal expressions 11, 115, 119, 129, 131, 152
Modal Extension Principle 180–181, 184
modal force 104, 140, 225, 323, 326–327
modal fragments 83–86
modal knowledge 148, 152, 167–173, 175–177; see also de re modal knowledge
modal logic (ancient Greek): Aristotle 334–335; "master" argument 332–333; modal conversion rules 335–336; modal proofs 337–340; modal syllogistic 335; modal syllogistic, models of 342; overview 331–332; possibility premises 336–337; possible premises, syllogisms with 340–342; standard deductive systems 309–310; world semantics 310–311
modal metaphysics 116–117, 121, 131, 159, 247, 260, 292–293

modal realism: genuine 115, 118; heavyweight 146–147; ideological 125, 129–130, 132–133
modal reliabilism (MR) 210, 217
modal set theory 292–307; completion problem 298–303
modal set theory, traditional modal metaphysics 292–293
modal statements 14, 131, 146–149, 151, 285
modal structuralism 283
modal vocabulary 117, 120–121, 149–150
modalism 118–122; conceivability theories 116–117; linguistic theories 117–118; overview 114; possible worlds 118; reductive motivations 115–116
modalist logical pluralism 322–323, 327
modality (Descartes): Cartesian metaphysics 355–356; contingency existence 356–357; Creation Doctrine 360–362
modality (Hume): apparent modal subjectivism 370–371; causal necessity, "impression" of 365–367; conceptual modality 367–370; empiricism 364–365; securing causal objectivity 371–373
modality (Kripke): aposteriori necessity 400–401; composition 402–403; genealogy 402–403; identity 401–402; modal logic 405–406; modal theory 406–407; modality, epistemology of 403–405; naming 401–402
modality (medieval philosophy) 344–354
modality (Quine): analyticity 394–398; essentialism 394–398; historical background 391–392; necessity 398–399; quantified modal logic 392–394
modality, logical pluralism 320–321; logical space 323–324; logics, plurality of 321–323
modality: dispositionalism 224–226; epistemology of 159, 208–211, 403–405; logic of 348–351; mathematics 281–290; metaphysics 116, 130–131, 159, 164, 182, 228, 319, 344–348; mosaic 222–224; natural kind terms, semantics of 245–247; natural kinds 239–249; overview 221–222; principle-based account 180–187; scientific structuralism 221–228; structuralism 226–228
moderate centering 88–89
moderate modal naturalism 83, 94
modus ponens 34, 116, 312
modus tollens 34, 121, 191
Montague, R. 42
Muller, F.A. 256
Müller, T. 259
multiplicity 6, 287, 289, 332, 372
Murray, Adam Russell 82
"must," modal expression 11

Naming and Necessity 71, 138, 188–189, 192
natural deduction 116, 194, 315
natural kind question 211
natural kinds 239–250
natural necessity,230–238
necessary: a priori, relationship with 202–204; term 143
necessitarianism 87, 94, 103–104
necessitism 22, 24–26
necessity: absolute 311–312, 367, 368; causal 364–365, 367–368, 370–373; conceptual 231, 370–371; logical 139, 221, 239, 281, 283, 308, 311, 391, 393, 407; mathematical 370–371; metaphysical 5–6, 65, 84, 130–132, 139, 146, 151, 159, 180–181, 184, 190, 196, 204, 267, 270, 273–274, 277, 285, 288, 303–304, 309, 311–314; modal 371; moral 371; relative 308, 311
Needham, Tristam 248, 276
Nelson, Alan 355
Neo-Conventionalism 141–144
"no origins" realism 50–51
Nolan, Daniel 40
nomic claim 231, 235–236
nomic stability 235–237
non-classical world 42
non-cognitivist 133
non-concrete objects 49
non-descriptivist 133
non-metaphysical modalism 320
non-normal worlds 42–43
nonproxy actualism 24
non-trivial counterfactuals 44
non-trivial essential properties 63
normative function 146, 149, 151, 153
norms 146–154
Nortmann, Ulrich 342
Norton, J.D. 255

objective modalities 188–190
object-language claims 150–151, 246
Ockham, William 350
Oderberg, D. 63
Okhamist tendencies 115
ontic structural realism 226
ontological dependence 105–107, 109–111; fundamentality 109–111; generic dependence 107; modality, connections with 108; one-sided 110; partial grounding 108–109
ontological modal realism (OMR) 125
ontology 50–51, 53–54, 82–83, 86, 90–92, 256–257
open worlds 45
ordinary knowledge 189, 192–193
"origin" theory 51
Osler, M. 358

Pairing 294
Pap, A. 117

paraphrase 43, 120–121
Parsons, C. 299, 303
Parsons's Principle 299, 302–303
Patterson, McCall 342
Pauli Exclusion Principle 223, 271
Peacocke, Christopher 180
Peano Arithmetic 13
Permutation Symmetry 223–224, 227
Phosphorus 131, 173–174, 186, 202–203, 266, 401, 403–405
physical modality 265–278
physical necessity 371
physics 222–228, 234, 251–255, 257
Placek, T. 259
placement problem 147
Plantinga, A. 22–23, 26, 78, 128–129, 131, 204
Platonist theories 50
plentitude 14
Plunkett, David 153
Poincaré group 223–224
Porphyry 344
possibilism 3, 21, 24–27, 127
possibility, language of 118
possible possibilities 89, 94
possible worlds 11–12; approaches 41; belief/centered worlds 15–16; conditionals 17–8; semantics 12–14; talk 11, 406; types 49
postclassical semantics: accessibility 87–88; counterparts of 89–90; world-relative domains 91–92
potential infinity 284–287
potentialist translation 285, 299–303
power theories 52, 54
Price, Huw 147–148
Pride and Prejudice 137
Priest, G. 43–45
primary conceivability 173
Principle of Plenitude 184
Pringle, John 230
Prior, A.N. 42, 49, 130
priority questions 206
priority 371
projection 236–237
property 16
propositional modal fragment 83–84
proxy actualism 26–27
pseudo-laws 225
pure model theory 42
pure sets 288, 295, 303–304
Putnam, H. 202, 243–244, 299, 403
Pythagorean theorem 201

