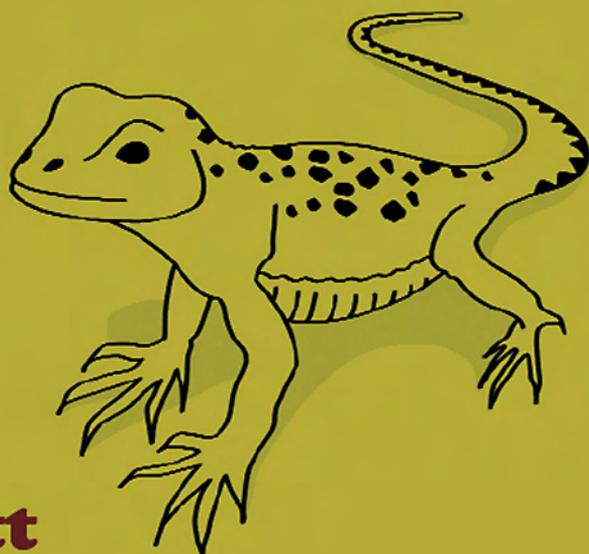


**What is this thing called
metaphysics?**



Brian Garrett

WHAT IS THIS THING CALLED METAPHYSICS?

'The book is a clear, no-nonsense introduction to metaphysics. It covers many of the most important contemporary issues and arguments. It is sophisticated without overwhelming readers with formalisms.'

David Robb, Davidson College, USA

Why is there something rather than nothing?
Does god exist? Who am I?

Metaphysics is concerned with ourselves and reality, and with the most fundamental questions regarding existence. This clear and accessible introduction covers the central topics in metaphysics in a concise but comprehensive way.

Brian Garrett discusses the crucial concepts in a highly readable manner, easing the reader in with a look at some important philosophical problems. He addresses key areas of metaphysics:

- existence
- causation
- God
- time
- universals
- personal identity
- truth.

What Is This Thing Called Metaphysics? contains many helpful student-friendly features. Each chapter concludes with a useful summary of the main ideas discussed, study questions, annotated further reading and a guide to Internet resources. Text boxes provide bite-sized summaries of key concepts and major philosophers, and clear and interesting examples are used throughout, whilst a helpful glossary explains important terms. This is an ideal textbook on metaphysics for undergraduates taking a first course in philosophy.

Brian Garrett is Senior Lecturer in Philosophy at the Australian National University, Australia. He is the author of *Personal Identity and Self-consciousness* also published by Routledge.

WHAT IS THIS THING CALLED METAPHYSICS?

BRIAN GARRETT

First published 2006
by Routledge
2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

Simultaneously published in the USA and Canada
by Routledge
270 Madison Avenue, New York, NY 10016

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

© 2006 Brian Garrett

This edition published in the Taylor & Francis e-Library, 2007.

“To purchase your own copy of this or any of Taylor & Francis or Routledge’s collection of thousands of eBooks please go to www.eBookstore.tandf.co.uk.”

All rights reserved. No part of this book may be reprinted or reproduced or utilized 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.

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

Garrett, Brian, 1961–

What is this thing called metaphysics?/Brian Garrett.

p. cm.

Includes bibliographical references and index.

1. Metaphysics. I. Title.

BD111.G37 2006

110–dc22

2006015183

ISBN 0–203–96844–1 Master e-book ISBN

ISBN10: 0–415–39391–4 (hbk)

ISBN10: 0–415–39392–2 (pbk)

ISBN10: 0–203–96844–1 (ebk)

ISBN13: 978–0–415–39391–1 (hbk)

ISBN13: 978–0–415–39392–8 (pbk)

ISBN13: 978–0–203–96844–4 (ebk)

'the object of a liberal education is to produce the perfected mind . . . while a vocational education is a training for work, a liberal education (if it is to be reckoned a training at all) is a training for leisure; it teaches, as Pattison put it, the art to live: it instructs a man how to live and move in the world and look upon it as befits a civilised being.'

John Sparrow, *Mark Pattison and the Idea of a University*
(Cambridge: Cambridge University Press) 1966

'Omit needless words'

W. Strunk and E. B. White, *The Elements of Style*
(Toronto: Macmillan) 1970

CONTENTS

Preface	xi
Introduction	xiii
1 God	1
<i>Introduction</i>	2
<i>The ontological argument</i>	4
<i>The cosmological argument</i>	7
<i>The teleological argument</i>	11
<i>The argument from evil</i>	13
<i>Concluding remarks</i>	16
2 Existence	19
<i>Introduction</i>	21
<i>The great sweep of being</i>	22
<i>Why does our world exist?</i>	23
<i>Modal realism assessed</i>	25
<i>Non-existent objects</i>	27
<i>What is existence?</i>	31
<i>Concluding remarks</i>	33
3 Universals and particulars	35
<i>Introduction</i>	37
<i>Property realism</i>	38
<i>Objections to traditional property realism</i>	41
<i>Russell's bundle theory</i>	42
<i>Varieties of nominalism</i>	43
<i>Concluding remarks</i>	48

CONTENTS

4 Causation	51
<i>Introduction</i>	53
<i>Preliminary questions</i>	54
<i>The constitutive question: what is causation?</i>	57
<i>Hume</i>	58
<i>The regularity theory</i>	61
<i>The counterfactual theory</i>	64
<i>A simpler theory</i>	65
<i>Concluding remarks</i>	66
5 Time: the fundamental issue	69
<i>Introduction</i>	71
<i>Setting up the debate: A series and B series</i>	72
<i>The A series is fundamental to time</i>	73
<i>Russell's reply</i>	75
<i>McTaggart's paradox</i>	76
<i>Diagnosis</i>	78
<i>A theory or B theory?</i>	81
<i>Concluding remarks</i>	82
6 Time: three puzzles	85
<i>Introduction</i>	87
<i>Prior's puzzle</i>	88
<i>Time without change</i>	90
<i>Time travel</i>	94
<i>Concluding remarks</i>	99
7 Free will	101
<i>Introduction</i>	103
<i>Fatalism</i>	104
<i>A more sophisticated fatalist argument</i>	107
<i>Theological fatalism</i>	110
<i>Free will and determinism</i>	111
<i>The case for incompatibilism</i>	112
<i>A compatibilist reply</i>	114
<i>Is the notion of free will incoherent?</i>	115
<i>Concluding remarks</i>	116
8 Personal identity	119
<i>Introduction</i>	121
<i>Theories of personal identity</i>	122
<i>What is a person?</i>	123

<i>What is it for a person to persist?</i>	125
<i>The complex view defended</i>	128
<i>Does identity matter?</i>	131
<i>The simple view</i>	132
<i>Concluding remarks</i>	134
9 Realism and anti-realism	137
<i>Introduction</i>	138
<i>Delineating the realist/anti-realist dispute</i>	140
<i>Case study I: character</i>	142
<i>Case study II: the past</i>	144
<i>Anti-realism: motivation and assessment</i>	146
<i>Anti-realism disproved?</i>	148
<i>Concluding remarks</i>	150
Glossary	153
Notes	163
Index	173

P

PREFACE

I began writing this book during a sabbatical from the Australian National University (ANU) in the first half of 2005. I am grateful to ANU for the leave from teaching. In subsequent seminars at ANU, working through draft chapters, I received useful feedback from my colleagues. Thanks especially to Jeremy Shearmur and Peter Roeper, but also to Havi Carel and Udo Thiel. Thanks also to those graduate students who attended the seminars: David Wall, Peter Grundy, Peter Eldridge-Smith, Luc Small, Paul Miller and Matt Cox.

I'm grateful to Routledge's Philosophy Editor, Tony Bruce, for his encouragement; to Priyanka Pathak, the Development Editor, for her advice and comments; and to Tim Crane for putting me in touch with Tony. I made a number of revisions as a result of critical feedback from three Routledge referees. In addition, my former student Robert Nichols wielded his editorial pen with characteristic incisiveness; John Gregory cast a civil servant's eye over the entire text; and Thomas Mautner allowed me to draw on his extensive knowledge of the history of philosophy.

In an attempt to make the book as user-friendly as possible, Routledge suggested the format whereby each chapter is replete with concept boxes, philosopher profiles, key quotes, glossed words, study questions, further reading, and Internet resources. Hopefully, the reader will find these useful. But if some key words and phrases are still unclear, I warmly recommend Thomas Mautner's *Dictionary of Philosophy* (second edition, London: Penguin, 2005) as an indispensable aid to further understanding.

I INTRODUCTION

When Andronicus of Rhodes edited Aristotle's works in the first century BC, he placed the books dealing with 'First Philosophy' after those on physics, and so they were called *metaphysics*. Aristotle characterized his 'First Philosophy' as the study of being *qua* being. This characterization, though cryptic, does capture one aspect of metaphysics: its craving for generality. The metaphysician is concerned to investigate the most general and ubiquitous features of reality (e.g., existence, space, time, causation, object, property). But metaphysical inquiry is not always utterly general. It also includes investigation of more specific topics, such as free will and personal identity.

In conducting their inquiries, metaphysicians draw on an array of data and techniques: rigorous thinking, conceptual analysis, empirical inquiry and the formal methods of logic and mathematics. Empirical results and formal techniques are more relevant to some metaphysical topics than to others. For example, empirical results are relevant to the debate over the nature of space and time; and formal techniques are relevant to the debate between realism and anti-realism. Nonetheless, questions in metaphysics are rarely solved by empirical results alone or by the purely formal techniques of logic and mathematics. We should see metaphysical issues as forming a spectrum with some issues at the more abstract or *a priori* end of the spectrum and others at the more science-relevant end of the spectrum.

The best way to appreciate the character of metaphysics is to engage with particular metaphysical disputes. In this book I have tried to introduce the central topics of metaphysics in a clear and accessible way, unencumbered by the details that tend to clog up philosophy journals. Often I will take a stance on some issue. But, where I do, I always give my reasons, and the reader is free to disagree. In the remainder of this brief introduction I will outline the main topics of each chapter.

In Chapter 1 we examine the three traditional arguments for God's existence – the ontological, cosmological, and teleological arguments, respectively –

WHAT IS THIS THING CALLED METAPHYSICS?

and one well-known argument against God's existence – the argument from evil. The discussion introduces and makes use of modal notions (notions of possibility and necessity), in particular the notion of a necessary being.

In Chapter 2 we look at a range of issues concerning existence. J. L. Austin once quipped that existence was 'like breathing only quieter'. It's hard to follow that remark, but in this chapter we look at the doctrine of modal realism (the view that other possible worlds and possible entities exist in just the way that our world and its inhabitants do); examine some answers to the old chestnut 'Why is there something rather than nothing?'; consider the view that there are non-existent objects; and finally address the question of whether existence is a property of ordinary objects (such as breathing) or not. Our discussion draws on some elementary logic and reveals various connections between these four issues.

Chapter 3 is concerned with standard, and ancient, metaphysical fare: the problem of universals. In order to account for the nature of properties, and the (alleged) differences between properties and particulars, do we need to conceive of properties as universals (identical in their instances)? Realists answer 'yes' to this question; nominalists answer 'no'. We consider some objections to standard varieties of realism and nominalism. We also look at a relatively recent version of nominalism – trope theory – which subverts the traditional paradigm and regards properties as a kind of particular.

In Chapter 4 we turn our attention to causation. After addressing various preliminary questions – e.g., about the kinds of entities that can be causes and effects, the temporal direction of causation, the logical properties of the causal relation – we ask the fundamental metaphysical question: what is the causal relation? What is the connection between two events such that one causes the other? We examine two reductive answers to this question: Hume's answer that causation is a matter of constant conjunction and Lewis's answer that causation is a matter of counterfactual dependence. We also look at a non-reductive answer: the view that causation is an irreducible species of natural necessity.

Chapter 5 is about time. As J. M. E. McTaggart pointed out early last century, positions in time can be distinguished in two quite different ways. Events can be placed in the B series (and ordered by the earlier-than and later-than relations). They can also be located in the A series (and ordered as past, present or future). Positions in the B series are permanent; those in the A series are constantly changing. The A theorist regards A series determinations as fundamental to time; the B theorist regards B series determinations as fundamental. McTaggart thought he had decisive objections to both theories, and happily concluded that time is unreal. Few have followed McTaggart's heroic course, but there is still a lively, and often scientifically informed, debate between A and B theorists.

In Chapter 6 we look at some subsidiary issues about time. Does the fact that we have temporally biased attitudes (e.g., thanking goodness when a painful



experience is over rather than yet to come) provide an argument in favour of the A theory of time? Is it a necessary truth, as many philosophers have held, that there can be no time without change? Finally, is it possible to travel into the past?

Chapter 7 is concerned with various challenges to the common sense belief that we have free will. The fatalist argues, on purely logical grounds, that freedom is an illusion: we cannot do other than we actually do. Hard determinists, as they are called, argue that free will and determinism are incompatible, and that since determinism is true, we lack free will. The arguments of the fatalist and the hard determinist are subjected to scrutiny. Others hold that free will requires a kind of self-determination which is logically unsatisfiable, and hence that the notion of free will is incoherent.

In Chapter 8 we address the questions of what is distinctive about persons and what it is to be the same person over time. The methodology employed relies heavily on imaginary thought experiments and our intuitive judgements about who is who in such thought experiments. This method has its limits but, within those limits, it can help to decide between rival theories of personal identity. As with causation, theories of personal identity can be divided into reductive and non-reductive varieties. All the standard reductive theories are subjected to detailed criticism, and a non-reductive theory – the simple view – is put forward for consideration.

The final chapter is concerned with a modern variant of an old debate. The realist versus idealist dispute is one of the oldest in metaphysics. In the 1960s the Oxford philosopher Michael Dummett proposed a new way of conceiving this debate. Drawing on various ideas of the later Wittgenstein concerning the public nature of meaning, Dummett suggested that we understand realism as a view about meaning. A realist about Fs (other minds, the past, mathematics, etc.) is one who holds that, in virtue of the meaning of F-sentences, F-statements are determinately either true or false, independently of whether we can ever come to know their truth value. The anti-realist about Fs denies such determinacy in truth value. In Chapter 9, we consider the motivation for anti-realism, the character of its commitments, and look at a technical argument which questions its very coherence.

GOD

1

• Introduction	2
• The ontological argument	4
• The cosmological argument	7
• The teleological argument	11
• The argument from evil	13
• Concluding remarks	16

INTRODUCTION

One of the oldest metaphysical questions is: does God exist? In discussing this question, we understand 'God' in the classical philosophical sense of a being who is all-powerful (**omnipotent**), all-knowing (**omniscient**), and wholly good. In this chapter, we will examine three of the best-known philosophical arguments for God's existence, and one familiar argument against God's existence. The arguments in favour of God's existence are known as the ontological, cosmological and teleological arguments, though there are many versions of each argument. The argument against God's existence is the argument from evil.

The ontological argument attempts to prove God's existence by reason alone. It is a purely *a priori* argument. The idea is that simply by grasping the concept or idea of God, together with an understanding of what that idea entails, we can prove that God exists. The ontological argument thus purports to be a deductively valid proof of God's existence from *a priori* knowable premises. It is intended to be as cogent and compelling as any proof found in logic and mathematics.

The cosmological and teleological arguments are not based on *a priori* knowable premises, but on **contingent**, empirical premises. The cosmological argument starts from the fact that the universe around us exists. It continues: since the existence of the universe is contingent (there might have been nothing, rather than something), there must be a non-contingent or **necessary being** who created the universe.



The teleological argument (or argument from design) also proceeds from an empirical premise: not the premise that the universe exists, but the more specific premise that the universe contains wonderfully complex entities, such as flowers and eyes and brains, the existence of which, it is claimed, supports the hypothesis of a benign and loving creator. Such marvellous entities most likely came into existence by design, not by chance. Notice that this argument is not deductive in character. The claim is not that the existence of complex structures such as eyes and brains logically implies the existence of God, but only that the existence of such structures makes it reasonable to believe that God exists. The postulation of a divine creator best explains the existence of such structures.

One important argument against God's existence is the argument from evil. This is an argument against the existence of God, conceived of as all-loving, all-knowing and all-powerful. The argument from evil claims that the existence of evil is incompatible with the existence of God. A weaker, evidential, version of the argument from evil claims only that the existence of evil provides evidence against the existence of God, and concludes, given the prevalence of evil, that it is irrational to believe in a loving God.

THE ONTOLOGICAL ARGUMENT

There have been many different versions of the ontological argument throughout the history of philosophy, but the first, and most discussed, is that presented in the eleventh century by St Anselm, Archbishop of Canterbury, in his *Proslogion*. Here is a crucial paragraph from which we can reconstruct his argument:

Thus even the fool is convinced that something than which nothing greater can be conceived is in the understanding, since when he hears this, he understands it; and whatever is understood is in the understanding. And certainly that than which a greater cannot be conceived cannot be in the understanding alone. For if it is . . . in the understanding alone, it can be conceived to exist in reality also, which is greater. Thus if that than which a greater cannot be conceived is in the understanding alone, then that than which a greater cannot be conceived is itself that than which a greater can be conceived. But surely this cannot be. Thus without doubt something than which a greater cannot be conceived exists, both in the understanding and reality.¹

A reconstruction might proceed as follows:

- (1) God is that than which nothing greater can be conceived.
- (2) God either exists in the understanding alone or exists both in the understanding and in reality.
- (3) If God existed in the understanding alone, a greater being could be conceived, namely, a being with all God's qualities who exists both in the understanding and in reality.
- (4) But God is that than which no greater can be conceived ([1]).

St Anselm (1033–1109)

Anselm was born in Aosta in Italy. He became a monk and was later appointed Archbishop of Canterbury. As much a theologian as a philosopher, Anselm is credited with putting forward the first version of the ontological argument for God's existence. Anselm's belief in God did not rest on his proof; he simply wanted to make manifest God's existence and nature. As he once said 'I do not seek to understand that I may believe, but I believe in order to understand.' His ontological argument has had a mixed reception: Aquinas and Kant rejected it; Duns Scotus and Descartes proposed their own versions of it. Although the argument has few adherents today, there is no consensus on where the reasoning goes astray.

- (5) So God cannot exist in the understanding alone (from [3] and [4]).
 (6) So God exists both in the understanding and in reality (from [2] and [5]).
 (7) So God exists (in reality) (from [6]).

Premise (1) is intended to be a purely definitional truth. For Anselm, the word ‘God’ simply means (among other things) ‘that than which no greater can be conceived’, just as ‘triangle’ means ‘three-sided, three-angled plane figure’ and ‘spinster’ means ‘unmarried woman’. So the fool could no more sensibly deny that God is that than which nothing greater can be conceived, than he could sensibly deny that triangles are three-sided or spinsters unmarried.

Premise (2) is also intended to be truistic, and an instance of the following supposedly quite general truth: for any F that has been conceived, either F exists in the understanding alone or F exists both in the understanding and in reality. Thus, unicorns and dragons exist in the understanding alone, men and horses exist both in the understanding and in reality.

Premise (3) is motivated by the following train of thought. Suppose we consider two beings alike in their properties, except only that the first being exists in the understanding alone, while the second exists both in the understanding and in reality. Then the second being is greater than the first – existence in reality is a great-making property. This principle, together with premise (1), yields premise (3). Given (1), (2) and (3), (7) quickly follows.

How might we criticize the argument? To start with, we should not assume that all definitions or stipulations are coherent. Some are not. For example, I might try to define ‘meganumber’ thus:

- (M) Meganumber is that natural number than which there is no greater.

Assuming that ‘greater’ here means ‘larger’, (M) is incoherent. There is no largest natural number since the natural number series is infinite.

Is there any reason to think that premise (1) is similarly incoherent? It will be if God’s great-making qualities are non-maximal (i.e., qualities which can always be possessed to a greater degree, such as height or weight). But, plausibly, God’s great-making qualities are maximal. Qualities such as omnipotence, omniscience and perfect goodness seem to be maximal; no being can be more powerful than an omnipotent being, for example. So we cannot criticize Anselm’s definition of ‘God’ as we did definition (M).

It might be objected that, even if premise (1) is coherent, it cannot have any ontological consequences. A mere stipulation cannot generate real entities. When the word ‘spinster’ was first introduced into the language and defined as ‘unmarried woman’, the definition did not guarantee that the world contained spinsters. That there are spinsters is due, not to any definition, but to the (non-verbal) fact that some women choose not to marry. Or again, that we have meaningful

WHAT IS THIS THING CALLED METAPHYSICS?

definitions of words such as ‘dragon’ and ‘unicorn’ should not lead anyone to think that such creatures exist. So how can premise (1), a mere definition, possibly have ontological consequences?

However, it would beg the question against Anselm to press this objection at this stage. For Anselm could reasonably reply that, although many definitions indeed have no ontological consequences, his definition does. If we are to criticize Anselm fairly, we must examine the subsequent steps in his reasoning.

Once we do, however, concerns arise immediately. It soon becomes apparent that Anselm has a quite bizarre understanding of what it is to possess a concept or idea. The first three sentences in the quote above suggest the following chain of thought. I first understand a word ‘F’ (a general term, say). In virtue of understanding ‘F’, an F exists in my understanding, and has all the standard qualities of an F. We can then inquire whether Fs also exist in reality.

Thus, if I understand the term ‘unicorn’, a unicorn exists in my understanding, and that unicorn has the qualities typically associated with unicorns (four legs, spiral horn, lion’s tail, etc.) only it lacks the quality of real existence. But this is incredible. When I understand the word ‘unicorn’, I do not have something four-legged and spiral-horned in my mind! Anselm has committed what we might call the *fallacy of reification*. He has identified understanding a word or grasping a concept with the mind’s containing the object conceived. But this is to confuse concept with object: the concept is in my mind, but its object is not. The principles underlying premises (2) and (3) are glaring examples of this fallacy. Once we recognize that it is a fallacy – that when I understand the word ‘God’ there is not something in my mind which is omnipotent, omniscient, etc. – we must reject those premises. They are based on an untenable construal of what it is to understand a word.

Although the above suffices to dispose of Anselm’s version of the ontological argument, it’s worth mentioning a response made to Anselm by one of his contemporaries, Gaunilo of Marmoutiers. In his *On Behalf of the Fool*, Gaunilo contemplates an island than which no more excellent can be conceived and then writes:

Now if someone should tell me that there is such an island, I should easily understand his words. . . . But suppose he went on to say, as if by a logical inference: ‘You can no longer doubt that this island which is more excellent than all islands exists somewhere, since you have no doubt that it is in your understanding. And since it is more excellent not to be in the understanding alone, but to exist both in the understanding and in reality, for this reason it must exist. For if it does not exist, any island which really exists will be more excellent than it; and so the island already understood by you to be more excellent will not be more excellent.’²

The ontological argument

St Anselm's ontological argument is a classic example of a rationalist argument. The argument attempts to show that we can establish a substantial conclusion – God's existence – by reason alone. This contradicts the empiricist principle (associated with British philosophers such as Locke, Berkeley and Hume) that reason alone can never produce substantial knowledge of reality. The empiricists held that knowledge of reality relies essentially on input from the senses. The ontological argument is ingenious. It attempts to prove the existence of God merely from the idea or definition of God as 'that than which no greater can be conceived'. It would be extraordinary if a definition of a word could prove the existence of anything beyond itself. Fortunately, Anselm's argument does no such thing.

Gaunilo is here attempting a parody of Anselm's proof. That is, he is employing reasoning analogous to Anselm's to establish an obviously absurd conclusion. Clearly the world does not contain a perfect island (that is, an island than which no greater can be conceived), or a perfect horse, or a perfect crocodile, etc. If arguments analogous to Anselm's are unsound, Anselm's argument must be unsound too. Note that parody arguments, though potentially effective in one respect, are deficient in another. If successful, a parody argument shows that the original (parodied) argument goes wrong, but it provides no diagnosis of where it goes wrong.

However, it seems that Gaunilo's parody fails because his argument is not analogous to Anselm's in a crucial respect. I mentioned earlier that in order for Anselm's definition of God to be coherent, God's great-making qualities must be maximal ones, i.e., qualities which cannot be possessed to a greater degree. But the qualities that make an island excellent (such as abundance of coconuts, number of palm trees, preponderance of dusky maidens, etc.) are plainly not maximal. In which case the description 'island than which no greater can be conceived' (like 'natural number than which there is no greater') expresses no coherent concept.

THE COSMOLOGICAL ARGUMENT

Although rejecting Anselm's ontological argument, St Thomas Aquinas (c.1225–74) advocated the cosmological argument for God's existence. This argument can take many forms, one of which is known as 'the argument from contingency'. The Jesuit philosopher and historian Frederick Copleston, in a famous debate with Bertrand Russell, outlined a version of the argument from contingency:

WHAT IS THIS THING CALLED METAPHYSICS?

the world is simply the . . . aggregate of individual objects, none of which contain in themselves alone the reason for their existence. . . . [S]ince no object of experience contains within itself the reason for its own existence, this totality of objects must have a reason external to itself. This reason must be an existent being. Well this being is either the reason for its own existence, or it is not. If it is, well and good. If not, then we must proceed further. But if we proceed to infinity . . . there is no explanation of existence at all. So . . . in order to explain existence, we must come to a being which contains within itself the reason for its own existence – . . . which cannot not exist.³

Later in the debate, Copleston gives a succinct recapitulation of the essence of the argument:

The series of events [comprising the history of the universe] is either caused, or it is not caused. If it is caused, there must obviously be a cause outside the series. If it is not caused, then it is sufficient to itself; and if it is sufficient to itself it is what I call necessary. But it [the universe] can't be necessary since each member is contingent, and . . . the total has no reality apart from its members.⁴

In order to evaluate these presentations, it might be useful to begin with some of Russell's critical comments. Both arguments make play with modal notions. The conclusion of Copleston's reasoning is that a necessary being exists (namely, God). Russell takes issue with this: '[t]he word "necessary" . . . can only be applied significantly to propositions. And . . . only to such as are analytic, that is to say such as it is self-contradictory to deny.'⁵ For Russell,

St Aquinas (c.1225–74)

Thomas Aquinas was born in Roccasecca in Italy. He studied the works of Aristotle at the University of Naples and then became a Dominican friar. Aquinas's best-known work is his *Summa theologiae* (1266–73). He held that faith and reason cannot conflict, since reason, if properly exercised, will never yield deliverances contrary to faith. Aquinas denied that the existence of God could be proved merely by reflection on the idea of God – hence he rejected Anselm's ontological argument. However, he did hold that the existence of God could be established from premises concerning the nature and existence of the universe (*Summa theologiae* 1a, qu. 2, art. 3). These are Aquinas's famous 'five ways' to prove the existence of God, one of which draws on the contingent existence of the universe (a version of the cosmological argument).

no proposition of the type 'A exists' can be analytic. Hence, talk of 'necessary beings', or indeed of 'contingent beings', is a solecism.

Of course, if the ontological argument is sound then 'God exists' is analytic, but that argument has been found wanting. However, even if we concede that 'God exists' is not analytic, there is still a reply to Russell. Since the work of Saul Kripke in the 1960s and 1970s, it is generally accepted that analyticity ('truth in virtue of meaning') is not the only source of necessity: there are necessary truths which are not analytic.⁶ Even before Kripke, some philosophers suspected that truths like 'nothing can be red and green all over' were necessary yet not analytic. But Kripke provided a theoretical framework in which to make sense of non-analytic necessities and provided more clear-cut examples: e.g., 'Tully is Cicero', 'water is H₂O', 'this lectern is made of wood'. If this modern orthodoxy is right, the non-analyticity of 'God exists' does not thereby count against its necessity.

But there is a deeper problem with Russell's response. It is surely true to say of any of the ordinary things we encounter: 'this might not have existed'. Indeed one can truly say it of oneself: I might not have existed (e.g., if my parents had not met). But then it follows immediately that 'I exist' is a contingent truth, and hence that I am a contingent being. So it makes sense to talk of contingent beings. In which case it must at least make sense to talk of necessary beings (whether or not there are any such beings).

However, this dispute is not really to the point. Although Copleston does use the phrases 'being . . . which cannot not exist' and 'necessary' to describe God, the relevant aspect of God for Copleston's purposes is not his necessity but his self-sufficiency: God contains within himself the reason for his own existence.

The trouble is that the notion of something containing the reason for its own existence is too obscure for us to do anything with it. Russell says that unless 'sufficient reason' means 'cause', he doesn't know what it means. But 'sufficient reason' can't mean 'cause', otherwise God would be being said to be 'self-caused', which Copleston agrees makes no sense. So Russell's complaint seems justified: if 'sufficient reason' doesn't mean 'cause', what does it mean? Hence, Copleston's first presentation of the cosmological argument needs clarification before we can assess it.

Copleston's second, slimmer, presentation of the argument does not make use of the notion of self-sufficiency until later on in the argument. It begins with the premise: the (entire) universe is either caused or not caused. Russell objects that this premise makes no sense. It makes sense to ask for the cause of this or that event (that is how we learn the word 'cause'), but it makes no sense to ask for the cause of all events, of the entire universe.

But the reason Russell gives for this claim is unconvincing. He points out that it is a fallacy to infer that a totality has some property just because each member of the totality has that property. And that is indeed a fallacy (the

WHAT IS THIS THING CALLED METAPHYSICS?

fallacy of composition).⁷ To use Russell's example: from the fact that every human being has a mother, we cannot validly infer that the entire human race has a mother. But Copleston does not commit such a fallacy. He does not reason: every event has a cause, so the totality of events must have a cause. He just thinks it makes perfectly good sense to ask whether the universe has a cause. And wondering whether the universe has a cause does not seem to involve the sort of category mistake involved in wondering whether the human race has a mother.

So let us grant Copleston his opening premise: the entire universe is either caused or uncaused. How does the argument proceed? Since the notion of self-sufficiency has been deemed too obscure, we shall have to operate with the notion of necessity. So we can recast Copleston's second presentation thus:

- (1) The universe is either caused or uncaused.
- (2) If caused, the cause lies outside the universe.
- (3) If uncaused, the universe is necessary.
- (4) The universe is not necessary.
- (5) The universe is caused (from [3] and [4]).
- (6) Hence, the cause of the universe lies outside the universe (from [2] and [5]).⁸

Since we are happy with the terminology used here, how should we respond? Premises (1) and (2) are uncontroversial: (1) is a truism (once we have allowed, *contra* Russell, that it makes sense to talk of the universe having a cause), and (2) seems right since if the cause of the universe lay within the universe, it (the cause) would be part of the universe. But no part of a thing can cause the thing itself, so the cause of the universe must lie outside the universe.

Premise (4), though it has been denied by some philosophers, is also plausible.⁹ Assuming that there might have been nothing rather than something, the existence of the universe is a non-necessary or contingent matter. Copleston's reason for (4) may seem to commit the fallacy of composition: he argues that the entire universe is contingent because each element of it is contingent. But his reasoning does not seem fallacious in this case. How could a totality be necessary if each of its elements is contingent? The fallacy of composition may be a fallacy with respect only to certain kinds of property, and the property of contingency may not be one of them.

So if the argument is to fail, it must fail at premise (3). Could one hold that the universe has no cause, yet is nonetheless contingent? *Prima facie*, yes. One might hold, as Aristotle did, that the universe has always existed, and yet is contingent.¹⁰ Second, many modern cosmologists hold that the universe, though it has not always existed, came from literally nothing: it has no cause. On this view, the universe had no cause, yet its existence was not necessitated. A defender of the cosmological argument must show that such views are untenable

William Paley (1743–1805)

William Paley studied and taught at the University of Cambridge, before becoming a clergyman in the Church of England. He wrote three major works. In *Principles of Moral and Political Philosophy* (1785) he defended utilitarianism (the view that we ought to do whatever maximizes human happiness) on theological grounds. In *A View of the Evidence of Christianity* (1794) he discussed belief in miracles and commented on the discussion of miracles in Section 10 of Hume's *Enquiry*. *Natural Theology* (1802) is Paley's best-known work in which he put forward his version of the argument from design, arguing that the complex and varied world around us is best explained as the result of divine creation.

and, until he does so, the cosmological argument will be less than wholly compelling. In addition, even if the argument were compelling, its conclusion implies nothing about the nature of the cause of the universe (e.g., whether the cause is good or loving).

THE TELEOLOGICAL ARGUMENT

The teleological argument for God's existence, also called the argument from design, attempts to argue for the existence of God, not simply from the contingent existence of the universe, but from the intricacy and complexity of its structure.

Although the argument had been around since the Stoics, William Paley (1743–1805) devised the best-known version of it, based on the following analogy. Paley says that when we encounter a stone in a desert, we do not suppose that it had a designer. But if we were to come across a complex instrument, such as a watch, we would assume it had a maker, and we would be justified in this assumption. For by examining its parts, we could see that they were 'framed and put together for a purpose'.¹¹ Hence we may reasonably conclude that 'the watch must have had a maker. . . . There must have been, at some place or other, an artificer or artificers who formed it for the purpose which we find it actually to answer; who comprehended its construction and designed its use.'¹²

Paley then argues that we can reason analogically from the universe we find around us, in particular from complex biological items, such as flowers, hearts and eyes, to the existence of a designer of the universe: God. If we are justified in moving from the existence of the watch to the existence of a watchmaker, then we must be equally justified in moving from the existence of complex biological structures to the existence of a creator. Furthermore, the beauty and symmetry of such structures points towards the benevolent nature of their creator.

WHAT IS THIS THING CALLED METAPHYSICS?

How should we respond to this argument? Note that, unlike the previous arguments, the teleological argument is not presented as deductively valid. Since neither creation *ex nihilo*, nor some random rearrangement of particles into a watch-like configuration, is logically impossible, the existence of a watch does not logically imply the existence of a watchmaker. Paley's claim is rather that, upon coming across a watch, it is reasonable to believe that it was made by a watchmaker. That is the best explanation of the existence of the watch. Analogously, on seeing the complex world around us, it is reasonable to believe that it was created by God.

The first thing to note is that even if Paley's reasoning were to be found persuasive, the most it would establish is that God created the universe, not that he still exists. It is quite consistent with Paley's argument that God 'left the building' soon after the act of creation. So his argument gives us no reason to think that God exists now, just as we have no reason to think that the watchmaker exists now. (Interestingly, Sir Isaac Newton believed that, without God, the planets would fly off in different directions. The regular motion of the planets, for Newton, thus made it reasonable to believe in the continued existence of God.)

Second, in David Hume's *Dialogues Concerning Natural Religion* (1779), published after Hume's death and before the publication of Paley's 1802 treatise, one of the imaginary participants, Philo, points out how little we are entitled to infer about the creator from observation of the (far from benign) world around us:

This world, for aught he [a defender of the argument from design] knows, is very faulty and imperfect, compared to a superior standard; and was only the first rude essay of some infant deity, who afterwards abandoned it, ashamed of his lame performance: it is the work only of some dependent, inferior deity; and is the object of derision to his superiors: it is the production of old age and dotage in some superannuated deity; and ever since his death, has run on at adventures, from the first impulse and active force which it received from him.¹³

Indeed, Philo suggests, the argument from design gives us no reason to suppose that there was only one creator. Why not conclude that just as '[a] great number of men join in building a house or ship, in rearing a city, in framing a commonwealth; . . . may not several deities combine in contriving and framing a world?'¹⁴

Third, and most important, we must remember that Paley was writing before Darwin's theory of evolution and the dawn of modern cosmology. For anyone in 1800, the only explanation for the order in the universe was divine creation. Paley's reasoning would have seemed then to be the merest common sense.



However, from our current vantage point, we have another explanation available to us – Darwin’s theory of evolution – and Paley’s reasoning no longer seems so inescapable.

If we are wondering how human beings, with all their intricate internal organs (such as eyes and hearts) came to be here, we now have two competing explanations: Darwin’s theory of evolution by natural selection and the divine explanation. The theory of evolution has enjoyed great explanatory success, and is so well-established that it is not seriously in doubt. It (currently) provides the best explanation for the existence and nature of human and other species. In which case, Paley’s reasoning lapses: the eye is not analogous to a found watch. The availability of the Darwinian explanation of the origin of the species has effectively scuppered the teleological argument.

THE ARGUMENT FROM EVIL

Thus far we have been criticizing arguments for God’s existence. Now we are going to consider a well-known argument against the existence of God: the argument from evil. This is an argument for the non-existence of God, on the assumption that God (if he exists) has the characteristics of omnipotence, omniscience and perfect goodness.

The basic thought behind the argument from evil is that the existence of evil, both natural evil (death and suffering caused by natural disasters) and moral evil (human beings’ cruelty to each other), are incompatible with the existence of God. Hence, given that evil exists, God does not.

However, we need to unpack things a little in order to become clear about the character of the alleged incompatibility. The argument from evil holds that the following propositions form an incompatible quartet:

- (1) God is omniscient.
- (2) God is omnipotent.
- (3) God is wholly good.
- (4) Evil exists.

If (1)–(4) are incompatible, they cannot all be true. Since (4) is undeniably true, one of (1)–(3) is false. But if any of (1)–(3) is false then God does not exist. If there is no being who is omniscient, omnipotent and wholly good, then there is no God. According to the argument from evil, there is no such being and so, no God.

Why think (1)–(4) are incompatible? Evil, both natural and moral, exists. If God is omniscient, he knows about the existence of evil; if he is wholly good, he wants to prevent evil; if he is omnipotent, he is able to prevent evil. So why doesn’t he? If there was a being who was omniscient, omnipotent and wholly

WHAT IS THIS THING CALLED METAPHYSICS?

good, the world would contain no evil. But the world does contain evil. So no such being exists.

This is a pleasingly straightforward argument. There have of course been many replies to it. These replies attempt, in different ways, to show how (1)–(4) might turn out, after all, to be compatible. Here is one line of reply. What if certain evils are *necessary* for the existence of certain goods? That is, without those evils there would not be those goods. Then the theist could argue that the value of the goods in question outweighs the cost of the accompanying evils. In which case, the existence of such evil need not count against God's love or knowledge or power.

A simple example illustrates the idea: I suffer pain in the dentist's chair, but that is necessary in order to have pain-free, healthy teeth in the future. However, this necessity – that I *must* have pain now in order to have less pain in the future – is a case of causal or technological necessity. At the current state of technology, painful visits to the dentist are necessary to have healthy teeth in the future. But God is omnipotent. He is not bound by causal or technological necessities. God's omnipotence means that he can do anything which is logically possible. He could have arranged things so that visits to the dentist were painless or equipped us with teeth that never decayed. So why didn't he?

What the 'necessary evils' defence requires, if it is to have a chance of being plausible, is evils which are **logically necessary** in order for there to be certain goods. Then, if it is desirable that the world contain such goods, God cannot be criticized for allowing such evil, since it is logically impossible to have such goods without the accompanying evil. Even God cannot do the logically impossible.

Examples of goods which seem logically to require certain evils are heroism, benevolence and sympathy. Heroism, benevolence and sympathy are possible only because the world contains, e.g., natural disasters, towards the victims of which other people can be heroic, sympathetic and benevolent.

However, we can make three rejoinders here. First, goods such as sympathy do not seem logically to require real disasters. The appearance of a disaster would be enough to elicit feelings of sympathy. If God is omnipotent, he surely could have arranged for the world to contain illusory tsunamis to which we could respond. But he chose not to. Why?

Second, the 'necessary evils' defence is too limited. The world contains many evils which have no accompanying or mitigating good. Many cases of death and suffering throughout human (and animal) history are never observed or recorded. Such evils obviously cannot be justified in terms of their production of goods in others.¹⁵

Third, there is something morally obnoxious about the 'necessary evils' defence. Are not the victims of natural disasters being used as a mere means towards others' moral improvement? 'The suffering of others is good for my soul' is hardly the dictum of a moral individual.



A second line of reply, attempting to show the compatibility of (1)–(4), is known as ‘the free-will defence’. God gave people free will, and that is a good. A world containing free people who sometimes freely choose good and sometimes freely choose evil is better, it is claimed, than a world of automata programmed always to do good. But free will comes at a cost: many evils enter our world as a result of free human action. Since God is not responsible for such evils, they do not count against God’s goodness or love or knowledge.

This reply is a variant of the first: in order for there to be a certain good (free will) it is necessary to put up with the bad consequences stemming from evil exercises of that good. However, the free-will defence is open to the following five objections.

- (1) The defence only addresses cases of moral evil (evil which results from free human choice). It does nothing to explain why God would allow natural evils.
- (2) It assumes that we have free will. Some philosophers have denied that we have free will, either because they are fatalists or they think that determinism is true or because they find the concept of free will confused (see Chapter 7).
- (3) It is not obvious that the good of free will outweighs the amount of evil that humans have actually produced. It’s implausible to think that the value of one person’s free will outweighs or compensates for any amount of evil produced by its exercise. Wouldn’t the world have been a better place if Hitler or Stalin, for example, had been automata, programmed to do good acts?
- (4) It’s not true that, if humans have free will, there must be bad consequences. Even if God gave us free will, he could have intervened after any evil choice was made, ensuring that no bad consequences ensued. This would surely be better than letting the bad consequences ensue – so why didn’t God intervene?
- (5) There is another, more radical, way in which God might have given us free will, yet ensured no bad consequences followed. Why did God not create beings who always freely chose the good?¹⁶ It might be thought that such beings are impossible, but that is far from clear. Do the angels not always freely choose the good? More soberly, we are beings who *sometimes* freely choose the good. Why would it be incoherent to imagine us *always* choosing the good? Surely there is no *a priori* upper limit on the number of times an agent might freely choose the good.

We can present this last objection in the following, picturesque, way. Imagine God surveying all the possible worlds, wondering which one to make actual.¹⁷ In some of those worlds, everyone freely chooses the good. In some, no one

WHAT IS THIS THING CALLED METAPHYSICS?

The problem of evil

The main reason why many reflective people do not believe in God is the existence of evil. They cannot understand how an all-knowing, all-powerful, all-loving God could allow evil into the world (e.g., the suffering that ensues after a natural disaster such as a flood or famine), nor why he doesn't prevent the evil consequences of people's malicious choices. Theists (i.e., those who believe in the existence of God) have offered various theodicies which purport to show how God's knowledge, power and love are consistent with the existence of evil. Kant believed that all theodicies were bound to fail (*Religion within the Limits of Reason Alone*, 1793), and our discussion seems to bear this out.

chooses the good. In yet others, some men choose good, some choose evil. Why did God not decide to actualize one of the worlds from the first category? If he was able to, and wanted to, and knew how to, he would have. But he didn't; so God is lacking in either benevolence or knowledge or power.

A final line of reply to the argument from evil points out that (1)–(4) are not formally inconsistent. Perhaps there is a reason why God permits evil, but that reason is unknown or unknowable to us. However, it's hard to distinguish this reply from wishful thinking.

CONCLUDING REMARKS

All the standard arguments for God's existence are open to objections. The existence of evil provides a strong argument against God's existence. It is therefore quite reasonable to believe that there is no being who is omnipotent, omniscient and wholly good.

STUDY QUESTIONS

- Does Anselm really commit the fallacy of reification in his presentation of the ontological argument?
- Was Russell right to question the coherence of the notion of a necessary being?
- Can you construct a more plausible version of the argument from design based on the massive improbability of a life-permitting universe?
- What is the most promising reply a theist might make to the argument from evil?



ANNOTATED FURTHER READING

- M. M. Adams and R. M. Adams (eds), *The Problem of Evil* (Oxford: Oxford University Press) 1992. An interesting, and sometimes demanding, set of essays devoted to the problem of evil. The introduction is very useful, and I would recommend especially the articles by J. L. Mackie, N. Pike, A. Plantinga and R. M. Adams.
- J. L. Mackie, *The Miracle of Theism* (Oxford: Clarendon Press) 1982. An excellent introduction to all the major arguments for and against God's existence. Highly recommended.
- A. Plantinga (ed.), *The Ontological Argument* (New York: Anchor Books) 1965. A useful little book, in two parts. In the first part, Plantinga selects presentations of the ontological argument from Anselm, Descartes, Spinoza and Leibniz, alongside critical commentaries by Gaunilo, Caterus, Gassendi, Kant and Schopenhauer. In the second part, Plantinga selects a (now rather dated but still interesting) collection of twentieth-century commentators. The most interesting contribution is that of N. Malcolm, with replies by A. Plantinga and P. Helm.
- A. Plantinga, *God, Freedom and Evil* (London: Allen & Unwin) 1975. Again a useful book, in two parts. In the first part, Plantinga puts forward his own ingenious theodicy, i.e., his solution to the problem of evil, based on his version of the free-will defence. In the second part, he outlines and criticizes the standard cosmological, teleological and ontological arguments for God's existence. He ends by proposing a version of the ontological argument which, he thinks, establishes not the truth but the rational acceptability of theism. Both discussions make play with the notion of a possible world.
- W. L. Rowe, *The Cosmological Argument* (Princeton, NJ: Princeton University Press) 1975. A very clear, detailed and historically informed discussion of all aspects of the cosmological argument.

INTERNET RESOURCES

- A. Plantinga (1998) 'God, Arguments for the Existence of', in E. Craig (ed.), *Routledge Encyclopedia of Philosophy*. Retrieved 31 May 2006 from <<http://www.rep.routledge.com/article/K029>>.
- G. Oppy, 'Ontological Arguments', *The Stanford Encyclopedia of Philosophy* (Fall 2005 edition), Edward N. Zalta (ed.). Retrieved 31 May 2006 from <<http://plato.stanford.edu/archives/fall2005/entries/ontological-arguments>>.
- M. Tooley, 'The Problem of Evil', *The Stanford Encyclopedia of Philosophy* (Winter 2004 edition), Edward N. Zalta (ed.). Retrieved 31 May 2006 from <<http://plato.stanford.edu/archives/win2004/entries/evil>>.

2 EXISTENCE

• Introduction	21
• The great sweep of being	22
• Why does our world exist?	23
• Modal realism assessed	25
• Non-existent objects	27
• What is existence?	31
• Concluding remarks	33

INTRODUCTION

Existence gives rise to a host of distinctively philosophical questions that lie at the heart of metaphysics. In this chapter we will address the following questions. First, what is the extent or sweep of being?¹ In particular, do non-actual or merely possible entities exist? Is what actually exists merely a small part of all that exists? Second, why does the universe exist? Why is there something rather than nothing? And why *this* something rather than some other? Third, are there non-existent objects? Why would anyone think that there are? Fourth, what is existence? In particular, is existence a property of ordinary objects, or not? Why does it matter how we answer this question? Answering questions 3 and 4 will require delving into complicated issues concerning meaning, reference and logical structure.

THE GREAT SWEEP OF BEING

It is very natural to think that what exists is just what actually exists. Of course, there might have been flying pigs, fire-breathing dragons, golden mountains, perpetual-motion machines and so on. But such things do not exist. They might have, but they don't. David Lewis disagrees. What might have existed, does exist – in some other possible world. Possible worlds exist too, in just the way that the actual world does. This is the infamous doctrine of **modal realism**. Lewis writes:

I advocate a thesis of plurality of worlds, or *modal realism*, which holds that our world is one world among others. There are countless other worlds, other very inclusive things. . . . They are isolated: there are no spatiotemporal relations at all between different things that belong to different worlds. Nor does anything that happens at one world cause anything to happen at another. . . . The other worlds are of a kind with this world of ours. . . . Nor does this world differ from others in its manner of existing. . . . The worlds are not of our own making.²

The doctrine of modal realism is captured in the following seven theses:

1. Possible worlds (including our world) exist.
2. They are the same kind of thing as our world.
3. Possible worlds, like the actual world, cannot be reduced to anything else.
4. There is nothing ontologically special about the actual world. Each world is actual *to its inhabitants*.
5. 'Actual' is simply an indexical, like 'here'.
6. Possible worlds are spatio-temporally isolated, and hence causally isolated, from each other.
7. Possible worlds are not mind-dependent.

It is important to appreciate what the doctrine of modal realism does and does not amount to. It is not the view that 'many worlds actually exist' (a view that is sometimes proposed by certain interpreters of quantum mechanics). On such a view the actual world is much bigger than we think it is, and that is not Lewis's view. Nor is modal realism the view that there are possible worlds or possible beings which do not exist. Lewis draws no distinction between 'there are Fs' and 'Fs exist'. His view is that the set of actually existing things is a small subset of all existing things, just as the set of people living in Canberra is a subset of the set of the world's total population, or the set of people existing in 1940 is a subset of the set of twentieth-century people.



Modal realism

The doctrine of modal realism proceeds in two stages. First, it analyses ordinary modal talk into possible world talk. The true sentence ‘there might have been blue swans’ is rendered as ‘there is a possible world containing blue swans’. Second, it endorses realism about possible worlds and their inhabitants. Other possible worlds exist in just the way that our world exists. Blue swans exist, but not in our world. Some philosophers deny that modal talk involves implicit quantification over worlds. Others accept the translation into possible world talk, but deny that other worlds exist. The actual world is privileged: it is the only world that exists. What then makes it true that there is a possible world containing blue swans if not the existence of a concrete world containing blue swans? Actualists, as they are called, postulate some surrogate truth-maker in place of a world: e.g., a complete and consistent set of sentences containing the sentence ‘swans are blue’. This set is what makes it true that there might have been blue swans. But is this plausible? Even if there had been no sets of sentences (e.g., if there had been no minds), there still might have been blue swans.

Each possible world corresponds to a different way our world might have been. Some possible worlds are like our world – they contain flesh-and-blood people and donkeys, just as our world does. Yet other possible worlds are more exotic and contain kinds of entities undreamt of in this world. Possible worlds are concrete worlds like our own, irreducible to anything else (e.g., to abstract objects such as sets of propositions).

According to modal realism, actual existence (existing in the actual world) is not in any way privileged. It is not the mark of the real since other worlds are equally real. The word ‘actual’ uttered by the inhabitant of any world simply refers to the world of the inhabitant. Each world is actual to its inhabitants, just as each place is ‘here’ to its occupants. ‘Actual’ is a mere **indexical**.

In addition, Lewis thinks that possible worlds are spatio-temporal unities, and are spatio-temporally isolated from each other. Possible worlds are not any spatial or temporal distance from each other.³ Finally, possible worlds are not fictions, made up by us. All possible worlds (including our own) exist independently of us: we neither caused them to exist nor sustain them in existence.

WHY DOES OUR WORLD EXIST?

Lewis’s modal realism has implications for one of the most puzzling questions about existence: why does our world exist? This question can be usefully divided into two distinct questions: why is there something rather than nothing? and why *this* something rather than some other?

Why is there something rather than nothing?

For centuries many have thought that this question demands an answer, and that it can only be answered by appeal to God. There is something rather than nothing because God chose to create this universe and the creatures within it. But even if we think that there is no God, the question still seems a sensible one. Some deny this, and hold that it can receive no interesting answer. There was a coincidence of initial conditions which allowed the physical universe to emerge and later allowed life to evolve. But there is no explanation of why just those initial conditions obtained. Many are content with this, but many are not. If the chance of life-permitting initial conditions obtaining is so incredibly low, must there not be an explanation of why those conditions actually obtained? In asking this question, it is assumed that there is only one world: the actual world. If this assumption is false, as modal realists believe, then our world is merely one among many life-permitting worlds. In that case, it is not so mysterious that our world is life-permitting: many worlds are, and ours happens to be one of them.

Lewis has a quick answer to the first question. For Lewis there could not be:

an absolutely empty world. A world is not like a bottle that might hold no beer. The world *is* the totality of things it contains . . . [I]f there isn't even the bottle, there's nothing there at all. . . . There can be nothing much: just some homogeneous unoccupied spacetime, or maybe only one single point of it. But nothing much is still something, and there isn't any world where there's nothing at all. That makes it necessary that there is something.⁴

Lewis doesn't regard this as an explanation of why there is something rather than nothing (he thinks there can be no such explanation). He simply thinks it shows that the question rests on a false assumption – the assumption that there might have been literally nothing, an absolutely empty world.

However, even if Lewis is wrong on this point and there is an empty world, his doctrine of modal realism would still have implications for the question: why is there something rather than nothing? If there is an empty possible world, we can reply: the 'rather than' is misplaced – there is nothing (in the empty possible world). According to modal realism, that world exists just as much as our world does.

What of the question: why *this* something rather than some other? Someone posing this question normally has in mind the more specific question: why is this world life-permitting and not life-denying? Derek Parfit nicely describes why we might find this question pressing:



For life to be possible, the initial conditions had to be selected with the kind of accuracy that would be needed to hit a bull's-eye in a distant galaxy. Since it is not arrogant to think life special, this appearance of fine-tuning needs to be explained. Of the countless possible initial conditions, why were the ones that allowed for life *also* the ones that actually obtained?⁵

One obvious answer is theistic: God, whose existence is necessary, arranged things that way. But obviously this is a good answer only if God exists. (See Chapter 1.) Another answer is that there is no answer. The actual universe just is life-permitting, and no explanation of this can be given. But this response is apt to seem unsatisfying. Fortunately, there is a third answer:

Our Universe may not be the whole of reality. Some physicists suggest that there are many other Universes – or, to avoid confusion, *worlds*. These worlds have the same laws of nature as our own world, and they emerged from similar Big Bangs, but each had slightly different initial conditions. On this *many-worlds hypothesis*, there would be no need for fine-tunings. If there were enough Big Bangs, it would be no surprise that, in a few of those, conditions were just right for life. And it would be no surprise that our Big Bang was one of those few.⁶

Although Parfit may not have had in mind Lewis's doctrine of modal realism when he wrote these words, the same point applies. If modal realism is true, and our world is one world among countless others, then it is not so surprising that our world supports life. Many worlds do, and our world happens to be one of them. Hence, by appeal to modal realism, the need for any theistic explanation is avoided. Since modal realism explains something which would otherwise be puzzling, this is a point in its favour.

MODAL REALISM ASSESSED

Lewis claims a number of virtues for modal realism, not the least of which is that it allows for a neat analysis of our modal talk (talk involving words such as 'might', 'must', 'possibly', 'necessary', etc.) and of our use of counterfactuals (conditionals of the form 'if A hadn't happened, B wouldn't have happened'). For example, we understand the sentence 'pigs don't fly' and we know what makes it true: how things actually are. Yet we also understand 'pigs might fly' and take it to be true. But what makes it true? Not, it might be thought, how things actually are. What then? Lewis provides an answer: how things are in another (very different) possible world. The sentence 'pigs might fly' is rendered by Lewis as 'there is a possible world in which pigs fly'.

WHAT IS THIS THING CALLED METAPHYSICS?

Since there is such a possible world, and since such a world is a concrete world like ours, the mystery of what makes such modal truths true is removed. Similarly sentences of the form ‘possibly P’ are rendered as ‘there is a possible world in which P’, and sentences of the form ‘necessarily P’ are rendered as ‘in all possible worlds P’. ‘Possibly’ and ‘necessarily’ are thus taken to be **quantifiers** over possible worlds.

A puzzle arises, however, in the case of modal claims about particular, actual individuals. Suppose I say (truly): Gore might have won the 2000 US presidential election. What makes this true? As before, Lewis says: how things are in some other possible world. But how are things in that world? Do they contain Gore himself, hands clenched in victory? That is one view. On the multiple-existence theory, individuals exist in more than one possible world.

However, this is not Lewis’s view. For Lewis, as for Leibniz, individuals are world-bound. Gore exists in one world (our world) and in no other. How, though, can a possible world not containing Gore be the truth-maker for the sentence ‘Gore might have won the 2000 US presidential election’? Lewis’s answer is that this other world contains a *counterpart* of Gore (someone very similar to Gore, but not Gore) who does win the 2000 election. Thus it’s true that Gore might have won the 2000 presidential election because there is a possible world, similar to our world, containing a counterpart of Gore who does win the election.

There is thus a debate between multiple-existence theory and counterpart theory. Saul Kripke famously complained that counterpart theory could not do justice to our modal claims:

[On counterpart theory] if we say ‘Humphrey might have won the election . . .’, we are not talking about something that might have happened to *Humphrey* but to someone else, a ‘counterpart’. Probably, however, Humphrey could not care less whether someone else, no matter how much resembling him, would have been victorious in another possible world.⁷

There has been much discussion over whether Kripke’s objection hits its target. Counterpart theorists retort that a counterpart of Humphrey winning the election is what it is for it to be true that Humphrey might have won the election. The multiple-existence theorist replies that it is not. It is unclear where victory lies. But we can make a weaker point. Multiple-existence theory provides a smoother semantics than counterpart theory for our modal talk. On the multiple-existence theory, ‘Gore might have won the election’ is true just if there is a possible world in which Gore wins the election. This is a more straightforward account than the counterpart theorist’s, and that is a point in favour of multiple-existence theory.