quantificational modal fragment 85–86
quantum theory, interpretations of: many-worlds interpretations 260; modal interpretations 260–261; overview 259–260
quasi-realist 133
Queen Victoria 61–64
Quine, W.V. 61, 71, 117–118, 169, 202, 321, 390
Quinean modal scepticism 131
Quine-Putnam indispensability argument 282

Read, Stephen 344
real possibility (Kant): beyond real/formal possibility 384–385; concept and intuition 381–382; logical vs. real possibility 379–381; overview 378–379
reductio ad absurdum 286
referential modal fragment 85
regularity theorist 366
relativized metaphysical modality 4, 82–83, 86, 92–96
reliability question 211
repeated instances 366
representational necessity 396
Rescher, Nicholas 342
Restall, G. 43
rigidity 103, 401–402
Roca-Royes, Sonia 157
Rodriguez-Pereyra, G. 101
Routley, R. 18
Roy, Tony 146
Russell, Gillian 141
Russell's Paradox 293–294, 297–298
Russellian proposition 180
Ryle, Gilbert 149–150

S4, argument against 313–314
S5, arguments against 313; arguments for 311–313
Salmon, Nathan 42, 78, 89, 147, 245, 313
salve veritate 201
SCARF: billiard balls 272–273; causation 273; collapsing stars 271–272; cosmic SCARF 268–269; different masses 273; particle quintumvirates 270–271; particles in a box 271; samples 269–270
Schaffer, Jonathan 274
Schechter, Joshua 161–163
Schlick, Moritz 149
Schnieder, B. 106, 111
Schwarz, W. 46–47
scientific structuralism 221–229
Scott-Montague semantics 18
secondary conceivability 173
Sellars, Wilfrid 149
sentence, validness of 310
sentences, meaning of 141
Separation 295
Shalkowski, Scott A. 1, 114
Shapiro, Stewart 281
Sidelle, A. 118, 132, 139–140, 152
Sider, Ted 115, 143
simple quantificational analysis 12–13, 15

simplest quantified modal logic 23
singleton set 243, 282, 293
Skiles, A. 63, 103–104
Sklar, L. 257
Smith, Robin 331
Sobel sequences 17
Socrates 63–64, 66, 71, 73, 101–102, 106–108, 110, 118, 131, 206, 243, 320
Sophisticated Traditional Conventionalism 139, 141
source questions 206
space-displacement invariance 236
spacetime points 269
spacetime theory: (in)determinism 258–259; ontology 256–257; overview 256; physically reasonable spacetimes 257–258
Special Theory of Relativity 223
specific modal status 201, 203
stability: linked to lawhood 232–233; natural necessity 234–235; sub-nomic 231–236
Stairs, Allen 147
Stalnaker, Robert 22, 26, 31, 36–37, 128
Stang, Nicholas 278
Strawson, P. 30
Stroud, Barry 366
structural properties 241, 301
subjunctive conditionals 31, 193
sub-nomic stability 231–236
sub-nomically stable set 233–237
subpluralities 302
Sundell, Tim 153
superset 233, 288
symmetries 223–228, 254
synthetic necessities, problems 138–141

theoretical identification 239–240, 247
theory-construction, principle of 53
thermodynamics: (ir)reversibility 254–255; adiabatic accessibility 255–256; overview 254
Thom, P. 350
Thomasson, Amie L. 146
Tichy, P. 36
time-displacement invariance 236
tracking 24, 142–143, 148
Tractatus 49, 149
transfinite cumulative hierarchy 297, 298
triggering 55
truthmakers 146–148, 150–151
truth-tracking intuitions 210
truth-values 41, 44, 52, 231, 283, 310, 324
Twin Earth, thought experiment 243–247

Uffink, J. 254, 256
uninformative 222
unreliability 216
urelemente 288

Vacek, M. 43
Vaidya, Anand Jayprakash 208
van Brakel, J. 248
van Fraassen, B. 33
van Inwagen, Peter 159, 174–175
variational principles 251, 253–254
Vetter, Barbara 147, 228
Vihvelin, Kadri 55–56

Williamson, Timothy 24, 45, 131, 200, 209
Wilson, Jessica M. 82
Wittgenstein, L. 49, 71, 133, 149
Wolff, C. 379
world, defined 14
worldly *vs* representational 101–102
woulds 37–38
Wright, C. 133, 161–163
Wyatt, N. 346–348

Yablo, Stephen 140, 143, 168, 174, 185–186
Yagisawa, T. 43
Yngvason, J. 251, 255–256

Zalta, Edward 25–26
Zermelo numbers 294, 296–297, 302–303
ZF 293–298