Many philosophers reject the backdrop of modal realism against which we have been conducting this dispute. Because of its ontological profligacy, it is a doctrine with few adherents. Many find it simply incredible. This is quite understandable. We are being asked to believe that there are worlds containing talking donkeys and flying pigs which exist in just the full-blooded way that our pigs and donkeys exist. But if that is our response, we should see Lewis as laying down a challenge: explain what makes our modal claims true without making any reference to possible worlds, realistically construed. Perhaps this can be done. For example, on the story view, a sentence such as ‘pigs might fly’ is true just if the sentence ‘pigs fly’ is a member of a complete and consistent story (set of sentences). Defenders of modal realism will reply that possibilities are objective and mind-independent. Even if there had been no sentences, there still might have been flying pigs. Hence the story view cannot be right. And so the debate goes on.

What should we conclude about Lewis’s doctrine of modal realism? It is a doctrine with significant explanatory advantages. It provides straightforward truth-makers for our modal claims. It undermines a presupposition of the puzzling question ‘Why is there something rather than nothing?’, and it allows a non-mysterious answer to the question ‘Why is this world life-permitting rather than life-denying?’ However, these advantages are offset by its ontological excess. We are asked to believe that dragons, goblins and golden mountains exist in just the way that horses, trees and rocks do. This is hard to accept. Nonetheless, anyone who rejects modal realism owes us an alternative account of the truth-makers for modal claims and needs to explain why, of all the possible worlds, our world should have the privilege of existing.

NON-EXISTENT OBJECTS

Lewis may disagree with common sense about what exists, but he doesn’t hold that there are things which don’t exist. Colin McGinn does. He is a modern defender of the view that there are non-existent objects. In a recent book he wrote:

we find it natural to talk in the following way. Not everything that we refer to exists: Venus does, Vulcan doesn’t; horses do, unicorns don’t. There are merely fictional entities as well as things that really exist. To exist is to have a property that only some of the things we refer to have – those that exist as opposed to those that are merely fictional.⁸

In propounding this view, McGinn is following in the footsteps of the Austrian philosopher Alexius Meinong, who happily embraced the non-existent

Alexius Meinong (1853–1920)

Alexius Meinong was an Austrian philosopher, heavily influenced by his teacher, Franz Brentano (1837–1917). From 1889 Meinong taught at the University of Graz, and made important contributions to philosophy and philosophical psychology. Meinong's work begins from his philosophy of mind and, in particular, from Brentano's thesis of intentionality: the thesis that it is a mark of the mental that mental states are directed towards objects. This led Meinong to a full-blown theory of objects that embraces possible objects (the golden mountain), impossible objects (the square triangle), and incomplete objects (something tall). Any subject of a true predication is an object. For Meinong, 'the square triangle is triangular' is true, so there is a square triangle. These objects are mind-independent, yet are all potential objects of thought. Meinong also believed in objective values, such as the good and the beautiful, detectable through emotions and desires.

and much else besides. Why did Meinong hold such an extraordinary view? Here is a famous passage from Bertrand Russell:

It is argued, e.g., by Meinong, that we can speak about 'the golden mountain', 'the round square', and so on; we can make true propositions of which these are the subjects; hence they must have some kind of logical being, since otherwise the propositions in which they occur would be meaningless. In such theories, it seems to me, there is a failure of that feeling for reality which ought to be preserved even in the most abstract studies. Logic, I should maintain, must no more admit a unicorn than zoology can; for logic is concerned with the real world just as truly as zoology, though with its more abstract and general features. To say that unicorns have an existence in heraldry, or in literature, or in imagination, is a most pitiful and paltry evasion.⁹

Russell's central task in the chapter from which this quote is taken, and in his more famous 1905 article 'On Denoting', is to frustrate Meinong's line of reasoning.¹⁰ From the mere fact that a grammatical subject term is meaningful, and features in true, meaningful sentences, it does not follow that it refers to something. Thus Russell rejects quite emphatically the view that meaning is reference. Emphatically, but not completely: Russell does think that there is a restricted range of genuine singular terms for which meaning is reference (e.g., the first person pronoun 'I', and terms referring to one's immediate experiences, such as 'this headache'). For Russell, these are the only genuine referring terms.¹¹ The meaningful use of such terms ('logically proper names', as they are known) guarantees that they have a reference.



Bertrand Russell (1872–1970)

Bertrand Russell was born in Monmouthshire, England, part of an aristocratic family (he inherited the titles of earl and viscount later in life). Both his parents died while he was an infant and he was raised by his strongly religious grandparents. He studied at Trinity College, Cambridge where, under McTaggart's influence, he was (briefly) an idealist. He later taught at Trinity as a contemporary of G. E. Moore, counting Wittgenstein among his pupils. Russell wrote extensively in almost every branch of philosophy, most significantly in logic, philosophy of mathematics, philosophy of language and mind, and epistemology, but also in social and political philosophy. The latter writings were intended for a general audience. A great prose stylist, he was awarded the Nobel Prize for Literature in 1950. Two of his books – *The Problems of Philosophy* (1912) and *A History of Western Philosophy* (1945) – were philosophical bestsellers. Among his most important and influential philosophical contributions were the theory of descriptions and Russell's paradox (which showed that the intuitive 'axiom of comprehension' – to each property there corresponds a set of exactly the objects with that property – leads to paradox).

All other grammatical subject terms (ordinary proper names, definite descriptions, etc.) are imposters. They appear to be in the business of referring, but that is not what they are doing at all. The grammatical structure of sentences containing such terms differs from their true **logical structure**. The point of Russell's theory of descriptions is to exhibit their logical structure.

In the case of some definite descriptions (i.e., phrases of the form 'the so-and-so') it is clear that grammatical structure and logical structure come apart. Consider:

- (1) The average family has 2.3 children.

This is grammatically subject-predicate (of the form 'Fa', where 'a' is the subject term and 'F' the predicate). Yet the subject term 'the average family' is a dummy singular term. Its function is not to refer to some particular family which is then said to have 2.3 children. Anyone who thought that would have misunderstood the sentence entirely. The logical structure of (1) is exhibited by:

- (2) The number of children divided by the number of families = 2.3.

(1) is simply a short way of expressing (2), and (2) is a long-division sum. It is of the form ' $a/b = c$ ', not 'Fa'. Moreover, in (2) there is no term purporting to refer to the average family, and that, for Russell, is a sure sign that the description 'the average family', though grammatically a subject term, is not a

WHAT IS THIS THING CALLED METAPHYSICS?

genuine singular term. (For Russell, genuine reference is ineliminable.) Though meaningful, its function is not to refer to anything.

This is enough to show that Meinong's reasoning is faulty. But Russell wanted to show that Meinong's reasoning fails quite generally, and not just in the case of special descriptions such as 'the average family'. Most people understand the conventions underlying constructions of the form 'the average so-and-so is F', and are unlikely to be fooled into thinking that the contained subject term is a real referring term. But in other cases, people might well be fooled. Hence, Russell proposed his theory to show that no description (definite or indefinite) is a genuine referring term. The grammatical structure of sentences containing descriptions is not at all their logical structure.

According to Russell's theory of descriptions, descriptions are disguised existential quantifiers. When they appear in sentences of the form 'An F is G' and 'The F is G', the descriptions, though grammatical subject terms, are not genuine singular terms. The real logical structure of 'An F is G' is captured by $\exists x(Fx \text{ and } Gx)$ (i.e., something is both F and G). The real logical structure of 'The F is G' is captured by $\exists x(Fx \text{ and } (y)(\text{if } Fy \text{ then } x = y) \ \& \ Gx)$ (i.e., there is an x which is F, and uniquely so, and x is G).

In the resulting analyses there occur no singular referring terms, only quantification (the existential quantifier 'there is', and the universal quantifier 'for all'), predication ('F' and 'G'), and identity ('='). Nothing corresponds to 'an F' or 'the F' in the analyses. Thus what appear to be referring terms turn out, under analysis, to function logically as quantifiers, and quantifiers are not referring terms. If I say 'there is a bald man in the room', I am not referring to any bald man, though what I say is true only if someone (anyone) in the room satisfies the description. Quantified sentences are satisfied, or not, by objects. If Bill is the world's only bald man, Bill's being bald makes true 'someone is bald'. But if Fred had been the world's only bald man instead, then Fred's being bald would have made true 'someone is bald'. In contrast, a sentence containing a genuine singular term is made true or false only by the states and doings of the object of reference. How things are with other objects is irrelevant.

In addition to his theory of descriptions, Russell had a further and more controversial thesis: ordinary proper names are disguised descriptions and, hence, disguised quantifiers. Despite appearances, ordinary proper names are not genuine singular terms or referring terms. The only genuine singular terms are Russell's logically proper names, referring to infallibly known current mental items. Thus the grammatical category of subject terms subdivides into the (few) genuine referring terms, and all the rest which are ultimately cashed out as quantifiers.

This is not the place to adjudicate the philosophy of language that Russell endorsed in 1905. His theory of descriptions is still widely accepted, though



not uncontroversial. His theory that ordinary names are disguised descriptions is widely rejected (largely thanks to the work of Saul Kripke).¹² The point of discussing Russell's work here is twofold. First, it shows how we can resist the assumption that to each meaningful subject term there must correspond an object, thus avoiding Meinongian ontological extravagance. Second, in claiming that ordinary names and descriptions are disguised quantifiers, Russell distinguishes surface grammar from logical structure. In doing so, he opens the door to the possibility that 'exists', as this word occurs in, e.g., 'George Bush exists', 'Superman does not exist' and 'tigers exist', though grammatically a predicate, is in fact a disguised existential quantifier.

WHAT IS EXISTENCE?

In asking 'What is existence?' our aim is to uncover the nature of existence, but our method will be logico-linguistic. Thus consider a true positive existential sentence, such as 'George Bush exists'. This is grammatically subject-predicate. One view – the property view – takes this at face value: the surface grammar is the real grammar. The logical form of 'George Bush exists' is as it appears: Fa . 'George Bush' refers to George Bush, the forty-third US president; 'exists' refers to the property of existence. George Bush has the property of existence. Hence, existence is a property of ordinary objects, alongside properties such as weight, height, baldness, etc.

On a different view – the quantifier view – the surface grammar is misleading: 'exists' is not really a predicate but a quantifier. The logical form of 'George Bush exists' is not Fa but $\exists x(x = \text{George Bush})$ (i.e., there is an x such that x is identical to George Bush). Hence, existence is not a property of George Bush or of any other ordinary object.¹³

Meinong and McGinn, who believe in non-existent objects, also defend the property view. This is not a coincidence. If a property is genuine, it is natural to suppose that some objects possess it and others lack it. If existence is a genuine property, it follows that some objects exist and some do not. The latter are the non-existent objects. Thus the view that existence is a property, like the liberal view of meaning and reference criticized in the previous section, dovetails with the doctrine of non-existent objects.

How are we to decide between the property and quantifier views? A. J. Ayer once wrote that, if the property view were true, 'it would follow that all positive existential propositions were tautologies, and all negative existential propositions self-contradictory.'¹⁴ It's not clear that Ayer's first complaint hits its target. Simply to hold that existence is a property does not seem to imply that, e.g., 'George Bush exists' is a tautology. Certainly, if one assumed that every meaningful name refers to an object, and that there are no non-existent objects, then

WHAT IS THIS THING CALLED METAPHYSICS?

indeed the meaningfulness of the name 'George Bush' ensures that an utterance of 'George Bush exists' could not fail to be true. But the doctrine that existence is a property plays no role here.

There is more force to Ayer's second complaint that, on the property view, negative existential propositions are contradictory. Consider the true sentence 'Superman does not exist'. On the property view, 'Superman exists' has the logical form 'Fa' ('a is F'); hence 'Superman does not exist' has the same form, only with ' \sim F' (not-F) in place of 'F': ' $(\sim F)a$ ' ('a is \sim F'). Any sentence of that form is subject to the rule of existential generalization: from $(\sim F)a$ it follows that $\exists x (\sim Fx)$ (i.e., there exists an x which is not-F). Thus, on the property view, from the truth of 'Superman does not exist' it follows that there exists something which does not exist. And this, as Ayer rightly says, is contradictory.

Of course, there is one way in which contradiction can be avoided. If there are non-existent objects, then the rule of existential generalization fails. According to friends of the non-existent, from the truth of 'Fa' we can only infer 'something is F', and 'something is F' does not imply 'there exists an object which is F'. Hence from 'Superman does not exist', we are entitled to infer only 'there is something which does not exist'. For defenders of the non-existent, this is not a contradiction: it is a truth! Many things lack the property of existence, and Superman is one of them.

However, the thesis that there are non-existent objects is hard to accept. First, it just seems silly to believe that there is a realm of objects that do not exist. Second, defenders of the non-existent must hold that the English expressions 'some' and 'there are' are not existence-implying. That is, 'some Fs are G' and 'there are Fs which are G' are taken not to imply that there exist Fs which are G. This is hard to believe. Third, the assumption that all names refer (to either existent or non-existent objects) is especially implausible in the case of a name such as 'Vulcan'. The astronomer Jean Leverrier introduced the name 'Vulcan' to refer to a planet between Mercury and the Sun, the presence of which would explain certain astronomical observations. It turned out that there was no such planet. 'Vulcan' is a paradigm case of reference failure, not a case of successful reference to a non-existent object.

However, even if Ayer's second complaint against the property view stands, the quantifier view fares no better. On the quantifier view, the sentence 'Superman does not exist' is rendered as ' $\sim \exists x (x = \text{Superman})$ ' (i.e., it's not the case that there is an x such that x is identical to Superman). Unfortunately, by the rule of existential generalization, we can infer: $\exists y \sim \exists x (x = y)$, which is a contradiction. What this shows is that the so-called problem of negative existentials is a problem for everyone.¹⁵ How can we acknowledge the truth of 'Superman does not exist' without either contradicting ourselves or embracing the non-existent?

Is there some other way of deciding between the property and quantifier views? Two considerations tell in favour of the quantifier view. First, given that we have rejected the doctrine of non-existent objects, what work does the alleged property of existence do? In the absence of the non-existent, it is a property of everything. Nothing lacks the property of existence. Yet it is a purely formal property and does no causal work. The weight, mass and velocity of a car are causally efficacious properties, but its existence is not (although, trivially, it would not have any properties unless it existed). The postulation of such a property is otiose.

Second, even defenders of the property view concede that the sentences ‘George Bush exists’ and ‘ $\exists x(x = \text{George Bush})$ ’ (necessarily) have the same truth value (they just think that the latter sentence does not display the logical structure of the former). So what is gained by, in addition, postulating existence as one of Bush’s properties? As noted, it does no explanatory work. Why not embrace the quantificational analysis on grounds of economy?¹⁶

On balance, then, the quantifier view is more plausible than the property view. And in rejecting the property view we should also reject its more egregious kin: the doctrine of non-existent objects and the view that all meaningful names and descriptions refer.

CONCLUDING REMARKS

In this chapter we have been kinder to the doctrine of modal realism than we have to the doctrine of non-existent objects. Some may find this strange: are both doctrines not equally odd and equally extravagant? Perhaps not. Lewis’s modal realism has things to be said in its favour. It accounts for the truth of our modal claims and it answers otherwise puzzling questions about why our world exists. But, apart from providing non-existent objects as truth-makers for sentences such as ‘Superman does not exist’, there is nothing to be said in favour of the doctrine of non-existent objects.

In addition, we questioned two of the motivations for believing in non-existent objects: the doctrine that meaning implies reference and the thesis that existence is a property of ordinary objects. Russell showed us the way to deny the first, and the considerations rehearsed in the previous section give us the grounds to deny the second.

WHAT IS THIS THING CALLED METAPHYSICS?

STUDY QUESTIONS

- Is modal realism a tenable position?
- Should we be puzzled that our universe exists?
- Should we countenance non-existent objects?
- Can you think of any objections to Russell's theory of descriptions?
- Is existence a property of Fido, alongside loyalty and friendliness?

ANNOTATED FURTHER READING

- S. Kripke, *Naming and Necessity* (Oxford: Basil Blackwell) 1980. One of the most brilliant and influential works of contemporary philosophy. Kripke offered a philosophy of language congenial to essentialism and, in doing so, rehabilitated metaphysical inquiry as central to philosophy.
- D. Lewis, *On the Plurality of Worlds* (Oxford: Basil Blackwell) 1986. Classic exposition and defence of modal realism.
- C. McGinn, *Logical Properties* (Oxford: Oxford University Press) 2000. McGinn's chapter 'Existence' provides a clear statement of a contemporary Meinongian position.

INTERNET RESOURCES

- P. Mackie (1998) 'Existence', in E. Craig (ed.), *Routledge Encyclopedia of Philosophy*. Retrieved 31 May 2006 from <<http://www.rep.routledge.com/article/X013>>.
- B. Miller (2002) 'Existence', *The Stanford Encyclopedia of Philosophy (Summer 2002 edition)*, Edward N. Zalta (ed.). Retrieved 31 May 2006 from <<http://plato.stanford.edu/archives/sum2002/entries/existence>>.
- L. N. Oaklander (2005) 'Negative Facts', in E. Craig (ed.), *Routledge Encyclopedia of Philosophy*. Retrieved 31 May 2006 from <<http://www.rep.routledge.com/article/N118>>.



UNIVERSALS AND PARTICULARS

• Introduction	37
• Property realism	38
• Objections to traditional property realism	41
• Russell's bundle theory	42
• Varieties of nominalism	43
<i>Predicate nominalism</i>	44
<i>Class nominalism</i>	44
<i>Resemblance nominalism</i>	45
<i>Tropes</i>	46
• Concluding remarks	48

INTRODUCTION

The problem of universals is one of the most venerable in metaphysics, dating back to Plato and Aristotle. At root, the problem concerns the nature of properties and relations. Are properties and relations universals, identical in their instances (as Plato and Aristotle thought), or can we explain, e.g., what it is for a sphere to be red, and hence what it is for two spheres to be the same colour, without appeal to universals?

Philosophers who hold that we must appeal to universals in order to explain the nature of properties and relations are traditionally called 'property realists'; those who deny this are traditionally called 'nominalists'. But there are varieties of each position: property realism has Platonic, Aristotelian and Russellian versions, while nominalism divides into predicate nominalism, class nominalism, resemblance nominalism and trope theory.

It would not be possible to discuss the problem of universals without also discussing the nature of that which has properties and stands in relations: particulars, objects or individual **substances** (as they are variously known). Traditional property realists about universals endorse the substance–attribute (particular–universal) ontology, and are thus committed to both categories of being. However, the notion of an individual substance is called into question by modern property realists (such as Bertrand Russell) and by some nominalists (such as the trope theorist, D. C. Williams).

WHAT IS THIS THING CALLED METAPHYSICS?

Universals

Since Plato, many philosophers have held that properties and relations are universals, 'wholly present' in their instances. If my billiard ball is red, that is because redness (the universal) 'inheres' in the ball (the particular). Similarly, if any other ball is red that is because the very same universal 'inheres' in that ball too. If a red ball is 1 foot from a white ball that is because *being 1 foot from* (the universal) inheres in the spatial distance between them. Some relations (the so-called internal relations, such as *being taller than* or *being redder than*) lack real being. That Fred is taller than Bill is necessitated by the fact that Fred is 6 foot tall and Bill is 5 foot tall. As long as we record the particular heights of Fred and Bill in the inventory of being, we do not need to add that Fred is taller than Bill. In contrast, external relations, such as *being to the left of* or *being one foot apart*, are not necessitated by their terms, and are genuine additions to being. On some views, internal relations are not universals (since not fundamental).

PROPERTY REALISM

We should first note two quite different conceptions of universals: Platonic and Aristotelian. On the Platonic view, universals are transcendent. That is, they exist outside space and time. They are changeless **abstract** objects. On the Aristotelian view, universals are immanent. They do not exist outside space

Plato (427–347 BC)

The founder of modern philosophy, Plato was born into an aristocratic family in Athens, and became a follower of Socrates. After Socrates' death in 399 BC, Plato began to write Socratic dialogues in memory of his teacher. A number of years later Plato founded the Academy, the first university-style institution in the West, which provided sanctuary to leading mathematicians, scientists and philosophers. Plato's two central doctrines were his theory of forms and his theory of the immortality of the soul. In contrast to the ephemeral, changing world we see around us, the forms are abstract and changeless, the true objects of knowledge. Thus the form of justice exists in an abstract realm, outside space and time. Individual human acts are just in virtue of 'participating' in the form Justice. We all had knowledge of the forms prior to our current physical incarnation. The forms are the foundation of Plato's metaphysics, but also play a key role in his political philosophy. In Plato's ideal city (outlined in his most famous work, *The Republic*) it is the philosophers who rule since they alone understand the forms, in particular the form Good.



Aristotle (384–322 BC)

Born in the Macedonian city of Stagira, Aristotle spent much of his later life in Athens. He enrolled in Plato's Academy in 367 BC, and remained there for twenty years, first as a student, then as a teacher and writer. He left Athens after Plato's death (347 BC) but later returned to set up his own school, the Lyceum. Aristotle was the most systematic philosopher of antiquity. He originated the conception of philosophy as an intellectual inquiry, divided up into distinct branches: logic, scientific inquiry (largely, biology and cosmology), metaphysics, the mind–body problem, ethics and politics, and literary criticism. Aristotle made original contributions to all these areas and, in doing so, helped to define what we now think of as philosophy. He was much discussed by the medievals after translations of his works into Latin appeared in the twelfth century. His influence was further reinforced, and legitimated, when Aquinas made Aristotelianism the basis for Catholic theology.

and time. They are located where their instances are located and nowhere else. Since the existence of abstract entities, outside space and time, does not depend on the existence of concrete entities in space and time, a consequence of the transcendent conception is that universals can exist uninstantiated (i.e., without at any time having instances). Thus the universal 'unicorn' exists, on the Platonic view, even though there are no unicorns. In contrast, on the Aristotelian view, universals cannot exist uninstantiated.

Plato was interested in mathematics and geometry. Ideal geometric shapes were the model for his universals (or 'forms' as he called them). Since, e.g., Pythagoras' theorem is true whether or not any physical item is exactly triangular, so universals exist whether or not they have any concrete instances. In contrast, Aristotle's model for universals was drawn from that of species in biology. If it makes no sense to think that a species might exist which never has any members, then it will seem to make no sense that a universal might exist without having any instances.

The essence of traditional property realism is reasonably clear: objects (particulars or 'individual substances' in the more traditional vocabulary) have, i.e., instantiate, properties; two (or more) objects can, quite literally, have one and the same property; hence properties are universals, which can be wholly present in two or more places at the same time.¹

One motivation for property realism, and certainly one of Plato's motivations, stems from considerations to do with the meaning of general terms. Suppose we begin by assuming a referential theory of meaning, according to which the meaning of a word is an entity, the entity referred to by the word. In the case of singular terms, e.g., ordinary proper names such as 'Socrates',

WHAT IS THIS THING CALLED METAPHYSICS?

'Red Rum' and 'Edinburgh' there is little difficulty identifying the entities which, according to the referential theory, are the meanings of these words: the man Socrates, the horse Red Rum and the city Edinburgh, respectively. But what of general terms such as 'horse' and 'city' (i.e., terms which apply to many entities)? What entities do these words refer to? They do not refer to one particular horse or to one particular city (for why that horse or that city?). That is why they are called general terms: they apply quite generally to all horses and all cities, respectively. Hence, thought Plato, if general terms don't name particulars, they must name universals (such as *cityhood* and *horseness*).

Few philosophers now accept this semantic argument for universals. First, the argument assumes that, if 'horse' refers, it refers either to a particular horse or to the universal *horseness*. But this assumption can be questioned: why not see 'horse' as referring to each horse? A defender of the semantic argument must rule this out if his argument is to succeed. Second, and more fundamentally, why should we accept the underlying assumption that the meaning of a word is some entity the word stands for? (See Chapter 2.) Names such as 'Santa Claus' and 'Odysseus', for example, are perfectly meaningful, yet there is no one to whom they refer (they are empty names). Rather than think of meaning as reference, maybe we do better to think of the meaning of a word as a function of its uses within some linguistic community. On such a view, the referent of a word, if it has one, is irrelevant to its meaning.

However, there is another, metaphysical, argument for universals. It runs as follows:

Consider two exactly similar red spheres. They have the same colour (amongst other similarities). That is, the colour of one sphere is literally the same as, i.e., **numerically identical** to, the colour in the other. What is present in one is also present in the other. Particulars cannot be (wholly) in two places at the same time, but universals can. No particular (such as a man or a horse) can be wholly present in two places at once. Of course, one part of a horse (e.g., its left front leg) can be, and will be, in a different place from another part (e.g., its right front leg). But the horse cannot be, in its entirety, in two places at once. In contrast, the essence of a universal is its repeatability: it can be wholly present in different places at the same time. Thus, in order to explain the truism that different objects can have the same property, wholly present in each object, we must appeal to universals.

The problem with this argument is that not every use of 'same' expresses strict numerical identity. Some uses do: e.g., when we say of two children that they 'have the same mother'. In such a case, we really do mean that the mother of one child is literally, numerically, the same as the mother of the other. But consider the use of 'same' in 'he has the same eyes as his father'. Here the



word ‘same’ expresses **qualitative identity** (i.e., striking similarity), not numerical identity. The sense of ‘same’ in ‘they have the same mother’ is quite different from its sense in ‘they have the same eyes’.

This observation has the potential to frustrate the metaphysical argument for universals. From the fact that X is the same F as Y it does not follow that there is some property of X which is numerically identical to (as opposed to merely similar to) some property of Y. The defender of the metaphysical argument will need to make out a case that the use of ‘same F’ in question expresses numerical identity. Such a case may be plausible with regard to substantive general terms (such as ‘man’ or ‘horse’), but less plausible with regard to qualitative general terms (such as ‘red’ or ‘round’).

OBJECTIONS TO TRADITIONAL PROPERTY REALISM

Neither the semantic nor metaphysical argument for universals is compelling. Is there a positive case against universals? One objection to traditional property realism is that it offers no account of the connection between a particular (e.g., Socrates) and his properties. A particular is said to instantiate various universals. But what is instantiation? Is it a relation? Is it primitive and unanalysable? Should we not find it mysterious?²

Second, explaining commonality in terms of instantiation of the same universal leads to an infinite regress. Aristotle pressed this objection against the Platonic, transcendent conception of universals. His ‘third man argument’ purports to demonstrate that Plato’s theory of forms leads to an infinite regress. (This argument can be found in Plato’s *Parmenides*, but Aristotle’s rendering is more familiar.)

Proceeding on the assumption that the form of F is itself an F (the self-predication assumption), the third man argument runs as follows. Plato attempts to explain what all individual men have in common, that in virtue of which they fall under the general term ‘man’, by citing a relation (‘participation’ or ‘imitation’) that each man stands in to the form Man. But if the form Man itself falls under the general term ‘man’, then we will need to postulate a further form Man1 in order to explain what all individual men and the form Man have in common. But if form Man1 is also a man then we will need to postulate a third form Man2 in order to explain why individual men, form Man and form Man1 are all men. And so on, *ad infinitum*. The regress is vicious since in order to explain commonalities at any one level we are forced up to the next level, and thus no genuine explanation is ever achieved.

Although Aristotle’s third man argument was directed against Plato’s theory of forms, many have thought that it yields an argument against all theories of universals. Even theories of universals, such as Aristotle’s, which do not make the self-predication assumption and conceive of universals as this-worldly

WHAT IS THIS THING CALLED METAPHYSICS?

are vulnerable to the following regress argument. Whenever two or more particulars instantiate a universal, those instantiations will themselves be instances of the universal Instantiation. What explains the commonality between different instances of Instantiation must be the presence of a further universal Instantiation¹. What explains the commonality between different instances of Instantiation¹ must be the presence of a further universal Instantiation². So on, *ad infinitum*, given only the realist's assumption that any commonality is to be explained as the instantiation of a universal.

We can present the regress slightly less gnomically as follows. Where A and B have some commonality (say, they are both F), the realist postulates a universal (Fness) which they both instantiate. But now the facts or states of affairs of A's instantiating Fness and B's instantiating Fness have a commonality (they are both cases of instantiation); so we must postulate a universal (Instantiation) which both states of affairs instantiate.³ But now the two complex states of affairs of {[A's instantiating Fness] instantiating Instantiation} and {[B's instantiating Fness] instantiating Instantiation} have a commonality; so we must postulate a further universal (Instantiation¹) which both states of affairs instantiate. And so on.

Third, there is a problem concerning that which has properties or instantiates universals (the particular or individual substance). The traditional property realist endorses the standard substance–attribute ontology (corresponding to the subject–predicate distinction in grammar). A substance (such as a particular horse) is not identical with its properties, but *has* those properties. A substance is not just the sum of its properties; it is something over and above, or under and beneath, its properties. But this raises an immediate epistemic worry. Since all we perceive and respond to are properties, how can we be acquainted with a substance?

RUSSELL'S BUNDLE THEORY

Bertrand Russell (1872–1970) is a property realist who rejects the traditional notion of substance and regards ordinary objects as bundles of universals. He thus avoids the standard objections to traditional property realism just rehearsed. For example, there can be no puzzle concerning the tie between substance and universal if there are no substances (as standardly conceived).

Why was Russell dissatisfied with the notion of substance? A substance or particular, he wrote,

cannot be defined or recognised or known; it is something serving the merely grammatical purpose of providing the subject in a subject-predicate sentence such as 'This is red'. And to allow grammar to dictate our metaphysics is now generally recognised to be dangerous.



... The notion of a substance as a peg on which to hang predicates is repugnant.⁴

Russell's response was to eliminate the category of substance and to conceive of ordinary objects as bundles of universals. Having dispensed with substances, he needed some account of what unifies and separates distinct objects. To explain the unity of a single object, he appealed to a notion of 'compresence', defined initially for the case of experiences. If I see a bird fly past and simultaneously hear a bell tolling, the two experiences are compresent (had by the same subject). More generally, the qualities that make up an object, such as my chair, are compresent with each other, such that no other quality in the universe is compresent with all those qualities. These qualities, for Russell, are universals: the quality of brownness in my carpet is strictly identical to the quality of brownness in my desk.

However, Russell's theory faces the following two problems. First, 'compresence' is only a label. We are not told how various universals manage to combine, without inhering in a substance, to form an object, such as a particular horse. Second, what makes objects distinct from each other, on Russell's view? As Russell points out, although men have some qualities in common (such as humanity), no two men are exactly alike. There is always some difference between two men, however trivial (e.g., differences in the number of hairs on their heads, or in the number of molecules composing their left hands, etc.). As Russell says, '[i]t is only the assemblage of qualities that makes the instance unique.'⁵ Thus, Russell can claim, whenever two individuals are distinct, there will be some difference in the corresponding bundles of universals in virtue of which they are distinct.

However, this claim merely invites the following objection. Even if no two objects are exactly alike, is it not possible for there to be two objects exactly alike? For example, could there not be a world containing only two exactly resembling spheres? Yet for Russell there cannot be such a world. If objects are bundles of universals, then identity of bundle implies identity of objects. If the 'two' spheres are composed of the same universals, they cannot be two in number. 'They' are one. But, intuitively, it is possible for there to be two exactly similar spheres. Russell's theory rules out this possibility, and this is an objection to his theory.

VARIETIES OF NOMINALISM

We have seen that there are problems facing traditional property realism and Russellian bundle theory. Are the prospects for nominalism any brighter? Let us examine the range of views that attempt to account for objects and their properties without appeal to universals.

Predicate nominalism

According to predicate nominalism, for x to be F is for the predicate 'F' to apply to, or be true of, x . What makes it true that x and y are both F is that 'F' applies to both x and y . Thus, what makes it true that a sphere is red is simply that the predicate 'red' applies to that sphere.

This theory is open to serious objections. First, there are surely properties in the universe that we will never discover, and for which there exists no natural language predicate. Predicate nominalism seems constitutionally incapable of acknowledging the possibility of such properties. Second, and more fundamentally, predicate nominalism seems to put the cart before the horse. Intuitively, x 's being F does not consist in 'F' applying to x ; rather, 'F' applies to x because x is F . If x is F (say, a mountain is a certain height), x would still have been F even if no languages had existed. (These objections also apply to concept nominalism: the view that x is F in virtue of the fact that x falls under the concept of F .)

Class nominalism

According to class nominalism, for x to be F is for x to belong to the class of F s. What makes it true that x and y are both F is that x and y are members of the class of F s. Thus, what it is for a sphere to be red is for the sphere to be a member of the class of red things.

Class nominalism avoids the first objection to predicate nominalism. Whether an object is a member of a certain class does not depend on whether we have discovered the class or have a word for it. (The class of electrons existed before we knew of it.) However, the second objection still seems to apply: intuitively x is a member of the class of F s because x is F , not *vice versa*.

In addition, class nominalism incurs two objections of its own. First, the relation of class membership is itself a universal, instantiated whenever an object is a member of a class. Hence, class nominalism is implicitly committed to universals, and so fails to be a genuine version of nominalism. Second, suppose that all and only F s are G s (e.g., suppose it were the case that all and only red things were round). Since the class of F s is the class of G s, it follows, on class nominalism, that the property of being F is the property of being G – an absurd result.

A related problem arises for any pair of empty general terms, such as 'unicorn' and 'dragon'. Applying the class nominalist strategy, we get: x is a unicorn if x belongs to the class of unicorns, and y is a dragon if y belongs to the class of dragons. However, the class of dragons is the class of unicorns (that is, the class with no members, the null class). In which case, the class



nominalist is committed to the absurdity that the property of being a unicorn is the same as the property of being a dragon. (This is a variant of the second objection since, if there are no Fs and no Gs, then trivially all and only Fs are Gs.)

Resemblance nominalism

According to resemblance nominalism, for x to be F is for x to be a member of a class of objects which resemble each other, where 'resemblance' is treated as a primitive, unanalysable relation. What makes it true that x and y are both F is that they resemble each other. A sphere is red because it is a member of a class of resembling objects.

However, there are three problems with this version of nominalism. First, an object such as a red sphere is a member of a number of resembling classes. For example, it is a member of the class of red things and a member of the class of spherical things. Yet, clearly, the sphere is not red in virtue of resembling spherical things; it is red in virtue of resembling red things. That is, it is red because it resembles other objects in respect of colour (rather than shape). In which case resemblance is no longer an unanalysable relation, but admits of respects. Moreover, what are respects if not universals?

Russell had a similar but more fundamental objection: the relation of resemblance is itself a universal. He wrote:

If we wish to avoid the universals *whiteness* and *triangularity*, we shall choose some particular patch of white or some particular triangle, and say that anything is white or a triangle if it has the right sort of resemblance to our chosen particular. But then the resemblance required will have to be a universal. Since there are many white things, the resemblance must hold between many pairs of particular white things; and this is the characteristic of a universal.⁶

In which case, as with class nominalism, resemblance nominalism has not avoided an implicit commitment to universals.

Finally, there is a problem with uniquely instantiated general terms. According to resemblance nominalism, an object is spherical if it resembles other objects in respect of shape. What if the universe had contained a single sphere and nothing else? Surely it would have been spherical? But the resemblance nominalist must deny this since there is nothing in that world for the sphere to resemble. This is surely an unpalatable consequence.

Tropes

Trope nominalism is the most interesting and most extreme version of nominalism. All other versions of nominalism agree with property realism about which things are particulars; they only disagree about the need to postulate further entities, universals. Trope theory disagrees with both moderate nominalism and property realism about which things are particulars: properties and relations are conceived of as particulars. Tropes (i.e., properties and relations conceived of as particulars) have been called ‘abstract particulars’ in contrast with more familiar particulars (ordinary concrete objects).⁷ Thus you, me and the Eiffel Tower are concrete particulars; your smile, the colour of my eyes, and the height of the Eiffel Tower are abstract particulars.

Put forward by G. F. Stout in the nineteenth century, trope theory was substantially developed and defended by Donald Williams and Keith Campbell in the twentieth century. For Williams, tropes are the basic items in the universe, the ‘alphabet of being’. ‘Any possible world, and hence, of course, this one, is completely constituted by its tropes and their connections of location and similarity.’⁸ According to trope theory, properties and relations are abstract particulars, not universals. Thus suppose we have two red billiard balls before us. They are exactly similar. In particular, they are both red, but not because they share a common property. Rather, the redness of one ball is an (unrepeatable) particular, numerically distinct from, but exactly resembling, the redness of the other ball. Each redness is a distinct abstract particular (a trope).

Tropes

Trope theory is a radical version of nominalism. Like all versions of nominalism, it denies the existence of universals. In addition, trope theory holds the extreme view that properties and relations are themselves particulars – abstract particulars. Thus the redness of a particular billiard ball is an abstract particular, located where the ball is and nowhere else. A different but exactly resembling billiard ball has a numerically different but exactly resembling redness trope. There is no colour property common to, or instantiated in, both balls. Similarly with all other properties and relations of the balls. On a standard version of trope theory, ordinary particulars (i.e., what all theories classify as particulars – you, me, the next winner of the Melbourne Cup, etc.) are composed of abstract particulars. A billiard ball is composed of a colour trope, a shape trope, a weight trope, etc. The universe consists of tropes. They are the ‘alphabet of being’ out of which all else is constructed. Questions remain for trope theory. Could a redness trope exist all by itself? If not, why not? What binds tropes together to form single objects? Are some tropes basic, and others reducible or eliminable?



Ordinarily we contrast a concrete particular with its properties: e.g., we contrast a man (Bill) with his baldness. We don't think of Bill's baldness as itself a particular, but that is what trope theory asserts. Others may be as tall as Bill, but they have their own height trope numerically distinct from Bill's height trope. Relations are also tropes: Bill's being a foot taller than Fred, and Anne's being a foot taller than Mary, are not the same relation: they are two exactly similar relation tropes.

On standard trope theory, ordinary concrete objects are not substances (as normally understood) but bundles of tropes. What is the relation between a concrete object and the tropes that constitute it? If we are to respect the common sense intuition that some properties of a concrete particular are accidental, the relation of a concrete object to its tropes cannot be that of set to member. Since baldness is one of Bill's **accidental** properties, we should not think of Bill's baldness as a member of the set of tropes which constitute Bill. If we did, then it would follow, since a set has its members essentially, that Bill is essentially bald (which he's not). In contrast, other properties of Bill may be **essential** to him (e.g., his humanity). Perhaps it is impossible for Bill to exist and be other than a human being. If so, Bill's humanity trope is essential to the bundle that is Bill.

The relation between an ordinary object and its tropes must presumably be some relation of whole to part, not set to member. Some parts are accidental, others essential. The mysterious 'is' of instantiation is thus replaced by the 'is' of 'is a part of'. The sentence 'Bill is bald' is true, not because Bill instantiates the universal *baldness*, but because a baldness trope is a part of the bundle of tropes that is Bill.

The great merit of trope theory is that it avoids many of the objections canvassed so far. It escapes the objections to other versions of nominalism since it doesn't analyse what it is for an object to have a property in terms of any relation between that object and something else (predicates, classes, classes of resembling objects, etc.). If a sphere is red that is because it contains a red trope, and this is not a matter of the sphere standing in any relation to anything else. Although trope theory helps itself to the notion of similarity or resemblance, it does not conceive of this relation, or any other relation, as a universal. A's resembling B and C's resembling D are distinct resemblance tropes, not a doubly instantiated universal. Thus, the problems we raised for traditional property realism do not arise on trope theory. For example, there is no regress worry since commonality is not explained in terms of instantiation, but in terms of resemblance.

However, there are problems for trope theory. First, trope theorists replace the instantiation relation with the part-whole relation. But what is this relation? We have been told nothing about it, other than that some parts (tropes) are essential to a bundle, others accidental. Rather than clarify matters, this raises another puzzle: how can a trope be essential to a bundle? What makes it essential? Is a bundle the sort of thing that can have essential parts?

WHAT IS THIS THING CALLED METAPHYSICS?

Second, even if there is no problem of instantiation for the trope theorist, there is the problem of what binds tropes together to form a single object. (A different kind of bundle theorist, Russell, had an analogous problem.) A bundle of tropes may be ‘co-located’, but what is the glue that unifies them? Is this a question for science or for metaphysics? Can it be answered at all?

Third, the strategy of the trope theorist is to exchange the instantiation relation for the part–whole relation. This works well enough in the case of predications involving concrete particulars, but what about predications involving tropes? We can truly say many things of a given redness trope: it is red, coloured, red or green, my favourite trope, in Canberra, persisted through 2006, etc. Since a trope is not a bundle, we cannot regard ‘this trope is red’ as made true by a bundle’s containing a redness trope. How then are we to understand these predications other than in terms of the instantiation of a property in a trope? Are we not then back with universals?

There are problems for trope theory, as there are for all other theories. Nonetheless, it is a novel account of the nature of objects and properties, and well worth further research.

CONCLUDING REMARKS

We have covered a lot of difficult ground in this chapter and our conclusions are the following. There are serious problems facing both traditional and Russellian property realism and serious problems facing nominalist accounts, both moderate and extreme. Trope theory may be the most promising, or least unsatisfactory, account of properties, though that project still has much in the way of unfinished business.

STUDY QUESTIONS

- What is the difference between Platonic and Aristotelian conceptions of universals?
- What is the distinction between numerical and qualitative identity?
- How might a defender of universals block the regress of instantiation?
- Is any version of nominalism defensible?
- What is a trope?



ANNOTATED FURTHER READING

- D. M. Armstrong, *Universals: An Opinionated Introduction* (Boulder, Col.: Westview Press) 1989. By far the best introduction to universals, written by one of their stoutest defenders. Armstrong also appeals to universals to explicate laws of nature: it is a law of nature that all Fs are Gs because the universals F and G are necessarily connected.
- K. Campbell, *Abstract Particulars* (Oxford: Basil Blackwell) 1990. A clear and trenchant defence of tropes.
- B. Russell, *The Problems of Philosophy* (Oxford: Oxford University Press) 1978, Chapters 9 and 10. Written in 1912, this book remains one of the classic introductions to philosophy. In Chapter 9 Russell argues for the existence of universals (construed transcendently) and offers a famous objection to resemblance nominalism. In Chapter 10 he considers how we know universals. He holds that sensible qualities and relations can be known directly, by acquaintance, and that all *a priori* knowledge is knowledge of relations between universals.

INTERNET RESOURCES

- J. C. Bigelow (1998) 'Universals', in E. Craig (ed.), *Routledge Encyclopedia of Philosophy*. Retrieved 31 May 2006 from <<http://www.rep.routledge.com/article/N065>>.
- C. Daly (2005) 'Properties', in E. Craig (ed.), *Routledge Encyclopedia of Philosophy*. Retrieved 31 May 2006 from <<http://www.rep.routledge.com/article/N121SECT4>>.
- C. Swyer, 'Properties', *The Stanford Encyclopedia of Philosophy (Winter 2000 edition)*, Edward N. Zalta (ed.). Retrieved 31 May 2006 from <<http://plato.stanford.edu/archives/win2000/entries/properties>>.

4 CAUSATION

• Introduction	53
• Preliminary questions	54
• The constitutive question: what is causation?	57
• Hume	58
• The regularity theory	61
• The counterfactual theory	64
• A simpler theory	65
• Concluding remarks	66

INTRODUCTION

Our language abounds with causal talk. We often use the word ‘cause’ itself – ‘smoking causes cancer’, ‘her words caused offence’, etc. – but we also use many verbs which presuppose causation: ‘they had to push their car’, ‘she pulled him towards her’, etc. Such talk raises a host of different questions, about which there is little consensus. The biggest question of all concerns the nature of causation itself. The modern discussion of this question was inaugurated by David Hume (1711–76). But there are preliminary questions about causation which should be addressed first and which will help set the scene for discussion of the more familiar constitutive question about the nature of causation.

WHAT IS THIS THING CALLED METAPHYSICS?

David Hume (1711–76)

Born in Berwickshire, Scotland, Hume is generally regarded as the greatest ever English-speaking philosopher. He was also a noted historian and essayist. Following in the tradition of Locke and Berkeley, Hume was an empiricist, a naturalist and a sceptic. His major philosophical works – *A Treatise of Human Nature* (1739–40), *Enquiry Concerning Human Understanding* (1748), *Enquiry Concerning the Principles of Morals* (1751), as well as the posthumously published *Dialogues Concerning Natural Religion* (1779) – were hugely influential, though criticized at the time as works of scepticism and atheism. Hume's eventual fame and fortune was largely due to the publication of his six-volume *History of England* (1754–62). His reputation for atheism prevented his election to chairs in Edinburgh (1745) and Glasgow (1752). He never held an academic post. An agreeable and clubbable man, Hume was friends with the leading intellectuals of his time. He could count Adam Smith, James Boswell, Denis Diderot, and Jean-Jacques Rousseau among his friends and acquaintances.

PRELIMINARY QUESTIONS

Here are six preliminary questions:

- (1) Are there two kinds of causation – singular and general – or is there one kind of causation holding between tokens and between types?
- (2) What are the *relata* of the causal relation? That is, when we say something of the form 'A caused B', what kinds of entity are A and B?
- (3) What is the logic of the causal relation?
- (4) What is the temporal direction of causation?
- (5) With what right do we describe one cause, from amongst the plethora of background conditions, as 'the cause' of an event?
- (6) What is the link between causation and laws of nature?

(1) It's worth distinguishing **singular causal claims** (her words caused offence) from **general causal claims** (smoking causes cancer). The former relate particular, datable phenomena (tokens), whereas the latter relate types of phenomena. Of course, singular and general causal claims can hardly be unrelated. The fact that smoking causes cancer can hardly be unrelated to the fact that particular sustained bouts of smoking cause particular people to develop cancer. Nonetheless, it is a further question whether in addition to singular and general causal claims there are also singular and general kinds of causation. Whatever



the answer to this question, in this chapter we will be concerned only with singular causation.

(2) What are the *relata* of cases of (singular) causation? Three standard candidates are: objects, events and facts. As to the first option, we do say things such as ‘the car killed the man’, ‘Mary killed Sally’, etc. suggesting that objects (the car and the man, Mary and Sally) are the *relata* of the causal relation ‘killed’. But arguably this is just a loose way of speaking. Strictly speaking, it was the impact of the car that caused the death of the man, and it was Mary’s thrusting the dagger that caused Sally’s death.

This suggests the second option: causes and effects are not objects but events (i.e., happenings or changes in objects). Donald Davidson is the modern philosopher most closely associated with this view.¹ According to this view, any singular causal claim, if it does not explicitly connect events, can always be rewritten so that it does connect events, thereby revealing the true logical structure of the causal claim. Thus ‘the car killed the man’ should be rewritten as:

- (a) the impact of the car caused the death of the man

making it explicit that the *relata* are events.

On the remaining option, facts are the causal *relata*. Whereas events are typically picked out by definite descriptions (‘the death of the man’), facts tend to be denoted by complete sentences (‘the man died’). According to this option, defended recently by Hugh Mellor, any causal sentence not explicitly connecting events, can and should be rewritten in terms of facts.² Thus, ‘the car killed the man’ should be recast as:

- (b) the man died because the car moved

making it explicit that the *relata* are facts.

If we put the first view to one side, which of the remaining views is correct? One argument tells in favour of the event view and against the fact view. Events are changes in concrete objects and thus just the kinds of things to do real causal work; in contrast, facts seem abstract and inert. Another argument tells in favour of the fact view and against the event view. Absences can be causes, and absences are not events. Mellor gives the following example: it can be true to say of Don while rock-climbing ‘he didn’t die because he didn’t fall’.³ That is, the fact that Don didn’t die was caused by the fact that he didn’t fall. If there can be negative facts, but not negative events, this example tells in favour of the fact view.

In this chapter, our examples will be presented in terms of event causation, but that should not be taken as support for the event view.

WHAT IS THIS THING CALLED METAPHYSICS?

(3) Relations can be classified as reflexive, symmetric and transitive.⁴ How does the causal relation fare on these logical dimensions? It is uncontroversial that causation is not reflexive: not every event causes itself. Is it irreflexive? That is, is it impossible for an event to cause itself? Not if **causal loops** are logically possible. If causal loops are logically possible (which they are if travel into the past is possible), then there can be cases where A causes B, and B causes A; so A causes itself. (See Chapter 6.)

Is causation symmetric or asymmetric? It is certainly not symmetric. That is, it's not the case that if A causes B, B causes A. Is it then asymmetric? That is, if A causes B does it follow that B does not cause A? Certainly in our world, if A causes B, B does not cause A. But if causal loops are possible, there can be cases where A causes B and B causes A. If this is a possibility, then causation is neither symmetric nor asymmetric – it is non-symmetric.

Is causation transitive? That is, if A causes B and B causes C does it follow that A causes C? Again, opinions differ. Here are two considerations that tell against transitivity. First, we may be unwilling to accept that the Big Bang causes me to type these words, yet that's what we'd be committed to if we thought causation transitive. Second, consider the following example. The boulder rolling down the hill causes the climber to duck, his ducking causes him to survive, yet we are inclined to deny that the boulder rolling down the hill caused him to survive.⁵ For these reasons, perhaps we should not regard causation as transitive.

(4) What is the direction of causation? In our world, if A causes B, A is earlier than B. But is this a necessary truth? As we shall see, Hume built it into his definition of 'cause' that a cause precedes its effect. But many have thought that this transforms a substantive issue into a matter of stipulation. For many philosophers, backwards causation is a theoretical possibility, and any theory of causation which ruled otherwise would thereby be suspect. Again this is controversial terrain, but it seems plausible to hold that, other things being equal, it is a point against a theory of causation if it excludes by definition the possibility of backwards, or even simultaneous, causation.

(5) We invariably talk of *the* cause of an event. For example, we say that the cause of the explosion was my striking the match. But there were many background conditions that had to be in place in order for that causal transaction to occur (e.g., the presence of oxygen, a dry match, flammable gas, etc.) Why are one of these not considered 'the' cause of the explosion?

A famous answer was given by J. S. Mill (1806–73). He thought that there was no distinction in reality between what we call 'the' cause and other background conditions; indeed the very distinction was 'capricious'.⁶ More recently, David Lewis (1942–2002) defended a similar view:

We sometimes single out one among all the causes of some event and call it 'the' cause, as if there were no others. Or we single out a few as the 'causes', calling the rest mere 'causal factors' or 'causal conditions'. Or we speak of the 'decisive' or 'real' or 'principle' cause. We may select the abnormal or extraordinary causes, or those under human control, or those we deem good or bad, or just those we want to talk about. I have nothing to say about the principles of invidious discrimination.⁷

'Capricious' and 'invidious' are strong words. Nonetheless, it has proven remarkably difficult to justify our ordinary talk of 'the' cause of some event. Better, perhaps, to see the world as shot through with causal transactions, only some of which are of interest to us. Selecting one cause as 'the' cause merely reflects our interests – e.g., what we regard as most explanatorily salient – and does not mark a distinction in reality between causes and 'mere' enabling conditions.

(6) What is the connection between causation and laws of nature? On one view, any true causal statement is simply an instance of a law of nature. The cement of causation is what distinguishes laws of nature (such as 'all men are mortal') from merely accidentally true generalizations (such as 'all men wear shirts'). However, as we shall see, it's far from obvious that every causal transaction is *ipso facto* an instance of a law of nature. Even so, it may still be plausible to hold that the necessity of a law of nature is the necessity of causation.

Let us now move on to the biggest question of all: what is the nature of the causal relation?

THE CONSTITUTIVE QUESTION: WHAT IS CAUSATION?

Consider four events A, B, C and D, where A precedes B and C precedes D. Suppose we judge that A caused B and that C caused D, but we don't think that A caused C or that B caused D. What distinguishes the {{A, B} {C, D}} pairs from the {{A, C} {B, D}} pairs? The first member of each pair occurred before the second member, yet we talk of cause only in relation to {A, B} and {C, D}. What do the {{A, B} {C, D}} pairs have which the {{A, C} {B, D}} pairs lack? In other words, if A causes B, what *makes it true* that A causes B? In virtue of what does A cause B?

Those who ask such questions typically assume that an informative answer can be given which does not presuppose the idea of causation. The task of the philosopher is then to sift through alternative answers and select the most plausible. This assumption is a reasonable one to begin with. If our use of a term is discriminatory (applying to some pairs of events but not to others), it's

Causation: reducible or basic?

If a property or relation is thought to be problematic, philosophers often attempt to domesticate it by reducing it to more basic, less problematic, components. Thus, for example, knowledge was long held to be reducible to justified true belief. Many have thought that causation is similarly reducible – either to regularities in nature or to counterfactual conditionals. On the regularity theory, A causes B if and only if A-type events and B-type events are regularly conjoined. On the counterfactual theory, A causes B if and only if, had A not happened, B wouldn't have happened either. Unfortunately, both theories are open to counterexamples, and so neither theory is tenable. But there is another possibility. Perhaps the causal relation is primitive, and so not reducible to more basic elements. On this view, our world is shot through with causal necessity. Such necessity underlies and explains the regularities we observe around us, but is not itself reducible to those regularities, or to anything else.

plausible to think that something informative can be said about what underlies our discriminations.

Thus what we seek are **necessary and sufficient conditions** for truth of sentences of the form 'A caused B'. We are looking for something of the form: A caused B if **and only if** X, Y and Z (where X, Y and Z do not presuppose the notion of causation). That is, in any case where A causes B, X, Y and Z are present, and in any case where X, Y and Z are present, A causes B. As noted, an assumption of this exercise is that X, Y and Z can be understood without reference to the notion of causation. This assumption may well prove to be false. Maybe there are no such X, Y and Z to be found.

HUME

Modern discussion of causation begins with David Hume. However, Hume was concerned not just with causation but with the origin of the idea of cause in our minds. This latter question was pressing given Hume's commitment to the following empiricist principle:

(EP) Every idea is either simple or complex. Complex ideas are constructions out of simple ones. Simple ideas derive from sense impressions (or from reflection).

In the *Treatise Concerning Human Nature*, Hume declares that the idea of causation involves the following relations between objects (or, as we will say, events):



Hume's projectivism

According to Hume, ideas are copies of impressions, and the idea of causation involves the idea of necessity. From what impression does the idea of necessity derive? Not from exposure to just one instance of an event of one type being followed by an event of another type. It is only after repeated exposure to such conjunctions that the mind forms the idea of necessary connection. The impression of necessity must therefore be an internal impression (consisting in feelings of expectation and anticipation) brought about by the mind's exposure to constant conjunctions. We mistakenly project this necessity back onto the world, giving rise to the illusion that necessary connections hold between events in the external world. So runs the standard interpretation of Hume on causation, according to which Hume is a projectivist rather than a realist about causation. However, recently some philosophers (such as Galen Strawson) have questioned this interpretation. Was Hume a regularity theorist? Did Hume really deny realism about causation?

- (i) Contiguity: causes and effects are contiguous in space and time.

If a brick hits a window and causes the window to smash, the two events – the brick hitting the window and the window smashing – are adjacent to each other in time and space. Of course, as Hume is aware, we also talk of my throwing the brick as causing the window to smash, and those events are not contiguous. But those events are linked by intermediate chains of cause and effect, each link of which connects contiguous elements.

- (ii) Priority: a cause must temporally precede its effect.

Hume considers whether 'precedes' might be weakened to 'precedes or is simultaneous with', but dismisses the proposal on the following grounds: 'if any cause may be perfectly contemporary with its effect, it is certain, . . . that they must all be so.'⁸ This would then lead to 'the destruction of that succession of causes, which we observe in the world; and indeed the utter annihilation of time.'⁹ However, this is not a compelling argument. Why does it follow from the claim that some causes are contemporaneous with their effects, that all must be?

It might be thought that Hume is merely claiming (i) and (ii) to be typical accompaniments of causation in standard everyday cases. But it is clear that he thinks them essential to causation quite generally. The claim that contiguity and priority are essential to all actual and possible instances of causation is far from trivial. Anyone who believes in the actuality (or even possibility) of action at a temporal or spatial distance would deny (i) and anyone who believes in

WHAT IS THIS THING CALLED METAPHYSICS?

the actuality (or even possibility) of backwards causation would deny (ii). (See Chapter 6.)

Hume accepts (i) and (ii) as essential to causation, but concedes that we cannot define causation only in terms of them. This is surely correct. There are many pairs of events, A and B, where A and B are contiguous, and A is prior to B, yet where A plainly does not cause B. (For example, suppose my dog barks whenever I throw a brick at a window. The soundwaves of the bark are prior to, and contiguous with, the window-smashing, yet they do not cause the window to smash.) What more is required? Hume suggests:

- (iii) Necessity: if A causes B there is a necessary connection between A and B.

This is an intuitively plausible condition on causation. If I throw a brick and it hits the window, we don't think that it just so happens that the window smashes; we think the window *must* smash. Cause and effect are necessarily connected. But what kind of necessity is this? Not logical necessity, as Hume reminds us. There is no logical contradiction in the eventuality of a brick just bouncing off a window. This may never happen, but it is not a logical impossibility. (Such an outcome is conceivable in the way that, e.g., round squares or married bachelors, are not conceivable. We know *a priori* that we will never encounter such entities.) The necessity in question is causal or natural necessity, and it is a further question whether this species of necessity admits of further analysis.

Given his commitment to (EP), Hume faces the following problem: how do we acquire the impression of causation from which the idea of causal necessity is derived? In the *Treatise*, Hume says that when he observes instances of cause and effect, all he perceives are the contiguity and succession of events. He has no impression of a necessary connection between them.

Hume leaves the matter there in the *Treatise*, but takes it up in the *Enquiry*, where he tells the following story of how we arrive at the idea of causal necessity. We cannot, from the observation of a single instance of an event of one type being followed by an event of another type, derive the idea of power or necessary connection. 'All events seem entirely loose and separate. One event follows another; but we never can observe any tie between them. They seem *conjoined*, but never *connected*.'¹⁰ But from the observation of multiple instances we can acquire the idea of necessary connection: 'when one particular species of event has always, in all instances, been conjoined with another, we make no longer any scruple of foretelling one upon the appearance of the other.'¹¹ In other words, upon exposure to many cases of brick-throwings being followed by window-smashings, and observing no counter-instances, we naturally expect future brick-throwings to be followed by window-smashings. 'This connexion, therefore, which we feel in the mind, this



customary transition of the imagination from one object to its usual attendant, is the sentiment or impression from which we form the idea of power or necessary connection.¹²

According to Hume, the idea of necessary connection is a copy of an impression in our minds, and not a copy of any feature in the external world. That is, upon repeated exposure to conjoined types of events in the world, we naturally form certain feelings of expectation and anticipation, and from those feelings we derive the idea of necessary connection. That idea is not a copy of any relation in the world. Hume is offering us an explanation of how we arrive at this idea without postulating necessary connections between events in the external world. Our natural belief that there is necessity in the world is a **projection** of the mind onto the world. And this, presumably, explains why, in defining ‘cause’ in the *Enquiry*, Hume makes no reference to the notion of necessity, but instead proposes a regularity theory of causation. If there is no natural necessity in the world, what makes our causal claims true cannot contain such necessity.

THE REGULARITY THEORY

Hume writes: ‘we may define a cause to be *an object, followed by another, and where all objects similar to the first are followed by objects similar to the second*. Or in other words, *if the first object had not been, the second never had existed*.’¹³

As many have observed, Hume’s phrase ‘or in other words’ is puzzling since the second italicized definition is quite different from the first. The second definition has given rise to a quite different approach to causation – the counterfactual theory – which we will look at shortly. It is the first definition that is our concern here. We can unpack Hume’s suggestion as follows:

- (RT) A caused B if and only if A preceded B and all A-type events are followed by B-type events.

In other words, once we see regular patterns in nature, constant conjunctions, we classify them as causal interactions. Is the regularity theory plausible?

Note that the regularity theory is intended as a reductive account of causation. What follows the ‘if and only if’ is taken to make no reference to causation. It is intended to tell us, in other and more familiar terms, what causation is. There are, therefore, two broad ways in which the regularity theory might be criticized. It may be true but fail to be reductive, or it may be false. And there are two ways in which it might be false. First, there might be a case where A caused B but either A did not precede B or not all A-type events were followed by B-type events. Second, there might be a case where A preceded B and all

A-type events were followed by B-type events but A did not cause B. Arguably the regularity theory can be criticized in all these ways.

First, there is a trivial way in which the regularity theory can fail from right to left. That is, there are cases where A precedes B and all A-type events are followed by B-type events, yet A does not cause B. Suppose that A precedes but does not cause B, and that A is an utterly unique event: the universe contains no other A-type events. This situation is surely possible. But the regularity theory implies that it is impossible. If A is followed by B, and A is the only A-type event in the universe, then (trivially) all A-type events are followed by B-type events. In which case, according to the regularity theory, the condition for causality is fulfilled, and so A caused B. But we began by assuming that A did not cause B. So the regularity theory, as stated, must be false.

Second, even when there are many A-type events, there are still scenarios in which all A-type events are followed by B-type events, yet in which there is no causation. Three centuries ago, Thomas Reid (1710–96) pointed out that ‘[i]t follows from [Hume’s] definition of a cause, that night is the cause of day, and day the cause of night. For no two things have more constantly followed each other since the beginning of the world.’¹⁴ This, of course, is the wrong result: day does not cause night, nor *vice versa*; rather day and night are joint effects of a common cause (the Earth’s rotation on its axis).

Might the regularity theory fail in the other direction too? That is, could there be cases where A causes B and yet either (i) A does not precede B or (ii) not all A-type events are followed by B-type events? Anyone who finds the notion of backwards causation coherent will hold that there could be cases where A causes B, yet B is earlier than A. As for (ii), there are philosophers of causation – called ‘singularists’ (such as Elizabeth Anscombe and Curt Ducasse) – who think that from the fact that A caused B it does not follow that there must be a true and exceptionless generalization to the effect that all A-type events are followed by B-type events.¹⁵

Anscombe offers four reasons in favour of singularism. First, nothing in our causal talk forces us to accept the universalism expressed in the regularity theory (the view that every causal interaction is underwritten by an exceptionless generalization). For example, one can perfectly well understand talk of someone contracting a disease through contact with an infected individual *without* taking a stand on whether that causal transaction is underwritten by some exceptionless generalization.

Second, Anscombe appeals to developments in twentieth-century physics. If quantum phenomena are **indeterministic**, and yet there are causal transactions at the quantum level, then there actually are cause–effect pairs which do not fall under exceptionless generalizations of the ‘All As are Bs’ variety.

Third, unlike Hume, Anscombe thinks that causation is directly observable in single instances. If it is allowed that we directly perceive material bodies (as opposed to fleeting sense-data) ‘then what theory of perception can justly

Elizabeth Anscombe (1919–2001)

Gertrude Elizabeth Margaret Anscombe studied at St Hugh's College, Oxford, and later attended lectures at Cambridge given by the Austrian philosopher Ludwig Wittgenstein. She became a close friend of Wittgenstein, translated his *Philosophical Investigations* (1953) and was nominated as one of his literary executors. She was subsequently appointed to Wittgenstein's old chair at Cambridge, which she occupied from 1970 until her retirement in 1986. Apart from her translations of Wittgenstein, Anscombe is best known for *Intention* (1957), *Three Philosophers* (1963) (written jointly with her husband, the philosopher and logician P. T. Geach), and three volumes of collected papers, ranging over topics in the history of philosophy, metaphysics, philosophy of mind, ethics and religion. A controversial figure, she opposed the entry of Britain into the Second World War (on the grounds that civilians would deliberately be killed), and was fiercely opposed to contraception and abortion.

disallow the perception of a lot of causality?¹⁶ This observability thesis clearly fits with singularism. If I can directly perceive causation in a single instance, what happens elsewhere can hardly be relevant to determining whether that sequence of events is a causal one.

Fourth, universalism has the odd consequence that whether an apparently causal transaction (say, a brick smashing a window) is genuinely causal depends on whether, at any time in the past or future, a similar brick-throwing has or has not been followed by a similar window-smashing. This is counter-intuitive. When we say that that brick-throwing caused that window-smashing, we are not making a prediction covering all space and time. Singularism avoids this problem.

The regularity theory is intended to be a reductive account of causation. It attempts to explain causation in other terms – precedence and regularity – neither of which are presumed to presuppose causation. On the face of it, neither the idea of A occurring before B nor the idea of all A-type events preceding B-type events presupposes the idea of causality. But matters are not so straightforward.

What is it for an event to be an event of a particular type? It is natural to understand the notion, as Hume does, in terms of similarity. Since A and B are particular events, they are unrepeatable. That window-smashing (e.g., the one that occurred on 1 January 2006) cannot occur again. But similar events (events of that type) can occur again.

Now similarity can mean either 'similarity in all respects' or 'similarity in some respects'. The former is hopeless for present purposes: no two window-smashings are similar in *all* respects, both intrinsic and relational. They may

WHAT IS THIS THING CALLED METAPHYSICS?

be caused by different people, or involve different windows; and even if they are caused by the same people, those people will be different in some way, even if just in age. Since the world is constantly changing, no two events at different times can ever be exactly similar in all respects.

Maybe by similarity we mean ‘similarity in some respects’, which fits better with ordinary usage (if I say that two pole-vaulters are similar, I don’t mean similar in all respects, but only in some respect, e.g., vaulting technique). So an A-type event is any event similar to A in some respect. But which respect? We may say ‘in the relevant respects’. But that is no answer: which respects are the relevant ones? The danger looms that the relevant respects can be specified only as the *causally* relevant respects: those respects relevant to bringing about a B-type event. But now the individuation of event-types presupposes the notion of causation, contrary to the reductive intention behind the regularity theory.¹⁷

THE COUNTERFACTUAL THEORY

After Hume defined his version of the regularity theory, he added the sentence: ‘Or in other words, *if the first object had not been, the second never had existed.*’¹⁸ As noted earlier, this sentence is mysterious since the idea it puts forward seems quite different from any version of the regularity theory. The notion of a counterfactual is more sophisticated, and logically quite different from, the notion of a regularity or constant conjunction. It points us to a new (and singularist) theory of causation, the counterfactual theory, the best-known advocate of which is David Lewis.¹⁹ According to the counterfactual theory:

(CT) A caused B if and only if if A hadn’t happened, B wouldn’t have happened.

The counterfactual theory makes essential use of counterfactual conditionals. We all use counterfactuals regularly and unthinkingly. We all understand the counterfactual conditional ‘if you had not been wearing a seatbelt, you would have been killed’. You were wearing a seatbelt (hence the **antecedent** of the conditional is counter to the facts), and you weren’t killed; but if you hadn’t been wearing a seatbelt, you would have been killed. The technical issue of how to analyse these conditionals is a difficult matter, but not one that need concern us here. All we need to draw on is our common sense understanding of the world and our grasp of the English language.

Two favourable points about the counterfactual theory are worth making. First, in most ordinary cases of causation, the counterfactual specified by the counterfactual theory is true. If I throw a brick at the window, causing it to

smash, it will typically be true that, had I not thrown the brick, the window would not have smashed. Second, the counterfactual theory does attempt to accommodate Hume's claim that the idea of causation involves the idea of necessity. The counterfactual theory captures the sense in which a cause is necessary for its effect: without the cause, there would have been no effect.

So should we accept the counterfactual theory? No. It is open to counterexamples. In particular, there can be cases where A causes B, but in which it is not true that, had A not happened, B would not have happened. So-called pre-emption cases are one example of this. Suppose that, as a result of hypnosis, Mary will jump out of the window if her phone rings at 8 a.m. Imagine that, knowing this, Bill rings her at 8 a.m., causing her to jump out of the window. Fred, who also dislikes Mary, is waiting in the wings, in case Bill should fail to call (but does not have to intervene). Here Bill's ringing Mary caused her to jump out of the window. But had Bill not called, the very same sequence of events would have ensued (since Fred would have called instead). So it's not true that had Bill not called, Mary's jump would not have occurred.

Another kind of case is that of overdetermination. Suppose that A and B together cause C, and that either would have caused C in the absence of the other. (Suppose that Mary has two phones, and that either ringing by itself would cause her to jump. Bill and Fred both phone at 8 a.m., causing her to jump.) C is not counterfactually dependent upon A (or upon B). If A hadn't happened, C would still have happened; and if B hadn't happened, C would still have happened. Hence, according to the counterfactual theory, A didn't cause C and B didn't cause C. But if neither A nor B caused C, then how did C come about? A defender of the counterfactual theory may say that A and B together cause C. It's true that had neither A nor B happened, C would not have happened. But, in a case where the outcome is overdetermined, how can A and B together cause C without A causing C and B causing C?²⁰

A SIMPLER THEORY

The flaws in the regularity and counterfactual theories are similar. They both fail (in different ways) to give proper place to the role of necessity in causation. Regularities can obtain accidentally, and A can necessitate B even though B is not counterfactually dependent on A. Both theories fix on contingencies which happen to be true in most cases of causation, but are not of the essence of causation. Causes and effects do typically exhibit a regular pattern, and effects are typically counterfactually dependent on their causes. But neither theory fully captures our notion of causation.

Since neither theory is satisfactory, perhaps we should consider a simpler, non-reductive, theory of causation, the simple theory:

WHAT IS THIS THING CALLED METAPHYSICS?

(ST) A caused B if and only if A (plus relevant background conditions) necessitates B

where the necessity involved is causal necessity, a relation which is taken to be primitive. Causal necessity should be distinguished from logical necessity (i.e., the kind of necessity exhibited by ‘either it’s raining or it’s not raining’). As Hume observed, if A caused B it is never logically contradictory to suppose ‘A but not B’. This non-logical species of necessity is held to be a basic ingredient in our ontology. Unlike the regularity and counterfactual theories, the simple theory offers a non-reductive account of causation: the notion of ‘cause’, and the allied notion of ‘causal necessity’, are not further analysable. We cannot explain the idea of causal necessity in other, more basic, terms.

One objection to the simple theory is epistemic. If all we observe are regularities, it will be urged, why should we believe in causal necessity? Following Anscombe, we may hold that causation can be directly observed in some cases, as when we will our arms to move, or see the knife cut the butter. Second, in other cases of causation we might appeal to the principle of ‘inference to the best explanation’. That is, postulating causal necessity best explains the regularities we observe around us. Why is it that whenever I throw a brick at a window it smashes? Because each window-smashing follows from each brick-throwing by causal necessity. Causal necessity is something we postulate to explain the world around us. It is a theoretical postulate of common sense. We certainly have good evidence of causality when we have regularities and counterfactual dependencies. But these are not what causality consists in. They are pointers to what lies beneath: causal necessity.

CONCLUDING REMARKS

We have seen that two purportedly reductive theories of causation, the regularity theory and the counterfactual theory, are open to objection. A natural alternative is the simple theory which has the following features.

- It holds that causation involves necessity (causal or natural necessity).
- It holds that the notion of causal necessity admits of no reductive analysis.
- It fits with the singularist view of causation.
- It fits with the view that causation is sometimes directly observable and, on other occasions, a theoretical postulate which explains the regularities we observe around us.

- Finally, the simple theory allows us to distinguish laws of nature from accidentally true generalizations in virtue of the fact that laws are true of (natural) necessity.

Such a theory is worth taking seriously.

STUDY QUESTIONS

- What are the *relata* of the causal relation?
- How plausible is Hume's empiricist principle (EP)?
- Is backwards causation possible?
- What is the most telling objection to the regularity theory?
- Why are cases of pre-emption and overdetermination problems for the counterfactual theory?

ANNOTATED FURTHER READING

- G. E. M. Anscombe, 'Causality and Determination', in E. Sosa and M. Tooley (eds) *Causation* (Oxford: Oxford University Press) 1993, pp. 88–104. A trenchant anti-Humean tract. Anscombe argues that causal relations do not presuppose exceptionless generalizations and that causation can be directly perceived. A difficult paper.
- P. Horwich, 'Lewis's Programme', reprinted in Sosa and Tooley, *op. cit.*, pp. 208–16. A useful, if demanding, critique of Lewis's counterfactual theory.
- D. Lewis, 'Causation', reprinted in Sosa and Tooley, *op. cit.*, pp. 193–204. The classic source of counterfactual approaches to causation. A difficult read.
- E. Sosa and M. Tooley, 'Introduction', in Sosa and Tooley, *op. cit.*, pp. 1–32. A readable introduction to the issues discussed in this chapter.
- G. Strawson, *The Secret Connexion* (Oxford and New York: Clarendon Press) 1989. Strawson makes a detailed case for interpreting Hume as a realist about causation.

INTERNET RESOURCES

- J. Faye, 'Backward Causation', *The Stanford Encyclopedia of Philosophy* (Fall 2005 edition), Edward N. Zalta (ed.). Retrieved 31 May 2006 from <<http://plato.stanford.edu/archives/fall2005/entries/causation-backwards>>.
- D. Garrett (2005) 'Hume, David', in E. Craig (ed.), *Routledge Encyclopedia of Philosophy*. Retrieved 31 May 2006 from <<http://www.rep.routledge.com/article/DB040>>.

WHAT IS THIS THING CALLED METAPHYSICS?

- P. Menzies (2001) 'Counterfactual Theories of Causation', *The Stanford Encyclopedia of Philosophy (Spring 2001 edition)*, Edward N. Zalta (ed.). Retrieved 31 May 2006 from <<http://plato.stanford.edu/archives/spr2001/entries/causation-counterfactual>>.
- J. Schaffer (2003) 'The Metaphysics of Causation', *The Stanford Encyclopedia of Philosophy (Spring 2003 edition)*, Edward N. Zalta (ed.). Retrieved 31 May 2006 from <<http://plato.stanford.edu/archives/spr2003/entries/causation-metaphysics>>.



5 TIME

The fundamental issue

- | | |
|--|----|
| • Introduction | 71 |
| • Setting up the debate: A series and B series | 72 |
| • The A series is fundamental to time | 73 |
| • Russell's reply | 75 |
| • McTaggart's paradox | 76 |
| • Diagnosis | 78 |
| • A theory or B theory? | 81 |
| • Concluding remarks | 82 |

INTRODUCTION

In this chapter, we are concerned with a fundamental question about our world: What is time? To focus our discussion, we will look in detail at some ideas of J. M. E. McTaggart (1866–1925). McTaggart's writings largely set the agenda for twentieth-century philosophical discussions of time. Although McTaggart's overall argument, with its conclusion that time is unreal, is generally rejected, his two subsidiary arguments still have supporters today.

J. M. E. McTaggart (1866–1925)

John McTaggart was born in Wiltshire, England. He studied and taught at Trinity College, Cambridge for most of his life (where his pupils included Bertrand Russell). In his most famous paper 'The Unreality of Time' (1908) he distinguished two ways of ordering events in time, the A and B series. He argued that the A series is fundamental to time, but that it leads to contradiction. He concluded that time is unreal. He also held that space is unreal. The spatio-temporal world was deemed an illusion, and reality was held to consist of immaterial souls, in ecstatic communion with each other. Along with F. C. Bradley (1846–1924), McTaggart was a leading member of the short-lived school of Hegel-influenced British Idealism which languished at the end of the nineteenth and early twentieth centuries.

SETTING UP THE DEBATE: A SERIES AND B SERIES

McTaggart begins by observing that:

Positions in time, as time appears to us *prima facie*, are distinguished in two ways. Each position is Earlier than some and Later than some of the other positions. . . . In the second place, each position is either Past, Present or Future. The distinctions of the former class are permanent, while those of the later are not. If M is ever earlier than N, it is always earlier. But an event, which is now present, was future, and will be past.¹

The first series of positions McTaggart labels the B series, the second the A series. It is certainly plausible to say that positions in time are distinguished by us in both ways, and that B-series positions are permanent in a way that A-series positions are not. This shows up in the unchanging truth value of B-series attributions. If it is true that my birth was earlier than 1970, then it always was true and always will be true that my birth was earlier than 1970. The truth value of the sentence 'my birth is earlier than 1970' never changes. In contrast, A-series attributions are temporary rather than permanent. My birth was future, then fleetingly present, then past for ever more. The truth value of the sentence 'my birth is future' changed from true to false.

Nonetheless, the A and B series are not independent time series. There are obvious and undeniable **truth-value links** between A series and B series attributions. For example, my utterance in 2006 'Hitler's death is past' is true if



A series/B series

McTaggart observed that positions in time belong to two series – what he called the ‘A series’ and the ‘B series’. Every event belongs to both series, and yet the series have quite different characteristics. Events in the A series are constantly changing their positions. Future events become present; present events become past, past events become more past. In contrast, an event’s position in the B series never changes. If an event occurred in 1961 it always was true and always will be true that it occurred in 1961. If one event is earlier than another event, that fact never changes. Three views are possible regarding the A and B series: the A series is fundamental to time; the B series is fundamental; neither series is fundamental. A defender of the first view is an A theorist. A defender of the second is a B theorist. A defender of the third view thinks that the debate has been misconceived: there are two languages of time but only one time series, which can be described in different ways.

and only if Hitler’s death is earlier than 2006; my utterance in 2006 ‘Blair’s death is future’ is true if and only if Blair’s death is later than 2006; and my utterance on 1 February 2006 ‘it’s raining now’ is true if and only if it’s raining on 1 February 2006. These truth-value links may tempt some to think that the A-series/B-series distinction is notational rather than metaphysical. That is, some might think that the A and B series stand to time as, e.g., centigrade and Fahrenheit stand to temperature, or as inches and centimetres stand to distance: mere notational variants. That is a possible view, but it is not McTaggart’s view. McTaggart holds that one series (the A series) is more fundamental to time than the other, and hence that the B theory – which holds that the B series is essential to time – is false.

THE A SERIES IS FUNDAMENTAL TO TIME

According to McTaggart, the fact that B-series positions are permanent should not lead us to think that they are more objective, or more essential to time, than positions in the A series. On the contrary, though A and B determinations are both essential to time, A-series determinations are more **fundamental** than B-series determinations.

The distinction between essential and fundamental is certainly a subtle one. In standard usage, ‘fundamental’ and ‘essential’ are interchangeable. What distinction, then, is McTaggart trying to draw?

It might be thought that McTaggart is claiming that, although the A and B series are both essential (or necessary) for time, only the A series is sufficient

WHAT IS THIS THING CALLED METAPHYSICS?

for time. In that sense, perhaps, the A series is more fundamental than the B series. However, McTaggart concedes that the B series ‘cannot exist except as temporal’, suggesting that the B series is sufficient to constitute time.² But he also says that the B series is ‘not by itself sufficient to constitute time’.³ What is going on here?

It may be that McTaggart has something like the following in mind. A description of reality only in B-series terms is necessarily incomplete. The set of B-series facts is a proper subset of the totality of temporal facts. This incompleteness means that the B series cannot be fundamental to time. In contrast, no such incompleteness afflicts the A series.

Why does McTaggart think that the A series is fundamental to time? McTaggart first presents his argument, then considers an objection to it by Bertrand Russell. His argument can be stated as follows:

- (1) Time necessarily involves change.
 - (2) Change is possible only in the A series.
- So: (3) Time fundamentally involves the A series.

McTaggart does not offer any argument for premise (1). He merely asserts that ‘there could be no time if nothing changed.’ Some philosophers, most notably Shoemaker, have held that it is possible for there to be time without change (see Chapter 6). However, McTaggart intends (1) to be a truism, since he counts the mere passage of time (or ‘pure becoming’) as change. So intended, this premise begs the question against the B theorist from the outset. Fortunately, however, McTaggart does not need premise (1). The truth of premise (2) by itself would demonstrate the superiority of the A series over the B series.

McTaggart does have an argument for premise (2). He appeals to the permanency of B-series locations and relations. Events never change their B-series location (it always was and always will be true that Hitler’s death occurred in 1945), nor their B-series relations to other events (it always was and always will be true that Hitler’s death is later than Caesar’s death). This, for McTaggart, suffices to show that the B series does not allow for change.

The only respect in which an event – such as the death of Queen Anne – can change is in the following respect: ‘[i]t was once an event in the far future. It became every moment an event in the nearer future. At last it was present. Then it became past, and will always remain past, though every moment it becomes further and further past.’⁴ (On McTaggart’s view the past is constantly changing: past events are becoming more past.) Hence, McTaggart concludes, change is only possible on the A series, and so the A series is fundamental to time.



RUSSELL'S REPLY

Russell did not think the A series fundamental to time. According to Russell,

past, present and future do not belong to time *per se*, but only in relation to a knowing subject. An assertion that N is present means that it is simultaneous with that assertion, an assertion that it is past or future means that it is earlier or later than that assertion. . . . If there were no consciousness . . . nothing would be past, present, or future.⁵

Thus the A series is subject-relative or mind-dependent, and hence not fundamental. Since Russell believes that the B series is fundamental to time, he is a B theorist. An A theorist, in contrast, holds that the A series is fundamental to time.

How then does Russell respond to McTaggart's argument against the B theory? Russell and McTaggart differ over how to characterize change. On Russell's characterization, premise (2) is false. Russell writes:

Change is the difference, in respect of truth or falsehood, between a proposition concerning an entity and the time T, and a proposition concerning the same entity and the time T*, provided that these propositions differ only by the fact that T occurs in the one where T* occurs in the other.⁶

Thus, there is change if, e.g., the proposition 'at time T my poker is hot' is true, and the proposition 'at time T* my poker is hot' is false. More simply,

Change

What is change? According to Russell, change is an object's having different and incompatible properties at different times. If my garden gate is green on Monday and then painted red on Tuesday, it has changed. According to McTaggart, change occurs not to objects but to events. No event can change its B series location: that is fixed and permanent. An event changes only in respect of its A series position: it was first future, then briefly present, then past for ever more. Hence, McTaggart concluded, change is possible only on the A series. Many have held that 'Russell change' is our ordinary notion of change. We talk of an object changing, and a change in an object is an event, a happening. Does the event need to change in order for there to be change? B theorists think not. Some A theorists think not too. They hold that 'Russell change' is a perfectly acceptable notion of change, but that, for other reasons, the A series is fundamental to time.

WHAT IS THIS THING CALLED METAPHYSICS?

there is change if my poker is hot at one time and not hot at some other time. And change, so understood, requires only the B series.

McTaggart has a reply to Russell. According to McTaggart, Russell's account is simply not an account of change. For the proposition 'at time T my poker is hot', if true, is always true, and the proposition 'at time T* my poker is hot', if false, is always false. And 'this makes no change in the qualities of the poker.'⁷ It is true at all times that the poker is hot at T and that it is not hot at T*. This permanency in truth value, for McTaggart, implies that there is no change in the poker.

McTaggart makes the same point in terms of facts. The fact that the poker is hot at T never changes; nor does the fact that it is not hot at T*. Yet 'there can be no change unless facts change.'⁸ And the only facts which can change are A-series facts. If T lies in the present, the poker is now hot and will be cold. The fact of its being hot will give way to the fact of its being cold, when the latter fact becomes present. Hence, concludes McTaggart, change is possible only on the A series.

How should we adjudicate this dispute? We are being offered two quite different accounts of change, which we can label 'McTaggart change' and 'Russell change'. McTaggart change occurs whenever an event or a fact alters its A-series position. Russell change occurs whenever an object has incompatible properties at different times. Is Russell change recognizable as a notion of change?

Prima facie, the answer is 'yes'. An object's altering its properties – e.g., my garden gate being green on Monday and then painted red on Tuesday – would ordinarily count as a change. In order for change to occur, we don't require that the event of the painting or the fact of the gate's being green change. We simply require the object to have changed its properties. In which case Russell's reply stands, and McTaggart's ground for premise (2) crumbles.

However, there is another objection to McTaggart's argument. In assuming that change requires change in an event or a fact, and given that the only way an event or fact can change is with respect to its A-series position, McTaggart has begged the question against his B-theorist opponent. No B theorist would accept a definition of 'change' which simply builds in reference to the A series. McTaggart's argument thus presupposes what it sets out to prove.

McTAGGART'S PARADOX

One might have thought that McTaggart would have taken himself to have established the A theory of time and left it at that. But that is not what happened. Having argued that time fundamentally involves the A series, McTaggart then proceeded to argue that the A series is contradictory, and hence that time is unreal. McTaggart happily embraced the conclusion that the



passing of time is an illusion and that nothing ever changes. Though this conclusion hardly commends itself to common sense, McTaggart's argument that the A series is contradictory – known as McTaggart's paradox – is an interesting and ingenious argument in its own right.

McTaggart writes:

Past, present and future are incompatible determinations. . . . But every event has them all. If M is past, it has been present and future. If it is future, it will be present and past. If it is present, it has been future and will be past. Thus all three characteristics belong to each event.⁹

We can represent McTaggart's reasoning as follows:

(4) Every event is past, present and future.

(5) No event can be past, present and future.

So: (6) The A series is contradictory.

The thought behind (4) is that no event escapes the passage of time: any event is future, then fleetingly present, then past for ever more. Every event occupies every A-series position. (There will be exceptions, if there is a first or last event. But this does not affect the argument since the first event would still be present and past, and the last event future and present, and these determinations are incompatible.)

The thought behind (5) is that past, present and future are incompatible determinations: if an event is past it is not present or future, if it is present it is not past or future, and so on. Nothing can possess incompatible characteristics. From (4) and (5), (6) follows.

Now it might seem that there is an obvious reply to this argument. Indeed McTaggart states the reply himself:

It is never true, the answer will run, that M *is* present, past and future. It *is* present, *will be* past, and *has been* future. Or it *is* past, and *has been* future and present, or again *is* future and *will be* present and past. The characteristics are only incompatible when they are simultaneous, and there is no contradiction to this in the fact that each term has all of them successively.¹⁰

Thus, there is a sense in which (4) is true, but, in that sense, (5) is false; and there is a sense in which (5) is true, but, in that sense, (4) is false. Consequently, the (4)–(6) argument is unsound, since it has incompatible premises.

But McTaggart has an ingenious response to this reply. We avoid the charge of contradiction in the three ground-level A-series positions (past, present,

WHAT IS THIS THING CALLED METAPHYSICS?

future) by invoking three second-level A-series positions (e.g., M is present, was future and will be past). But there are nine positions in this second-level series (is past, is present, is future, was past, was present, was future, will be past, will be present, will be future) and every event occupies each of these A-series positions. Some combinations of these nine positions are incompatible (e.g., is present and is past). We can avoid these contradictions by distinguishing more complex tenses, and moving up to third-level A-series positions. But some of these twenty-seven positions will be incompatible. To avoid contradiction we must move up to a fourth level, and so on. We can escape contradiction by moving up a level, but at every level a contradiction remains. ‘And, since this continues infinitely, the first set of terms never escapes from contradiction at all.’¹¹ The ‘obvious reply’ is not so obvious after all.

DIAGNOSIS

McTaggart’s paradox, as we have labelled it, has a Janus-like quality: on one day it can seem compelling, on the next, sophistical. But I think the A theorist (who is the intended target of the paradox) has a reply. In what follows, I draw on some ideas of Michael Dummett and Paul Horwich.

In developing his paradox, Dummett suggests, McTaggart makes an implicit assumption.¹² He assumes that a consistent and complete description of reality is, in principle, possible. That is, irrespective of one’s position in time, it is possible to give a description of reality which is consistent and includes all truths. What McTaggart’s paradox reveals is that the A theory is incompatible with this assumption. When we try to specify all the A-series truths, as opposed

M. A. E. Dummett (1925–present)

Sir Michael Dummett was born in London in 1925. After serving in World War Two, he studied at Christ Church, Oxford before being elected a Prize Fellow of All Souls in 1950. In 1979 he became Wykeham Professor of Logic and a Fellow of New College, a position he held until retirement in 1992. A committed Catholic and social activist, Dummett was especially involved in anti-racism campaigns in Britain during the 1960s. The two major influences on Dummett’s philosophy are Frege and Wittgenstein. He has written three substantial tomes on Frege’s philosophy of language and philosophy of mathematics. Dummett has also developed the view that traditional debates in metaphysics can be advanced, and perhaps even resolved, by focusing on debates in the theory of meaning. Here the Wittgensteinian doctrine of the public nature of meaning (and the corresponding impossibility of a private language) has had a great effect on Dummett’s thinking.



to specifying those which are true from one's current perspective, we end up in contradiction. Thus, if the A theory is true, a complete and consistent description of reality is impossible. Given the assumption that such a description is possible, McTaggart concluded that the A series is contradictory. In contrast, a description of reality in B-series terms is consistent, but incomplete (according to McTaggart, it leaves out the fact of change). Thus, in a different way, the B theory flouts McTaggart's assumption.

But does the A theorist have to accept McTaggart's assumption that a complete and consistent description of reality is, in principle, possible? Certainly, he must accept the consistency requirement. But need he, indeed should he, accept the completeness requirement, that is, the requirement that reality can be completely described, independently of one's temporal perspective?

On reflection, the answer seems clearly 'no'. It is of the essence of the A theory that the fundamental temporal facts (i.e., the tensed facts which record an event's position in the A series) change as time passes. According to the A theory, reality is constantly being (re)partitioned: facts that are future, become present; facts that are present become past; facts that are past become more past. Reality is constantly changing. If it is now 2006 I can state the facts as they are from that perspective including, e.g., the fact that my death is future. In 3006 a statement of the facts will include the fact that my death is past. There is only a contradiction if we assume that there must be some (perspective-neutral) description which includes both facts. The A theorist should deny that there can be any such description: any description of temporal reality, from within time, is necessarily incomplete. In which case, the A theory can escape McTaggart's paradox.

Dummett admits that it is not easy to give up the belief that there must be a complete description of reality, 'that of anything which is real, there must be a complete – that is, observer-independent – description.'¹³ Nonetheless he is willing to take the moral of McTaggart's reasoning to be that we should 'abandon our prejudice that there must be a complete description of reality.'¹⁴

Paul Horwich, in contrast, thinks that the completeness requirement should not be given up, and so concludes that McTaggart's paradox does refute the A theory of time.¹⁵ Why does Horwich regard the completeness requirement as sacrosanct? He thinks the A theorist's idea that there is 'a variation, from one time to another, as to which facts obtain' trades on 'an idiosyncratic and unmotivated conception of *fact*.'¹⁶ In other words, abandoning the completeness requirement (as the A theorist does) violates our ordinary understanding of what a fact is, and hence the requirement should be respected.

Horwich offers an example to support his claim. He says that:

we do not regard

X is to the left of Y

WHAT IS THIS THING CALLED METAPHYSICS?

and

X is not to the left of Y

as explicit descriptions of facts. Rather we suppose that whenever such claims are true, they are partial accounts of facts whose explicit descriptions take the form

X is to the left of Y relative to Z

and

X is not to the left of Y relative to W. . . .

The general point is that we reserve the term ‘fact’ for those aspects of reality whose explicit descriptions are sentences that are true *simpliciter* – and not merely true relative to some contexts or points of view, and false relative to others.¹⁷

How should an A theorist reply? It is significant that Horwich’s example is a spatial one. This, the A theorist will say, makes all the difference. For it is plausible to suppose that there can be a complete and consistent description of all spatial facts. The completeness intuition is robust in the case of spatial facts. Two people with very different positions in space can still agree on all the spatial facts. This, after all, is why maps are useful: we don’t need a different map for each location. So Horwich is right that, in his example, the canonical description is ‘X is to the left of Y relative to Z’ rather than ‘X is to the left of Y’, but that is because spatial facts admit of a complete (observer-independent) description. But temporal facts do not, and Horwich’s example merely serves to highlight a fundamental difference between time and space. Or so the A theorist may plausibly reply.

Another way of putting the A theorist’s reply is as follows. Although one could formulate an ‘A theory’ of space – the view that here and there, near and far, are fundamental spatial properties, not reducible to ‘B series’ spatial properties and relations such as ‘at Oxford’, ‘50 miles north of London’, etc. – such a theory has no plausibility whatsoever. As Dummett puts it: ‘the use of spatially **token-reflexive expressions** [i.e., expressions such as ‘here’ and ‘there’] is not essential to the description of objects as being in a space. That is, I can describe an arrangement of objects in space although I do not myself have any position in that space.’¹⁸

I can give a complete description of the spatial facts from any spatial perspective or from none, and hence without using terms such as ‘here’. In contrast, Dummett suggests, I cannot give a complete description of temporal facts without using perspectival terms such as ‘now’, ‘past’ and ‘future’. In which case the completeness requirement, though plausible in the spatial case, is implausible in the temporal case. By the same token, the idea of a perspectival



fact is not ‘idiosyncratic’, but exactly what tensed facts would be like if the A theory is true.

A THEORY OR B THEORY?

Let us take stock. We have described two temporal series, the A series and the B series, and we have said that there are two competing theories of time, the A theory and the B theory. According to the A theory, the A series is fundamental to time. According to the B theory, the B series is fundamental. McTaggart presented two arguments: the first against the B theory (claiming that it was incomplete because it could not account for the fact of change), and the second against the A theory (claiming that the A series is contradictory). Having taken himself to have demolished both theories of time, McTaggart concluded that time is unreal.

However, we found neither argument to be convincing. There is a perfectly sensible notion of change definable in B-series terms, and the A theorist has the resources to obstruct the derivation of McTaggart’s contradiction. Since neither theory has been refuted, we can ask again, which theory should we accept: the A theory or the B theory?

The A and B theories are not just terminological variants; they offer very different pictures of reality. According to the A theory, reality is constantly changing simply by virtue of the passage of time. Events are constantly exchanging their A-series positions. My birth was future, then briefly present, then past for ever more. This dynamic picture of reality is sometimes presented in terms of a moving *now*, illustrating the flow of time.

Different versions of the A theory are possible. According to presentism, only the present is real. According to C. D. Broad, the present and past are real, but the future is unreal. On Broad’s view, the sum total of reality is accreting as time passes. According to McTaggart, an A theorist should hold that past, present and future are equally real.¹⁹

However, only the first two versions seem in the spirit of the A theory, and Broad’s view is the more plausible.²⁰ It is of the essence of the A theory that the present is assigned a privileged position, and it cannot have such a position if past, present and future are equally real. The *now* earns its keep by doing serious ontological work: as it glides over events, it gives them being. In holding that future is unreal, the A theorist allows a sense in which the future is open: at present a number of future paths are possible, and the *now* closes off all but one as it moves along.²¹

This picture of reality contrasts sharply with that of the B theory. According to the B theory, there is no moving *now*, time does not flow, and past, present and future are equally real. Though we know more about the past than we do about the future, this is not because the future is unreal, but because

WHAT IS THIS THING CALLED METAPHYSICS?

knowledge is causal and there is little or no backwards causation in our universe. On the B theory, 'now' does not refer to a moving entity, it is a pure indexical. An utterance of 'now' simply refers to the time of utterance, just as an utterance of 'I' refers to the utterer, and an utterance of 'here' to the place of utterance. The present time is no more privileged over other times than I am privileged over other people or the place I currently occupy is privileged over other places.

The A and B theories thus constitute very different and incompatible views of reality. According to the (best version of the) A theory, time literally flows, and the future is unreal. According to the B theory, time does not flow, and past, present, future are equally real. A theorists tend to emphasize disanalogies between time and space; B theorists regard time and space as analogous dimensions.

I will not attempt to decide between the A and B theories here. Some see in Dummett's remarks above not just a defence of the A theory against McTaggart's objection, but a positive case in favour of the A theory and the view that time is not like space. Others take developments in modern physics – in particular, special-relativity theory – to show that there is no absolute and unique *now*. What is happening now is relative to one's frame of reference. What is present in one frame of reference may be past in another. This in turn is taken by some to undermine the A theory.²² But the dispute is still very much alive, and demonstrates how issues in metaphysics can draw on empirical results as much as on *a priori* reflection.

CONCLUDING REMARKS

In this chapter, we distinguished two ways of ordering events in time: in the A series and in the B series. The question arose as to which series, if either, was fundamental to time. The A theory holds that the A series is fundamental, the B theory that the B series is fundamental. We looked at two interesting arguments by McTaggart against each theory and found both arguments wanting. It is an open question which theory of time is correct, though some considerations raised in the next chapter favour the A theory.

STUDY QUESTIONS

- How is the A series to be distinguished from the B series?
- Why did McTaggart think that change was possible only on the A series?
- What is McTaggart's paradox?
- In what ways is time like space?
- Does the A theorist ascribe perspectival truths (such as 'it's raining now') an unwarranted ontological significance?

ANNOTATED FURTHER READING

- M. Dummett, 'A Defence of McTaggart's Proof of the Unreality of Time', in his *Truth and Other Enigmas* (Cambridge, Mass.: Harvard University Press), 1980, pp. 351–7. This is not a defence of McTaggart's paradox, but a plea for the paradox to be taken seriously. Dummett thinks that McTaggart's paradox, though it does not show time to be unreal, does oblige us to give up an intuitive principle about the description of reality.
- P. Horwich, *Asymmetries in Time* (Cambridge, Mass.: MIT Press) 1987, Chapter 2. A clear and engaging discussion of all the issues discussed here. Horwich takes McTaggart's paradox to refute the A theory, but disagrees with McTaggart over whether change is possible on the B series.
- R. Le Poidevin and M. MacBeath (eds), *The Philosophy of Time* (Oxford: Oxford University Press) 1993. A useful introduction by the editors, and reprints of classic discussions of time-related issues by (amongst others) McTaggart, Prior, Mellor, Shoemaker, Dummett, Lewis and Quinton.
- J. M. E. McTaggart, *The Nature of Existence*, Vol. II (Cambridge: Cambridge University Press) 1927, Chapter 33. (Reprinted as 'Time is not Real' in R. C. Hoy and L. N. Oaklander (eds), *Metaphysics: Classic and Contemporary Readings* (Belmont, Calif.: Wadsworth) 1991, pp. 43–54.) McTaggart's discussion oriented the twentieth- and twenty-first-century philosophical debate on time.
- H. Mellor, *Real Time* (Cambridge: Cambridge University Press) 1981, Chapter 6 (second edition, 1998). Like Horwich, Mellor takes McTaggart's paradox to undermine the A theory, and holds that change is possible on the B series. A vigorous defence of the B theory.
- D. C. Williams, 'The Myth of Passage', *The Journal of Philosophy* (1951) Vol. 48, pp. 457–72. (Reprinted in Hoy and Oaklander, *op. cit.*, pp. 55–63.) A spirited defence of the B theory of time and the 'block' universe.

WHAT IS THIS THING CALLED METAPHYSICS?

INTERNET RESOURCES

- H. Dyke (2005) 'Time, Metaphysics of', in E. Craig (ed.), *Routledge Encyclopedia of Philosophy*. Retrieved 31 May 2006 from <<http://www.rep.routledge.com/article/N123>>.
- N. Markosian (2002) 'Time', *The Stanford Encyclopedia of Philosophy (Winter 2002 edition)*, Edward N. Zalta (ed.). Retrieved 31 May 2006 from <<http://plato.stanford.edu/archives/win2002/entries/time>>.
- R. Le Poidevin (2005) 'Presentism', in E. Craig (ed.), *Routledge Encyclopedia of Philosophy*. Retrieved 31 May 2006 from <<http://www.rep.routledge.com/article/N120>>.
- L. Sklar (1998) 'Time', in E. Craig (ed.), *Routledge Encyclopedia of Philosophy*. Retrieved 31 May 2006 from <<http://www.rep.routledge.com/article/Q107>>.



TIME

Three puzzles

- **Introduction** 87
- **Prior's puzzle** 88
- **Time without change** 90
- **Time travel** 94
 - Backwards causation* 96
 - Causal loops* 96
 - Grandfather paradox: can a time traveller change the past?* 97
- **Concluding remarks** 99

INTRODUCTION

In this chapter I want to look at three interesting puzzles concerning time, two of which have implications for the debate about the nature of time. I will discuss Arthur Prior's 'Thank goodness that's over!' puzzle, Shoemaker on the possibility of time without change, and Lewis on the possibility of time travel. Both Prior's argument and the possibility of time without change lend support to the A theory.

WHAT IS THIS THING CALLED METAPHYSICS?

Time travel

Many television shows (e.g., some episodes of *Star Trek*), movies (e.g., *Terminator*) and science-fiction stories (e.g., H. G. Wells's *The Time Machine*) depict an individual travelling in time. What distinguishes a time traveller from an ordinary person is that the personal time of a time traveller differs from external time. A time traveller can travel 1,000 years into the past or the future in only one hour of personal time. That is, the time traveller's hair grows the amount it normally does in one hour, his watch tells him that one hour has elapsed, etc. Yet really 1,000 years have passed. A time-travelling world is strange indeed. If travel is into the past, it contains backwards causation, and may contain metaphysical oddities, such as causal loops. Some think that, if travel into the past is possible, a time traveller who formed the intention to change the past could go back in time and change the past. This is the premise of the movie *Terminator*. But this premise is false. No such intention can ever be fulfilled, since 'changing the past' is a contradictory description. There is thus an inconsistency at the very heart of *Terminator*.

PRIOR'S PUZZLE

In 1959 Prior posed the following puzzle for the B theory of time. He wrote:

One says, e.g., 'Thank goodness that's over', and not only is this, when said, quite clear without any date appended, but it says something which it is impossible that any use of a tenseless copula with a date should convey. It certainly doesn't mean the same as 'Thank goodness the date of the conclusion of that thing is Friday, June 15, 1954', even if it be said then. (Nor, for that matter, does it mean, 'Thank goodness the conclusion of that thing is contemporaneous with this utterance'. Why should anyone thank goodness for that?)¹

That is, the B theory does not have the resources to give utterances of 'Thank goodness that's over' their intended content. Such utterances can have that content only if the A theory is true.

However, Prior has slightly misstated the puzzle. The B theory is a theory of the nature of time, not a theory about the meaning of tensed sentences (i.e., sentences containing A series terms such as 'past', 'present' and 'future'). The B theory does not claim that a tensed sentence, e.g., my utterance of 'Hitler's death is past', means the same as its tenseless counterpart, *viz.*, 'that utterance of "Hitler's death is past" occurred after Hitler's death'. The claim of the B theory is that tensed truths (those expressed using 'past', 'present', 'future' and other A series terms) **reduce** to tenseless truths (those expressed



A. N. Prior (1914–69)

Arthur Norman Prior was born in New Zealand. He was Professor of Philosophy at Manchester University and later a fellow of Balliol College, Oxford. He died suddenly in 1969, only three years after arriving at Balliol. Prior made major contributions to the development of tense logic and the metaphysics of time. He explored analogies between the logic of necessity and possibility and the logic of past, present and future. The tense operators ‘it has been the case that p’ and ‘it will be the case that p’ were held to function analogously to the modal operators ‘it is necessary that p’ and ‘it is possible that p’. Thus an ordinary past-tensed sentence such as ‘Bill was bald’ was thought to be most perspicuously represented as ‘it has been the case that: Bill is bald’. Prior thought that tensed propositions were necessary for the expression of temporal facts, and so was an A theorist about time. Prior often drew on the work of ancient and medieval logicians and was by all accounts an indefatigable and inspiring teacher.

using B series terms and relations). Similarly, on the B theory, tensed facts reduce to tenseless ones. In general, the reduction of one fact to another need not involve any claim of synonymy between sentences.

Fortunately, as Hugh Mellor has observed, Prior’s puzzle can be restated so that it does engage directly with the B theory:

Suppose you have just had a painful experience, e.g., a headache. Now it is over, you say with relief ‘Thank goodness that’s over.’ What are you thanking goodness for? On the face of it, the fact that the headache is no longer a present experience, i.e., is now past. That is presumably why you made your remark after the pain, and not during or before it. Can this . . . still be explained [on the B theory]?²

So the puzzle is this. On the B theory, all temporal facts are tenseless (we can speak of tensed facts if we like, but they are reducible to tenseless ones). Tenseless facts are fixed and unchanging. They hold at all times. It can therefore never be appropriate to thank goodness for a tenseless fact at one time *rather than* another. But it is appropriate to thank goodness when a headache is past, and not when it is present or future. So if I thank goodness when a headache is past I am thanking goodness for a tensed fact (that the headache is past); but I am not thanking goodness for any tenseless fact; therefore tensed facts are not reducible to tenseless ones, and the B theory is false.

Mellor’s own response to his version of Prior’s puzzle is to claim that when I say ‘Thank goodness that’s over’ after the ending of a painful headache, I am not thanking goodness for any fact, hence I’m not thanking goodness for

WHAT IS THIS THING CALLED METAPHYSICS?

the tensed fact that my headache is past, but merely expressing relief (not relief about anything, just relief). However, this response is implausible. Even if relief does not always have an intentional object (cases where one is relieved but not relieved about anything in particular), it usually does, and it surely does in the kind of case we are considering. When I say ‘Thank goodness that’s over’ after the ending of a painful headache, I am relieved about something in particular, *viz.*, the fact that my headache is past. Hence, Mellor’s redescription of this case is not plausible.

What Prior’s example shows is that some of our attitudes presuppose changing temporal facts. I thank goodness because the fact of my headache’s being present gives way to the fact of its being past. Since, on the B theory, all temporal facts are tenseless and unchanging, the B theorist can make no sense of such an attitude.

Moreover, Prior’s puzzle generalizes. The problem is not just that the B theory can make no sense of utterances of ‘Thank goodness that’s over’. It can make no sense of many of our attitudes to our own past and future experiences. For example, we feel more and more relieved as some shameful episode recedes into the past. How are we to make sense of this in B-series terms? Why should I be relieved about the fact that, e.g., 2005 is five years after the shameful events of 2000? Or that 2006 is six years after those events? Such tautologies cannot ground feelings of relief.

More generally, consider the fact that we care more about future pains than we do about past ones. We much prefer that some unpleasant experience be behind us than yet to come. We are, as Parfit says, ‘biased towards the future’.³ How can we account for this attitude on the B theory? The relations ‘earlier than’ and ‘later than’ are perfectly symmetrical. Why should the fact that a pain is later than the time of utterance, rather than earlier, be a reason for caring more about it?

In contrast, the A theorist can at least say the following. Time has an intrinsic direction from past to future, and this underwrites our asymmetrical concern. It is reasonable to care more about future pains because they are moving towards us, and it’s reasonable to care less about past pains since they are moving further and further away from us. But what can the B theorist say? Prior’s puzzle is thus just one of a family of puzzles that pose a special difficulty for the B theory.

TIME WITHOUT CHANGE

What is the connection between time and change? It is undeniable that change implies time: change is a temporal process. But does time imply change? That is, is it possible for there to be a period of time during which nothing in the entire universe changes?⁴ We are not asking whether there has been or will be



Time without change

Many philosophers (Aristotle, Hume and McTaggart, to name but three) have thought that time implies change. That is, if time passes, then somewhere in the universe some amount of change, however small, must occur. Given how McTaggart understood ‘time’ and ‘change’, it is a definitional truth, not in need of argument, that time implies change. Aristotle did offer an argument for the impossibility of time without change, though not a good one. He argued that there cannot be time without change because awareness of the passing of time necessarily involves awareness of change. But this ignores the possibility that time might pass without our being aware of it passing, and during such a period there may be no change. Of course, we cannot be directly aware of changeless time, since our continued awareness is itself a change. But perhaps we can have indirect evidence that a period of change-free time has elapsed. Shoemaker’s imaginary world nicely illustrates this possibility. In showing how people in his world might come reasonably to believe that a year of time without change has elapsed, Shoemaker lends further support to the thesis that time without change is a logical possibility.

such a period of change-free time in our universe. We are simply asking whether the idea as much as makes sense, whether it is logically or conceptually possible for there to be a period of change-free time. And if it is possible, could people in such a possible world be able to predict when a period of change-free time has elapsed? Sydney Shoemaker’s interesting paper ‘Time without Change’ makes an ingenious case for answering this question positively.⁵

Some famous philosophers have insisted that there cannot be time without change – Aristotle, Hume and McTaggart, to give three illustrious examples. However, as will be recalled from the previous chapter, McTaggart counted an event’s becoming more past as a change. On this view, the mere passage of time counts as change, and so time without change would be impossible by definition.

In his discussion, Shoemaker does not count these **McTaggart changes** as genuine changes, and understands ‘change’ as ‘ordinary change’: an object changes when it changes its colour, height, weight, momentum, position in space, etc. So understood, the question of whether there can be time without change no longer immediately answers itself. It is also the sense of ‘change’ in which Aristotle and Hume denied that there could be a period of time without change.

Why think that time necessarily involves change (in the ordinary sense of ‘change’)? Shoemaker quotes Aristotle’s ground:

when the state of our own minds does not change at all, or we have not noticed it changing, we do not realise that time has elapsed, any

WHAT IS THIS THING CALLED METAPHYSICS?

more than those who are fabled to sleep among the heroes in Sardinia do when they are awakened; for they connect the earlier 'now' with the later and make them one, cutting out the interval because of their failure to notice it.⁶

Though not transparent in meaning, these remarks suggest the following train of thought. There cannot be time without change because our awareness of the passing of time necessarily involves awareness of change (either in our own minds or in the observable world).

However, this is not a convincing argument. To start with, it is structurally flawed. Its conclusion is metaphysical (time necessarily involves change), yet its premise is epistemic (concerning a condition on our awareness of time passing). How could a metaphysical conclusion follow from such an epistemological premise? Is Aristotle assuming that time passes only if we are aware of it passing? Moreover, Aristotle's premise – our awareness of the passing of time necessarily involves awareness of change – is ambiguous. It ignores the distinction between direct and indirect awareness. Obviously, I cannot be directly aware of time without change, since my awareness would itself constitute change, but why could I not have indirect evidence that a certain period of change-free time has elapsed? To use Shoemaker's analogy, I cannot directly verify 'at some time t , the universe contains no minds at t ', since any attempt to do so would be self-undermining, but I can surely have indirect evidence, from cosmology, that justifies me in accepting this statement.⁷

Shoemaker then proceeds to provide an example of a world in which there is a period of time without change and in which people in that world would have good reason to believe that there has elapsed a certain period of time without change. Shoemaker imagines a world divided into three spatial regions: A, B and C. There is typically interaction between the people in all three regions, and people can freely move from one region to the others. But there is the following oddity: every so often one of the regions 'freezes' for a period of one year. Thus, e.g., when A freezes, the people in B and C can see that no events occur in A. When the year is over, everything in A jumps back to life. People in A continue conversations with each other as if no time had elapsed. Of course, things would look odd to any A occupant who was looking into B or C prior to the freeze. Just after the end of the freeze, it will appear to such an occupant as if many big changes have occurred instantaneously. However, once the B and C occupants explain what has happened, the A occupant's bewilderment will be somewhat lessened. What happens to the A region also happens periodically to the other regions, and such freezes are verified by people in the unfrozen regions.

We do not yet have an example of time without change. As we have described this world so far, whenever there is a frozen region, there are always two unfrozen regions in which there is change. However, the more observant



members of this world begin to notice a certain regularity to the freezes: A freezes every three years, B freezes every four years, and C freezes every five years. From this information they can work out, by elementary arithmetic, that every sixty years there is a global freeze: A, B and C are frozen simultaneously for one year.⁸ Now we have what we were after: a world in which there is a period of time without any change, and in which inhabitants of that world can predict when there will be such a period.

Shoemaker's world seems possible, but one question is pressing. What causes the end of a freeze? In the case of a local freeze, say in A, it may be events in B or C that cause the freeze in A to end. But what causes the end of a global freeze? When the entire universe is frozen, what could cause it to unfreeze? Not events occurring during the year of global freeze, for there are none. Nor can the cause be any event prior to, or simultaneous with the beginning of, the freeze, for then the freeze would be over as soon as it had begun.

However we answer this question we must, it seems, give up the principle 'no action at a temporal distance'. That is, the kind of causality at work in our imaginary world must violate the following principle:

- (P) If an event is caused, then any temporal interval preceding it, no matter how short, contains a sufficient cause of its occurrence.

As Shoemaker elaborates: 'To suppose (P) false is to suppose that an event might be caused directly, and not *via* a mediating causal chain, by an event having occurred a year earlier, or that an event might be caused by such and such's having been the case for a period of one year'.⁹ Although we take our world to be governed by principle (P) – that is, if A at t_1 causes B at t_2 we expect there to be a spatio-temporally continuous chain of causes and effects linking A with B – Shoemaker sees no conceptual barrier to giving up (P). That is, there could be a world, such as the one we are imagining, containing a relation worth calling 'causation', which violates principle (P).

In a (P)-violating world, the mere passage of time has causal efficacy. Given that causally efficacious changes are genuine, mere temporal changes in such a world are genuine. In other words, McTaggart changes, which Shoemaker earlier excluded from the realm of genuine changes, have to be regarded as genuine, at least in (P)-violating worlds. The passage of time in such a world is causally efficacious. In which case, Shoemaker's imaginary world is not one in which there is time without change.

Nonetheless, Shoemaker has succeeded in describing a world in which there is time without (ordinary) change, and in which periods of change-free time can be predicted by inhabitants of that world. It's just that, in such a world, ordinary change is not the only kind of genuine change. There can be periods of time without ordinary change, but not periods of time without change.

WHAT IS THIS THING CALLED METAPHYSICS?

We can end by noting that the thesis Shoemaker has made so plausible – that there can be periods of time without ordinary change – poses a difficulty for the B theorist. How, for the B theorist, can a year of change-free time pass if there are no events in that year to stand in B-series relations to other events? The passage of time, for the B theorist, just consists in events ordered in the B series. As with Prior’s puzzle, we have another consideration that tells against the B theory and in favour of the A theory.

TIME TRAVEL

For many, time travel is the most exciting topic in the philosophy of time. From a philosophical point of view, the most fundamental questions about time travel are: is time travel possible, and what would a time-travelling world be like? In asking these questions we are not assuming that time travel has or ever will occur in this world. We are merely asking whether it is possible. And by ‘possible’ we mean possible in the widest possible sense: i.e., logically or conceptually possible. We are not concerned with whether time travel is physically possible (i.e., compatible with the actual laws of nature), though that is an interesting question, but whether it as much as makes sense. We can all enjoy stories and movies involving time travellers, but do such stories represent real possibilities? In answering this question I will draw on David Lewis’s excellent article ‘The Paradoxes of Time Travel’.¹⁰

David Lewis (1941–2002)

David Lewis taught briefly at UCLA before moving to Princeton University in 1970. Lewis wrote on many areas of philosophy, but is best known for his work on counterfactual conditionals and the philosophy of modality (possibility and necessity). According to Lewis, a counterfactual conditional of the form ‘if A had been the case, B would have been the case’ is true just if some possible world where A and B are both true is ‘closer’ to our world than any world where A and ~B are both true. Lewis takes ‘closeness’ to be matter of similarity and takes a realist view of possible worlds: possible worlds exist in just the way that our world exists. There is nothing special or privileged about the actual world, since each world is actual to its inhabitants. Some have complained that ‘closeness’ cannot be understood in terms of similarity, and many have complained that Lewis’s realist view of possible objects and possible worlds is incredible (since it asks us to believe that golden mountains and talking donkeys exist in just the way that our mountains and donkeys exist).



First, we must define ‘time travel’. As Lewis says: ‘[i]nvariably, it [time travel] involves a discrepancy between time and time.’¹¹ Time travel can be to the past or to the future: in either case, the time traveller’s journey may have taken, e.g., an hour, yet he may have ended up hundreds of years into the past or the future. The idea of a discrepancy between time and time may sound incoherent, yet Lewis avoids incoherence:

by distinguishing time itself, *external time* as I shall call it, from the *personal time* of a particular time traveller: roughly, that which is measured by his wristwatch. His journey takes an hour of personal time, let us say: his wristwatch reads an hour later at arrival than at departure. But the arrival is more than an hour after the departure in external time.¹²

It is important to realize that the distinction between personal and external time is not that between two dimensions of time. Lewis writes of personal time that it:

isn’t really time, but it plays the role in [the time traveller’s] life that time plays in the life of a common person. . . . We may liken intervals of external time to distances as the crow flies, and intervals of personal time to distances along a winding path. The time traveller’s life is like a mountain railway. . . . [w]e are not dealing here with two independent dimensions. Just as distance along the railway is not a fourth spatial dimension, so a time traveller’s personal time is not a second dimension of time.¹³

The distinction between personal and external time allows us to make sense of talk of personal identity in the case of a time traveller. We want to say that the time traveller who steps into his time machine in 2006 is the *same person* as the man who steps out of the machine, one hour later in personal time, in 1900. It is often thought that relations of **mental and/or bodily continuity** make for personal identity over time. (See Chapter 8.) In the case of a non-time traveller the phrase ‘over time’ is not ambiguous. But in the case of a time traveller it is ambiguous between ‘over external time’ and ‘over personal time’. Lewis suggests that we should regard the personal identity of a time traveller as consisting in mental and/or bodily continuity with respect to personal time. This allows us to agree with the intuitive verdict that the person stepping into the time machine in 2006 is the same as the person who steps out of the machine in 1900.¹⁴

Thus far, we have drawn an important distinction between personal time and external time which allows us, *prima facie* at least, to make sense of time travel. However, there may still lurk puzzles and paradoxes hidden in the very

WHAT IS THIS THING CALLED METAPHYSICS?

idea of time travel and of agents (persons) travelling in time. These puzzles and paradoxes always tend to be presented with respect to travel into the past and not the future. The question we have to address is whether these puzzles and paradoxes present genuine objections to the possibility of time travel. I will argue that they do not, and thus that there is no conceptual barrier to travel into the past or future. Here are some of the puzzles.

Backwards causation

Time travel into the past necessarily involves backwards causation (with respect to external time). As Lewis says, of a traveller into the past: '[y]ou may punch his face before he leaves, causing his eye to blacken centuries ago.'¹⁵ Or again, the time traveller pressing the ignition button in his time machine in 2006 causes the time machine to arrive in 1900. Is it an objection to the very possibility of travel into the past that it requires backwards causation? Only if the idea of backwards causation, the idea of an effect preceding its cause, is incoherent. Aside from theories which simply stipulate that an event counts as a cause only if it precedes its effect, none of the leading theories of causation rules out the possibility of backwards causation. Hence it is not reasonable to object to travel into the past simply on the grounds that it inevitably involves backwards causation. Indeed, if travel into the past is not ruled out by other considerations, we could appeal to the possibility of time travel as an argument for the possibility of backwards causation.

Causal loops

Aside from backwards causation, travel into the past also gives rise to the possibility of causal loops. Causal loops are

closed causal chains in which some of the causal links are normal in direction and others are reversed. . . . Each event on the loop has a causal explanation, being caused by events elsewhere on the loop. This is not to say that the loop as a whole is caused or explicable. It may not be.¹⁶

A nice example of a causal loop involves the transfer of information. Imagine a time traveller who goes back in time a few years and talks to his earlier self. They discuss time travel, and:

in the course of the conversation his older self told his younger self how to build a time machine. That information was available in no



other way. His older self knew how because his younger self had been told and the information had been preserved [in memory]. His younger self knew, after the conversation, because his older self [had told him]. But where did the information come from in the first place? Why did the whole affair happen? There is simply no answer.¹⁷

Are causal loops impossible? If they are, then travel into the past must be impossible too. However, there is no reason to think that causal loops are impossible. We are happy to entertain the possibility of many uncaused and inexplicable events: 'God, or the Big Bang, or the entire infinite past of the universe, or the decay of a tritium atom'.¹⁸ If these are possibilities, why not also causal loops? The possibility of causal loops shows that worlds in which people travel into the past are strange and very unlike our world, but not that such worlds are impossible.

Grandfather paradox: can a time traveller change the past?

One of the most famous objections to time travel is that a time traveller could change the past. Since it is impossible to change the past, it is concluded that time travel is impossible too.

It is certainly true that changing the past is impossible. To change the past is to make it true that some event that happened didn't happen or to make it true that some event that didn't happen did happen. But it can never be true that some event both happened and didn't happen – not even God can make that true. There is nothing special about the past in this respect. It is equally impossible to change the present or the future. No one can make it the case that an event happens and does not happen, or that some event will happen and will not happen. Of course, we can affect or bring about the future (by choosing to act in certain ways now), but we cannot change it, in the sense just defined. Those who believe in the possibility of travel into the past certainly commit themselves to the possibility of affecting or bringing about the past. But are they committed to the possibility of a time traveller changing the past?

Lewis thinks not, but begins by outlining the case for thinking that a time traveller can change the past with the following example:

Consider Tim. He detests his grandfather, whose success in the munitions trade built the family fortune that paid for Tim's time machine. Tim would like nothing so much as to kill Grandfather, but alas he is too late. Grandfather died in his bed in 1957, while Tim was a young boy. But when Tim has built his time machine and travelled to 1920, suddenly he realizes he is not too late after all. He buys a rifle; he

WHAT IS THIS THING CALLED METAPHYSICS?

spends long hours in target practice; he shadows Grandfather to learn the route of his daily walk.¹⁹

Tim can kill Grandfather: he has a high powered rifle; he is a good shot; weather conditions are perfect, etc. Yet Tim cannot kill Grandfather: Grandfather died in his bed in 1957, so he cannot have died in 1920. Consistency demands, despite Tim's best efforts, that he somehow fail to kill Grandfather. Why does he fail? 'For some commonplace reason. Perhaps some noise distracts him at the last moment, perhaps he misses despite all his target practice, perhaps his nerve fails, perhaps he even feels a pang of unaccustomed mercy.'²⁰ Hence, concludes Lewis, it is wrong to think that a time traveller can change the past.

Three comments are in order here. First, it may be thought that Lewis has just replaced one contradiction with another. Has the contradiction 'Grandfather dies in 1920 and in 1957' not been replaced by the contradiction 'Tim both can and cannot kill Grandfather'? In which case, travel into the past still implies a contradiction.

Lewis has a nice reply to this objection. There is no contradiction since 'can' is equivocal. Relative to one set of facts, Tim can kill Grandfather (e.g., facts about Tim's rifle, his shooting ability, the weather conditions, and so on). But relative to another, more inclusive, set of facts (including the fact that Grandfather was not killed in 1920), Tim cannot kill Grandfather. There would only be a contradiction if Tim can and cannot kill Grandfather, relative to the same set of facts. But that is not something to which a defender of the possibility of time travel is committed.

Second, Lewis's story is a causally self-undermining one: Tim is attempting to eliminate one of the causes of his own existence. This makes Tim's attempt doubly impossible. It is self-undermining as well as an attempt to change the past. But the self-undermining aspect is inessential to Lewis's solution. Even if Tim tried to assassinate Grandfather's partner, he would also fail, for the very same type of reason (given that, e.g., the partner lived until 1950).

Third, Lewis is quite right that Tim's story, if it is to be consistent, must continue along the lines he suggests. Somehow Tim fails: his gun jams; he shoots someone else; he misses, etc. But there is something unsatisfying about this. Suppose that day after day, Tim attempts to shoot Grandfather, and each time something goes wrong. One day it's too windy, the next day his finger slips, the following day his gun jams, and so on. Each mishap has an explanation, but the entire sequence of them does not. Given that Tim has, by all usual criteria, the desire and the means to kill Grandfather, the resulting sequence of mishaps would constitute a highly improbable string of events. Similarly for any other time traveller who attempted to 'change the past'. This is certainly odd, but perhaps not impossible. Such improbable sequences do not show the world of a time traveller to be impossible, but they do confirm, what we know already, that such a world is very different from ours.

Finally, it's worth pointing out that the topic of time travel intersects with the A theory/B theory debate discussed in the previous chapter. If presentism is true, and only the present is real, time travel ought to be impossible. If past and future are both unreal, where would the time traveller go? If Broad's view is true – past and present real, future unreal – travel into the future ought to be impossible. (Hence, if we were visited by travellers from the future, this would confute both these versions of the A theory.) Only the B theory, which holds that past, present and future are equally real, allows for travel into the past and future. Some may take our discussion of time travel as evidence in favour of the B theory, but this is perhaps a question-begging strategy. Better to conclude: if the past is real, then there is no conceptual barrier to travel into the past (likewise for the future).

CONCLUDING REMARKS

We have looked at three interesting puzzles about time – Prior's puzzle, time without change and time travel – and we have arrived at some interesting conclusions. Prior's puzzle yielded a nice argument in favour of the A theory. Shoemaker successfully described a possible world in which inhabitants of that world could know that time passed for one year without any (ordinary) changes taking place. This result also favours the A theory since it is unclear how the B theory can make sense of time passing in the absence of events (changes) which can be ordered in the B series. Finally, our discussion of time travel showed that, assuming the reality of the past, a world in which travel into the past occurs is very different from our world, but possible nonetheless.

STUDY QUESTIONS

- Could the B theorist reasonably claim that our temporally biased attitudes are irrational?
- Can the A theorist explain satisfactorily why we care more about future experiences than about past ones?
- Can we generate a spatial analogue of Prior's reasoning ('Thank goodness that's not happening here!') in defence of an 'A theory' of space? Would this cast doubt on Prior's original argument?
- Is Shoemaker's imaginary world really a possible world? If so, does it pose a problem for the B theory?

WHAT IS THIS THING CALLED METAPHYSICS?

- Are there any genuine paradoxes of time travel?
- Could a time traveller be his own father?

ANNOTATED FURTHER READING

- P. Horwich, *Asymmetries in Time* (Cambridge, Mass.: MIT Press) 1987, Chapter 7. A fairly technical, but still useful defence of Kurt Gödel's claim that travel into the past is physically possible.
- D. Lewis, 'The Paradoxes of Time Travel', *American Philosophical Quarterly*, Vol. 13, No. 1, 1976. Reprinted in his *Philosophical Papers*, Vol. II (Oxford: Oxford University Press) 1986, Chapter 18. The classic philosophical treatment of time travel. Lewis's article is very readable, and free of technicality.
- H. Mellor, *Real Time* (Cambridge: Cambridge University Press) 1985. The most sustained defence of the B theory available, containing a response to Prior's 'Thank goodness that's over' argument.
- A. N. Prior, 'Thank goodness that's over', *Philosophy*, Vol. 34, 1959. Clear statement of a now classic problem for the B theory.
- S. Shoemaker, 'Time without Change', in R. Le Poidevin and M. MacBeath (eds), *The Philosophy of Time* (Oxford: Oxford University Press) 1993. A clear and thorough presentation of the case for the possibility of time without change.

INTERNET RESOURCES

- P. Horwich (1998) 'Time travel', in E. Craig (ed.), *Routledge Encyclopedia of Philosophy*. Retrieved 31 May 2006 from <<http://www.rep.routledge.com/article/Q108>>.
- F. Arntzenius and T. Maudlin (2005) 'Time Travel and Modern Physics', *The Stanford Encyclopedia of Philosophy (Summer 2005 edition)*, Edward N. Zalta (ed.). Retrieved 31 May 2006 from <<http://plato.stanford.edu/archives/sum2005/entries/time-travel-phys>>.

7 FREE WILL

• Introduction	103
• Fatalism	104
• A more sophisticated fatalist argument	107
• Theological fatalism	110
• Free will and determinism	111
• The case for incompatibilism	112
• A compatibilist reply	114
• Is the notion of free will incoherent?	115
• Concluding remarks	116

INTRODUCTION

In this chapter, we will be concerned with two attacks on the thesis that we have free will. It may seem outrageous to attack this thesis, since it seems to us that we act freely virtually all the time. However, since the beginning of philosophy our free will has been called into question.

The first attack on free will consists of a collection of arguments and considerations known as fatalism. The arguments of the standard fatalist are purely logical or *a priori* in character: they rest on no empirical premises. There is, however, another version of fatalism – theological fatalism – which rests on assumptions about God's existence and nature.

The second line of attack on free will rests on an empirical premise – that of determinism. According to this second attack, our universe is deterministic and the truth of determinism is incompatible with, and hence excludes, free will. The thesis of determinism, if true, is a contingent, empirical truth. Though the arguments of the fatalist are less than compelling, the attack on free will stemming from determinism is more unsettling. In a further twist, Galen Strawson has recently argued forcefully that free will is an incoherent notion, hence compatible with neither determinism nor indeterminism.

The issue of free will is important, not just as an issue in metaphysics, but also because it is generally assumed that moral responsibility requires free will. Hence, if free will is an illusion, so too is moral responsibility, thereby undermining those social, moral and legal practices that presuppose such responsibility.

FATALISM

The conclusion of fatalist reasoning, as its name suggests, is that we are prisoners of fate: we cannot do anything other than we actually do. The simplest version of fatalism draws on two *prima facie* plausible principles:

- (i) There will be only one actual future.
- (ii) For any proposition P, if P is true now, it was true at any past time that P.

The first principle is intended to capture the uncontroversial claim that there are not two or more actual futures. There is only one way the universe will evolve after any given time, even though we may not know what way that is. The second principle is a version of the timelessness of truth: if it is true now that, e.g., Tony Blair won the 2005 UK General Election, then it was true at any arbitrary past time, say 900 BC, that Blair would win the election in 2005.

How might these principles generate a fatalist conclusion? Suppose P is a future contingent statement. It is about some time in the future, and it is not necessarily true. Suppose that P concerns one of my actions. In particular, let P = I will vote Liberal next year. Now there is only one actual future (by principle [i]), and suppose that I will indeed vote Liberal next year. So it is true now that I will vote Liberal next year. By principle (ii), at any arbitrary past time, say 1800, it was true then that I will vote Liberal next year. But in that case, concludes the fatalist, it is inevitable that I will vote Liberal next year.

Fatalism

Fatalists hold that it was true 100 years ago that I would type these words today. Theological fatalists hold that God always knew I would type these words today. They conclude that it was inevitable that I type these words today. I could not but have typed them. This action, indeed all actions by all people, are thus unfree. Freedom is an illusion. Fortunately for us, the fatalists' reasoning is confused. They want us to think that some fact about the past or about God's prior knowledge necessitates or makes it true that I will type these words today. But intuitively the direction of dependence is the other way round. It was true 100 years ago that I would type these words today because I freely type them now. Does this not make a fact about the past depend on a fact about the present? No, because the fact about the past is only superficially or grammatically about the past. It is really a fact about the present disguised as a fact about the past. Similarly, if God knew I would type these words today, that is because I freely chose to type them today. The fatalist's argument is broken-backed.

Determinism

The thesis of determinism has traditionally been thought to threaten freedom of thought and action. If determinism is true then, given the laws of nature and the state of the universe at any arbitrary past time, it is determined that you read these words now. It is physically impossible that you not read these words, and so your reading these words is not free. Compatibilists hold that determinism does not compromise our freedom. Libertarians hold that determinism does compromise our freedom and, for that reason, determinism is false. But is it not for science to tell us whether determinism is true? Anyway, how can indeterminism provide a congenial home for free will? Carefully reasoned and predictable actions hardly seem the result of indeterministic or partly random happenings. Others hold that compatibilists and libertarians are both wrong. Freedom is incompatible with determinism and with indeterminism. That is, free will is logically impossible. I cannot be the 'final root of all my determinations' in the way required by the concept of free will. Ascriptions of moral responsibility, in as much as they presuppose free will, are without foundation.

I am powerless to do otherwise. Similarly for any other action of mine in the future that I take to be free. Freedom is an illusion.

Thus from:

- (1) I will vote Liberal next year; and
- (2) in 1800 it was already true that I was going to vote Liberal next year;

the fatalist concludes

- (3) it is inevitable that I will vote Liberal next year.

It is not only future actions that are unfree. We can run the very same reasoning, modifying the tenses, to show that none of my past actions were free either. Thus:

- (1*) I voted Liberal last year; and
- (2*) in 1800 it was already true that I was going to vote Liberal last year;

from which the fatalist concludes

- (3*) it was inevitable that I was going to vote Liberal last year.

WHAT IS THIS THING CALLED METAPHYSICS?

Similarly for any other past action of mine which I took to be freely executed. Hence the fatalist concludes, generalizing from my case to that of others, no action of any agent is ever free.

This conclusion is outrageous. How might we resist it? David Lewis offers the following diagnosis:

Fatalists – the best of them – are philosophers who take facts we count as irrelevant in saying what someone can do, disguise them somehow as facts of a different sort that we count as relevant, and thereby argue that we can do less than we think – indeed, that there is nothing at all that we don't do but can. . . . [A fact such as (2)] . . . is an irrelevant fact about the future masquerading as a relevant fact about the past, and so should be left out of account in saying what, in any ordinary sense, I can do.¹

In assessing what someone can do now, facts about the past are relevant. Thus whether I can now swim is determined by past facts, such as, whether I ever took swimming lessons. Thus the past fact:

(4) I never took any swimming lessons

is a determinant of the truth value of:

(5) I can now swim.

The truth of (4) explains the falsity of (5). Statement (4) is a fact entirely about the past; (5) is a fact about the present. The fatalist sees the relation between (2) and (1) as analogous to that between (4) and (5). Just as the past fact that I never took any swimming lessons explains that and why I cannot now swim, so, the fatalist thinks, the past fact that in 1800 it was true that I was going to vote Liberal next year explains that and why I cannot but vote Liberal next year. But therein, according to Lewis, lies the fatalist's mistake: (2) looks like a fact about the past, but it is not. It is really a fact about next year disguised to look like a fact about 1800.

Lewis's point is plausible. We do not think that (2) is about 1800 in the way that, e.g.,

(6) in 1800 the population of London was greater than 1 million

is about 1800. Nonetheless, one might feel that more needs to be said. Since Lewis is not disputing the truth of (1) and (2), why exactly are we not entitled to infer (3)? What if the fatalist replies that, for him, (2) is relevant to the assessment of (1), and (1) and (2) imply (3). How do we then respond?

Some commentators have pointed out that the fatalist makes a peculiar assumption about the direction of dependence between (1) and (2).² In order to infer (3) from (1) and (2), the fatalist must assume that my voting Liberal next year *depends upon* the fact that in 1800 it was already true that I was going to vote Liberal next year. But intuitively the direction of dependence is precisely the opposite: its being true in 1800 that I will vote Liberal next year depends on the fact that I will vote Liberal next year. Of course, if (2) depends on (1), it looks as if a past fact depends on a future fact. It is here that Lewis's point is relevant: (2) is not really a past fact, it is a future fact in disguise. Once we realize that (2) depends on (1) and not *vice versa*, and that (2) is not a fact about the past, there is no reason to think that (1) and (2) imply (3). We can therefore happily regard the fatalist's (1)–(3) argument as invalid.

A MORE SOPHISTICATED FATALIST ARGUMENT

Might the fatalist present his argument in a more sophisticated form? Let us consider Michael Dummett's discussion of fatalism in 'Bringing about the Past'. He writes:

The standard form of the fatalist argument was very popular in London during the bombing. The siren sounds, and I set off for the air-raid shelter in order to avoid being killed by a bomb. The fatalist argues, 'Either you are going to be killed by a bomb or you are not going to be. If you are, then any precautions you take will be ineffective. If you are not, all precautions you take are superfluous. Therefore it is pointless to take precautions'.³

Dummett is discussing this argument because he noticed that the main argument against the rationality of attempting to bring about the past, for example, by retrospective prayer, is exactly similar to the fatalist's argument, except of course for the reversal of tense. Thus, in response to a father (unaware of his son's fate) praying to God now in order to help bring it about that his son was saved from drowning yesterday, Dummett imagines the following being said:

Either your son has drowned or he has not. If he has drowned, then certainly your prayer will not (cannot) be answered. If he has not drowned, your prayer is superfluous. So in either case your prayer is pointless: it cannot make any *difference* to whether he has drowned or not.⁴

Dummett's thought is that since we do not accept the fatalist's argument nor should we accept this argument, thus allowing us to at least make sense

WHAT IS THIS THING CALLED METAPHYSICS?

of the idea of doing something now in order that something else should have happened earlier (at least when one does not know whether the earlier event has occurred).

Of course, it might be objected that the two arguments are not analogous precisely because of the difference in tense. Since the past is real, the argument against retrospective prayer is sound. Since the future is unreal, the fatalist's argument is unsound. We can call this the 'Aristotelian solution'.⁵

In order to evaluate this solution, let's first write out the fatalist's argument:

- (1) Either you are going to be killed by a bomb or you are not going to be killed by a bomb.
- (2) If you are going to be killed by a bomb, any precautions you take will be ineffective.
- (3) If you are not going to be killed by a bomb, any precautions you take will be superfluous.

So:

- (4) It is pointless to take precautions.

According to the Aristotelian solution, future contingent statements lack a truth value. The Aristotelian solution presupposes the unreality of the future: there is no reality to make future contingent statements true or false now, though future necessary statements are true or false now. It is natural for the Aristotelian to characterize the unreality or 'openness' of the future in terms of a tree model. At the present moment, there are many possible future branches (consistent with how things are now); as the present (or moving *now*) glides along it lops off all but one.⁶ (See Chapter 5.)

The Aristotelian may then hold that a statement about the future is true if it is true in each branch; false if it is false in each branch; neither true nor false otherwise. Thus, a future contingent statement, such as 'it will rain tomorrow', is not now either true or false since it is true in some branches and false in others. In contrast, the future necessary truth 'either it will rain tomorrow or it will not' is now true since it is true in every branch, and the future necessary falsehood 'it will and will not rain tomorrow' is now false since false in every branch. It should be noted that this account violates classical logic since a disjunction ('either it will rain tomorrow or it will not') is held to be true though neither disjunct is true.

How does the Aristotelian solution find fault with the fatalist's argument? Not by denying its first premise. But although (1) is true, neither of its disjuncts is true, and so the antecedents of (2) and (3) are not true. The Aristotelian can now hold either that **conditionals** with antecedents that lack a truth value are themselves untrue (in which case the fatalist's argument has two

untrue premises) or he can hold that the argument from disjunctive syllogism (P or not- P ; if P then R ; if not- P then R ; so R) is not valid if ‘ P ’ and ‘not- P ’ lack a truth value (in which case the fatalist’s argument is invalid). Thus, the fatalist’s argument is either unsound or invalid. However, on the Aristotelian solution, there is no analogous fault in the argument against retrospective prayer: ‘he has drowned’ and ‘he has not drowned’ are both determinate in truth value.

But there is a deeper flaw in the fatalist’s argument. Even if future contingent statements have a truth value, premise (3) is not remotely plausible. It is sophistical to argue that if you are not going to be killed, any precautions you took were superfluous. It might be the case that you were not killed precisely because you took precautions.

In order for it to be true that precautions were superfluous, the following **counterfactual conditional** would have to be true: if you had taken no precautions, you would (still) not have been killed. But the truth of this counterfactual does not follow merely from the truth of ‘you will not be killed’. The principle ‘if P then had Q been the case, P would (still) have been the case’ (for arbitrary P and Q) is an obviously false principle. Suppose I was in a car crash, but saved because I was wearing a seat belt. It would be crazy to reason: I wasn’t killed, so had I not been wearing a seatbelt, I still wouldn’t have been killed.

What does follow from any truth P , at least on some interpretations of the **indicative conditional**, is: if Q then P . That is, ‘if P then if Q then P ’ is a tautology, assuming that ‘if . . . then’ in English is understood as the **material conditional**. On this interpretation of indicative conditionals, the following is true: if you will not be killed, then if you take no precautions, you will not be killed. But this does not imply that precautions were superfluous. As we have seen, superfluity requires, not the truth of an indicative conditional, but the truth of a counterfactual conditional, and that counterfactual does not follow merely from the truth of ‘you will not be killed’.

Consequently, the case for (3) is fallacious, and we should simply deny that premise. The mere truth (if true) of ‘you will not be killed’ does nothing to show that any precautions you took were superfluous. For the same reason, we should reject the earlier argument against the pointfulness of retrospective prayer. The premise ‘if he has not drowned, your prayer is superfluous’ is open to precisely the objection just lodged against premise (3).

Nonetheless, there may be unusual circumstances in which a fatalistic attitude to one’s life seems justified. Here is one familiar fantasy: you are browsing in the library one day, and come across a strange-looking volume entitled *This is Your Life*. On reading it, you realize that this really is the book of your life. It contains accurate descriptions of your upbringing, career, appearance, behaviour, even your innermost thoughts. Here in black and white are truths about

WHAT IS THIS THING CALLED METAPHYSICS?

yourself that (surely!) no one else could possibly have known. The book looks old, and the date of its printing is well before you were born. You also notice that the book does not stop at the present time; indeed that is only the halfway mark. Impatiently, you begin to read about your future life . . .

Would it not be natural to adopt a fatalistic attitude to one's life in this circumstance? And is that not because there are truths about how your life actually will go, as the fatalist claims? We have seen that fatalism is a logical trick, so the mere existence of truths about your life cannot justify a fatalistic attitude towards it. However, you may think that the only way the author of the book could *know* all these truths about you was if he was also the author of your life, i.e., he is the controller and you are his puppet. In that case, a 'fatalistic' attitude is justified, since none of your actions are your own and hence none are free. But what justifies this attitude is not the existence of truths about you, but the fact that your thoughts and behaviour are controlled by someone else. Consequently, fatalism, as understood here, is not vindicated by this fantasy.

THEOLOGICAL FATALISM

There is another version of fatalism, much discussed in the Middle Ages, which arises from the supposed tension between God's omniscience and human free will. If God knows all truths (past, present and future), then he knows what every human being will do. How then can any human action be free? Clearly this puzzle only arises within a theological context in which it is assumed that God exists and that he is omniscient. Nonetheless, from within that context, does theological fatalism give rise to any new puzzle?

A. J. Ayer thinks not:

If the fact that someone knew what I was going to do tomorrow would not make it necessary [inevitable] that I should do it, then the fact that someone knew what I was going to do, not only tomorrow but all the days of my life, would not make these actions necessary either. . . . Neither does it make any difference whether the person to whom the foreknowledge is attributed is taken to be human or divine. . . . All these considerations are irrelevant.⁷

According to Ayer, if its now being true that I will vote Liberal next year does not make it inevitable that I will, then knowledge of that truth (divine or otherwise) does not make it inevitable either. Nor are matters changed if many more truths about me are known.

There are presentations of theological fatalism which may seem to introduce a novel consideration. Steven Cahn writes:

If it is true that I will perform a particular action, then God, who knows all truths, knows I will perform that action. But if I could refrain from that action, presumably I could confute God's knowledge, which is impossible. But if I cannot refrain from the action, it is not free.⁸

But does this really raise a new point? Could we not have presented the same reasoning in a non-theological context? For example: 'If it is true now that I will perform a certain action, A, then I will perform that action. But if I could refrain from action A, presumably I could change the past, which is impossible. But if I cannot refrain from the action, it is not free.'

The reply to this argument is that it's not true that if I could refrain from action A, I could change the past. In this world I do action A. In a different possible world, in which I do B instead, it's true in that world that I will do B. These are two possible worlds containing different truths and different actions. In neither world is the past changed.

The same reply can be made to the theological version: it's not true that if I could refrain from action A, I would confute God's knowledge. Rather, in the world in which I do B, God's knowledge in that world is different from his knowledge in this world (since the worlds contain different truths). In the A world, God knows that I will do A; in the B world, God knows that I will do B. This does not confute God's knowledge. In addition, the theological argument makes the standard fatalist assumption about the 'direction of dependence' diagnosed above. God's knowledge that I will do A depends on my freely doing A, not *vice versa* as the theological fatalist supposes.

FREE WILL AND DETERMINISM

Fatalism is not the only threat to free action. Since Thomas Hobbes (1588–1679), philosophers have worried that determinism threatens free will. Some – the compatibilists (such as Hobbes and David Hume) – regard the tension as illusory and hold that free will and determinism are compatible. (Indeed, some compatibilists hold that free will requires the truth of determinism.) Others – the incompatibilists – hold that the tension is real enough: free will and determinism are incompatible. However, incompatibilists then divide into two camps: the libertarians, who conclude that, since we are free, determinism must be false; and the hard determinists, who conclude that, since determinism is true, we lack free will.

Finally, there are those philosophers (such as Galen Strawson) who hold that the truth or otherwise of determinism is irrelevant to whether we have free will. The concept of free will – whether compatibilist free will or libertarian free will – is internally incoherent, and can be shown to be so on purely

Thomas Hobbes (1588–1679)

Hobbes is best known as a political philosopher. In his book *Leviathan*, published in 1651, Hobbes argued for rule by an absolute sovereign as a bulwark against civil disorder. Hobbes wanted to show how politics could be a science. His account of the nature of matter was the foundation for his theory of human psychology, which in turn provided the foundation for his theory of politics. For Hobbes, the entire universe, including human thought and behaviour, is just matter in motion. All causation is mechanistic and necessary: everything is necessitated by what has gone before. Nonetheless, Hobbes held that his mechanistic view of the universe is consistent with the existence of God (conceived as a material being), with conscious mental states (also conceived as physical), and with human freedom. Hobbes was thus one of the first compatibilists: ‘Liberty and Necessity are consistent’, he wrote, ‘the actions which men voluntarily do . . . proceed from liberty, and because every act of man’s will . . . proceeds from some cause, and that from another cause, in a continual chain, [such voluntary actions] proceed from necessity’ (*Leviathan*).

a priori grounds. Hence free will is an illusion. Since moral responsibility presupposes free will, moral responsibility is an illusion too.⁹

THE CASE FOR INCOMPATIBILISM

In order to appreciate the case for incompatibilism, we first need to characterize determinism. Determinism is the thesis that, given the **laws of nature**, and the state of the universe at any past time *t*, it is physically impossible for the history of the universe (before and after *t*) to be other than it is. Another way of formulating determinism is in terms of causation. It can be rendered as the thesis that every event has a cause, where causes are understood to necessitate or determine their effects.

Both the regularities we observe around us, and the rise of Newtonian science, reinforced belief in determinism. But the thesis of determinism is not a logical truth. If true, it is a contingent, empirical truth. Even so, determinism is a thesis that might be criticized on philosophical grounds. For example, if we accepted a regularity theory of causation (of the sort associated with Hume), we would have a reason to deny that causes necessitate their effects. But, as we saw in Chapter 4, the regularity theory is open to objections.

Of course, many think that quantum mechanical considerations reveal our universe not to be deterministic. But it’s unclear that indeterministic or partly random goings-on at the subatomic level are relevant to free will. If an agent’s

action is the result of an indeterministic process, how can we explain why he did that action rather than some alternative? How can an agent be morally responsible for an action that resulted from an indeterministic process (over which he can have no control)? Why is such an agent not merely lucky if he does the right thing? These questions raise significant problems for libertarian conceptions of free will.

However, even if determinism fails at the subatomic level, determinism at the macro-level may still pose a threat to free action, so the argument for incompatibilism still needs to be addressed. Why is free will thought to be incompatible with determinism? Suppose that determinism is true, and suppose I freely do action A (say, wave my arm in greeting). According to the incompatibilist, if A is free, I could have refrained from doing A. I could have done otherwise. But, according to determinism, given the laws of nature and the state of the universe at any time prior to my doing A, the future of the universe could not have been other than it actually is. So I could not have refrained from doing A. In which case, I did not do A freely; and similarly for every action of every other person in the universe.

Here are two frequently encountered replies to this incompatibilist argument. First, it is sometimes thought that the argument goes through only on the assumption that persons and their mental states are purely physical. However, this is not so. Most actions, such as arm-waving, involve movement of matter, and such movement is governed by deterministic laws. Thus even if persons and their mental states were non-physical, and even if the non-physical fell outside the scope of determinism, my arm could not have been in any other spatio-temporal location than its actual one when I waved. So I still could not have refrained from waving my arm, even if the mental causes of my arm's moving were non-physical.

Second, some philosophers have thought that human actions are not governed by causal laws, but belong to a different sphere. Human actions, unlike mere bodily movements, are explained by reasons, and reasons are not causes. We operate with two explanatory grids: the movement of matter is explained by causal law, but human action is explained by the ascription of reasons. In which case, the truth or otherwise of determinism is irrelevant to the existence and understanding of free human action.

However, it is controversial to claim that reasons are not causes. Was my wanting to greet someone (a reason) not the cause of my waving? But there is a more basic problem with this response – the same problem that beset the previous response. Even if human actions and bodily movements belong to different logical categories, they are still connected: most human actions, such as waving, involve bodily movements. So any deterministic constraints on bodily movement will equally be constraints on human action. Thus, neither the 'mental states are non-physical' nor the 'reasons are not causes' lines yield any convincing response to the argument for incompatibilism.

A COMPATIBILIST REPLY

Where then do compatibilists think the incompatibilist's argument goes wrong? The standard compatibilist response is to deny that there is any incompatibility between free will and determinism. Free will is to be contrasted, not with universal causation or determinism, but with *constraints* of various familiar kinds. In saying this, compatibilists take their view to be vindicated by facts about the ordinary usage of the words 'free' and 'unfree'. We say that a man is unfree only when someone holds a gun to his head, or claps him in irons, or suchlike. It is the presence of such specific constraints that robs a man of his freedom, not the fact of universal causation.

However, there are two reasons for dissatisfaction with this response. First, ordinary usage is not quite as the compatibilist takes it to be. If someone holds a gun to my head, and, in response, I open the safe, my action is still free. I could have chosen not to comply with the robber's wishes (unwise though that might have been). The temptation to describe this as a case of unfree action may stem from a confusion between free will and moral accountability. I certainly would not be held morally accountable for the theft of the money in the circumstances described, but that does not mean that my action would have been unfree. This is not a devastating objection to compatibilism, but it does show that the cases of constraint with which free action should be contrasted are those in which one is literally deprived of the power of choice or action (e.g., physical restraint, kleptomania, 'acting' under hypnosis, etc.).

Second, and more importantly, the compatibilist is attempting to draw a distinction within the realm of causes, between those causes compatible with free will (determining causes) and those causes incompatible with free will (constraining causes). But is this not an arbitrary distinction? An agent subject to either kind of cause could not have done otherwise; so why are determining causes consistent with free action while constraining ones are not?

A thought experiment devised by Harry Frankfurt may provide a better reply to the incompatibilist's argument and, in doing so, point towards a different (compatibilist) conception of free will.¹⁰ Here is a simple version of Frankfurt's thought experiment. Let us suppose that Smith plans to rob his local bank, waits until the appropriate time, and duly executes his plan. In terms of the actual course of events this is like any other bank robbery, except for one detail. Unbeknownst to Smith, an evil, robbery-loving demon has been monitoring Smith's brain and was prepared to intervene if Smith had shown any hesitation about committing the robbery. If Smith had attempted to change his mind, he would have found himself unable to do so. As it happens, Smith very much wanted the money and never changed his mind, so the demon never had to intervene in the actual course of Smith's life.

We have a robust intuition that Smith's action of robbing the bank was free, and one for which he was fully morally responsible. The planning and execu-

tion of the robbery were entirely of Smith's devising. The demon did not cause Smith to do anything. Yet Smith could not have done otherwise: in respect of the robbery, he could not have chosen or acted differently. Doesn't this show that free will does not require the ability to do otherwise? In which case, the compatibilist now has a reasoned reply to the incompatibilist's argument. Free will and determinism have not been shown to be incompatible, since free will does not require the ability to do otherwise. Rather, free will requires only that we act on our own beliefs and desires.¹¹

IS THE NOTION OF FREE WILL INCOHERENT?

However, Frankfurt's example does not mean that compatibilism is true. Even if determinism is no threat to free will, free will and determinism will be incompatible if free will is impossible.¹² Galen Strawson has recently argued that the notion of free will is incoherent.¹³ Free will and moral responsibility require a conception of self-determination which is logically unsatisfiable. J. G. Fichte (1762–1814) nicely captured this paradoxical notion of self-determination:

What I desired was this: that I myself, that of which I am conscious as my own being and person, . . . that this 'I' would be independent, would be something which exists not by another or through another, but of myself, and, as such, would be the final root of all my determinations.¹⁴

Galen Strawson (1952–present)

Galen Strawson, son of the late Oxford philosopher P. F. Strawson, is currently a professor of philosophy at the City University of New York Graduate School. He was educated at Cambridge, Oxford and the Sorbonne, and he previously taught at Oxford and Reading. A creative, rigorous and independent thinker, Strawson has produced three important books to date. In 1986 he published *Freedom and Belief* in which he argued that free will and ultimate moral responsibility are incoherent notions. Free will is an illusion. In 1989 he published *The Secret Connection* in which he argued that Hume did not hold a regularity theory of causation and believed in natural necessity. In 1995 he published *Mental Reality* and argued therein that conscious experience is the distinctive mark of the mental and that there is no conceptual connection between a subject's possession of mental states and his behaviour or dispositions to behaviour. He called his view a 'naturalized Cartesianism'. In all three works Strawson argued ingeniously against established orthodoxies.

WHAT IS THIS THING CALLED METAPHYSICS?

Why think self-determination impossible? Strawson's thought is that self-determination requires you to be ultimately morally responsible for what you do. To be ultimately morally responsible for what you do, you must be ultimately responsible for what you are in some mental respects (e.g., character). But it is impossible to be ultimately responsible for the way you are in mental respects (or in any other respect). For this would require that you intentionally brought it about that you had a certain mental nature. But this in turn would require that you had a prior mental nature which you intentionally brought about, which in turn would require a prior mental nature which . . . and so on, *ad infinitum*. Hence, on pain of an infinite regress, it is impossible to be ultimately responsible for the way you are in any mental respect, and so it is impossible to be ultimately responsible for what you do. In the absence of such responsibility there can be no self-determination and hence no freedom.¹⁵

This is a challenging argument. It will be a sound argument if any worthwhile notion of free will or moral responsibility must have this self-determining or self-creating character. The challenge, therefore, is whether a defender of free will can produce a notion of free will, not requiring full self-determination, which nonetheless grounds moral responsibility. This may not prove an easy challenge to meet.

CONCLUDING REMARKS

We have examined the two best-known attacks on free will: fatalism and determinism. The arguments of the fatalist were exposed as sophistry. They do nothing to show that it is inevitable or fated that we do what we do. The classical incompatibilist argument (attempting to show that free will and determinism are incompatible) was also found wanting. The biggest threat to free will may stem from Galen Strawson's contention that our notions of free will and ultimate moral responsibility are logically unsatisfiable.

STUDY QUESTIONS

- Outline the two fatalist arguments criticized in this chapter. Can either be improved upon?
- If we reject the fatalist's arguments should we also accept that it can sometimes be rational to attempt to bring about the past?

- Is the argument of the theological fatalist interestingly different from that of his non-theological counterpart?
- How would you defend libertarianism about free will?
- Might we justify our normal social and moral practices (thanking, blaming, punishing, etc.) without appeal to the (allegedly) incoherent notion of ultimate moral responsibility?

ANNOTATED FURTHER READING

- A. J. Ayer, 'Fatalism', in his *The Concept of a Person* (London: Macmillan) 1963. A clear and no-nonsense critique of both standard and theological fatalism.
- A. J. Ayer, 'Freedom and Necessity', in his *Philosophical Essays* (London: Macmillan) 1965. A modern defence of compatibilism about free will.
- R. Clarke, *Libertarian Accounts of Free Will* (Oxford: Oxford University Press) 2003. A thorough overview of libertarian attempts to make sense of free will.
- M. Dummett, 'Bringing about the Past', in his *Truth and Other Enigmas* (Cambridge, Mass.: Harvard University Press) 1978. A classic, but difficult, discussion of whether it can ever be rational to attempt to affect the past with, *en route*, a comparison with, and diagnosis of, the fatalist's argument.
- G. Strawson, *Freedom and Belief* (Oxford: Oxford University Press) 1986. A clear and readable defence of the radical view that free will and ultimate moral responsibility are illusions.
- P. van Inwagen, *An Essay on Free Will* (Oxford: Clarendon Press) 1983. A clear and thorough defence of incompatibilism, with a useful discussion of fatalism.
- G. Watson, *Free Will* (Oxford: Oxford University Press) 2003. A useful collection of the best work on free will in the past half-century.

INTERNET RESOURCES

- E. Craig (1998) 'Fatalism', in E. Craig (ed.), *Routledge Encyclopedia of Philosophy*. Retrieved 31 May 2006 from <<http://www.rep.routledge.com/article/N096>>.
- M. McKenna (2004) 'Compatibilism', *The Stanford Encyclopedia of Philosophy (Summer 2004 edition)*, Edward N. Zalta (ed.). Retrieved 31 May 2006 from <<http://plato.stanford.edu/archives/sum2004/entries/compatibilism>>.
- T. O'Connor (2005) 'Free Will', *The Stanford Encyclopedia of Philosophy (Summer 2005 edition)*, Edward N. Zalta (ed.). Retrieved 31 May 2006 from <<http://plato.stanford.edu/archives/sum2005/entries/freewill>>.
- H. Rice (2002) 'Fatalism', *The Stanford Encyclopedia of Philosophy (Winter 2002 edition)*, Edward N. Zalta (ed.). Retrieved 31 May 2006 from <<http://plato.stanford.edu/archives/win2002/entries/fatalism>>.
- G. Strawson (1998, 2004) 'Free will', in E. Craig (ed.), *Routledge Encyclopedia of Philosophy*. Retrieved 31 May 2006 from <<http://www.rep.routledge.com/article/V014>>.

WHAT IS THIS THING CALLED METAPHYSICS?

L. Zagzebski (2005) 'Foreknowledge and Free Will', *The Stanford Encyclopedia of Philosophy* (Fall 2005 edition), Edward N. Zalta (ed.). Retrieved 31 May 2006 from <<http://plato.stanford.edu/archives/fall2005/entries/free-will-foreknowledge>>.

PERSONAL IDENTITY

- Introduction 121
- Theories of personal identity 122
- What is a person? 123
- What is it for a person to persist? 125
- The complex view defended 128
- Does identity matter? 131
- The simple view 132
- Concluding remarks 134

INTRODUCTION

What is a person? Are we persons at all times at which we exist? What is it for the same person to persist over time? Is personal identity an important relation? These are the key questions that will be addressed in this chapter.

In asking questions about personal identity, we mean ‘identity’ in the sense of strict numerical identity, as governed by **Leibniz’s Law** (that is, A is counted as identical to B only if everything true of A is true of B and *vice versa*). We are not concerned with identity in the qualitative sense, as when we speak of ‘identical twins’. Numerically, the twins are two people, not one.

Note also that in asking what it is for the same person to persist over time, we are asking a constitutive question not an evidential one. We are asking what it is for person A at time t1 to be the same person as person B at time t2, and we are hoping for an informative answer to this question. We are not asking why we believe that A and B are the same, or what evidence we typically rely on in arriving at judgments of personal identity. Such evidence (e.g., physical appearance, voice, fingerprints, etc.) is never a logical guarantee of personal identity. (It is possible that someone might copy my fingerprints, but that would not make that person me.) We are seeking a condition which, if it obtains, logically guarantees identity, and, if it fails to obtain, guarantees non-identity. Though separate, the constitutive and evidential questions are not unrelated: ideally, an answer to the constitutive question should harmonize with our standard uncontroversial ways of telling who is who in everyday cases.

Criteria of identity

A criterion of identity for Fs tells us what the identity over time of Fs consists in, and hence tells us what changes an F can survive, and what changes destroy an F. It is assumed that the criterion of identity for Fs will not presuppose the notion of F-identity. Criteria of identity are thus standardly conceived as reductive in character. Some think the criterion of identity for artefacts such as ships is spatio-temporal continuity: a ship at t1 is identical to a ship at t2 if and only if the earlier ship is spatio-temporally continuous (traces a continuous path through space and time) with the later ship. The criterion is reductive – spatio-temporal continuity does not presuppose ship-identity – and tells us what changes a ship can survive (those changes, yet to be specified, that do not disrupt its spatio-temporal continuity). Although the criterion may be open to objection – can a ship not survive discontinuity (e.g., dismantlement and reassembly)? – the intent behind it is clear enough. Is there a criterion of identity for persons? Many say ‘yes’, but defenders of the simple view think not: for them, personal identity is primitive or basic. Persons are special precisely because they lack criteria of identity.

THEORIES OF PERSONAL IDENTITY

There are many theories of personal identity over time. They can usefully be divided into two categories: the simple view and (various versions of) the complex view. According to the simple view, endorsed by Joseph Butler, Thomas Reid and, more recently, Richard Swinburne, no substantial, informative answer to the constitutive question is possible.¹ Personal identity is primitive and unanalysable. We cannot say, in other terms, what is it for the same person to persist over time. The simple view may seem unappealing, but it can be motivated.

According to the complex view, endorsed by the majority of contemporary philosophers, personal identity is further analysable, typically in terms of continuities of one sort or another. Some versions of the complex view analyse personal identity in terms of psychological continuity, other versions appeal to physical continuities, such as the continued existence of the body or brain.

A further distinction can be drawn between those versions of the complex view which offer a **reductive** analysis of personal identity, and those which offer only a non-reductive analysis. An analysis of personal identity is reductive if the fact of a person’s existence and continued identity can be understood without reference to the concept of a person. Derek Parfit is the best-known defender of reductionism about persons. He thinks that reality can be completely described in impersonal terms. As well as having a radical view about the nature of personal identity, Parfit also has a radical view about its importance.

Joseph Butler (1692–1752)

Joseph Butler was born in Wantage, England and educated at Oriel College, Oxford where he studied the ideas of Locke and Shaftesbury. He corresponded with the English philosopher and theologian, Samuel Clarke, and pursued a career within the Anglican Church, eventually being appointed Bishop of Durham. His major works are *Fifteen Sermons* (1726) and *Analogy of Religion* (1736). In the earlier work, Butler criticized Shaftesbury for leaving conscience out of his account of morality and offered a refutation of psychological hedonism or egoism. In the later work, Butler argued that it is reasonable to accept arguments in favour of natural religion or deism (i.e., systems of thought that attempt to establish the existence of God on the basis of rational insight, independently of any revelation).

Derek Parfit (1942–present)

Educated at Eton and Balliol College, Oxford, Parfit was elected a Prize Fellow at All Souls College, Oxford in 1967. He has remained there ever since, but regularly visits Harvard and New York universities. Parfit has published a number of landmark papers, beginning with 'Personal Identity' (1971), but his most significant work to date is his book *Reasons and Persons* (1984). Much of this rich and ingenious work is devoted to undermining the self-interest theory of rationality. Parfit offers many arguments against this theory, concluding that it can be no less rational, e.g., to want to benefit others at one's own expense or to want to make some intellectual advance. Parfit's views on personal identity also undermine the self-interest theory. If, as he claims, identity is not what matters, then the self-interest theory loses all force. In the final part of his book, Parfit discusses puzzles and paradoxes arising from our ability to affect both the identity and quality of life of future generations. We need a new theory of beneficence in order to give a satisfactory account of these matters, but Parfit admits that he has not yet found such a theory.

Parfit holds that personal identity is unimportant – identity is not what matters – and he takes this to have significant implications for rationality and morality.²

WHAT IS A PERSON?

A person is a mental being. But not all mental beings are persons: my cat is not. A person is a certain kind of mental being – a self-conscious mental being.

WHAT IS THIS THING CALLED METAPHYSICS?

It is hard to improve on Locke's definition, which also serves to elucidate the notion of self-consciousness. According to Locke, a person is 'a thinking, intelligent being, that has reason and reflection, and can consider itself as itself, the same thinking thing, in different times and places.'³ A person is thus a being of some psychological sophistication, capable of engaging in tensed and counterfactual first-person thoughts (thoughts such as 'I will take luncheon at Tetsuya's today', 'I was at the cinema last week', 'I might have dined at the Oriental last night, but chose not to', etc.).

This definition has the required result that typical adult human beings count as persons, and also makes clear why we value persons more than non-persons, since the life of a self-conscious being has more value than that of a non-self-conscious being (which is not to say that non-self-conscious beings have lives of no value). Nonetheless, it's fair to say that the definition, though true as far as it goes, does not tell us what ontological category, or category of being, persons belong to. Are persons immaterial souls, or human beings, or bodies, or brains, or bundles of perceptions?

Descartes thought that we each have an immaterial soul that survives bodily death.⁴ Hume thought that persons were not substances, but bundles of perceptions, and he famously compared the self to a republic. He wrote: 'I cannot compare the soul more properly to any thing than to a republic or commonwealth, in which the several members are united by the reciprocal ties of government and subordination, and give rise to other persons, who propagate the same republic in the incessant changes of its parts.'⁵

Parfit has made much of this analogy. Just as the constituents of a republic (its citizens and territory) can be understood without reference to the concept of a republic, so the constituents of the self or person (thoughts and experiences) can be understood with reference to the concept of a person. Hume's analogy thus provides the model for Parfit's reductionism about persons.

On yet other views, we are to be identified with our bodies, or with our brains, or with human beings (if human beings are different from human bodies). There is thus a wide range of views as to which entities are the bearers of self-consciousness.

Although we are persons, it should not be assumed at the outset that we are essentially persons, or that 'person' is, in David Wiggins's terminology, a substance sortal rather than a phase sortal.⁶ Indeed, according to one recent theory, animalism, even if non-animal (e.g., bionic or immaterial) persons are possible, we are essentially animals (human beings), but not essentially persons.⁷ On this theory, I am identical to the human being in my shoes. Since that animal can exist in a persistent vegetative state, irretrievably devoid of mentality, I can exist in such a state, though no person can exist in that state. In which case, I can exist but not as a person, and so 'person' is not a substance sortal. On this theory, and contrary to what has traditionally been assumed, the

question of *our* identity is not the same as the question of *personal* identity, and so the wrong questions have often been asked.

How are we to decide between these views of the nature of persons? A number of strategies are available. Discussions of Descartes's dualism or Hume's theory of the self are heavily theoretical in character. Deciding between the other views relies more on the methodology of thought experiments. Here the rationale is straightforward. If a theory holds that persons are essentially Fs, and we can think up a possible scenario in which a person continues to exist yet is not an F, then clearly persons cannot be identical to, or essentially, Fs. Hence, we can best answer the question 'What is a person?' by first answering the question 'What is it for the same person to exist over time?'

WHAT IS IT FOR A PERSON TO PERSIST?

Let us state more precisely various versions of the complex view and see what problems they face. We will then discuss Parfit's view that personal identity is not what matters and his reductionism about persons. Finally, we will turn our attention to the simple view.

Here, to begin with, are three versions of the complex view, all of which conform to most of our actual judgements of personal identity over time:

- (1) Body Criterion: A at t1 is the same person as B at t2 if and only if A's body is the same as B's body.
- (2) Brain Criterion: A at t1 is the same person as B at t2 if and only if A's brain is the same as B's brain.
- (3) Psychological Criterion: A at t1 is the same person as B at t2 if and only if A and B are psychologically continuous.

The body and brain criteria are fairly straightforward. It is assumed that the identity over time of bodies and brains is unproblematic, consisting in spatio-temporal continuity.⁸ It is also assumed, in order to generate a clean contrast between the two criteria, that 'body' means 'body minus brain'. Although names of persons ('A' and 'B') appear in the right-hand side of these two criteria, it can reasonably be assumed that such occurrences are eliminable. A's body, for example, could be specified simply by using spatio-temporal coordinates, without reference to A. Thus, the body and brain criteria offer reductive analyses of personal identity. A problematic relation – personal identity – is reduced to an unproblematic one.

The psychological criterion requires more explanation. We can say that A and B are psychologically continuous just if the mental states of the later B (i.e., B's memories, beliefs, character, projects, intentions, desires and so on) are causal descendants of A's mental states. Thus, B may remember an

WHAT IS THIS THING CALLED METAPHYSICS?

experience had by A, B may act so as to fulfil an intention had by A, B's sense of humour may have been inherited from A, and so on. When there are enough psychological links, or overlapping chains of such links, we may speak of psychological continuity. Psychological continuity is a transitive relation, though psychological connectedness (i.e., the direct psychological links that make up continuity) is non-transitive.⁹ This implies that I may be psychologically continuous with someone with whom I am psychologically very dissimilar.

Whereas the body and brain criteria are obviously reductive analyses of personal identity, it is controversial whether the psychological criterion is reductive. This depends on whether the notion of psychological continuity can be fully understood without reference to persons. Can pains and thoughts be completely described without reference to the subject who has them? Are the contents of memories and intentions not essentially person-involving (I remember that *I* tasted ice-cream, I intend that *I* holiday next month, etc.)? These are among the deepest questions in the topic of personal identity.

However, we can first ask whether any of these criteria are plausible. The criteria, recall, are intended to tell us the essence of personal identity. They are answers to the constitutive question and purport to be necessary truths, true of any person in any possible circumstance. Thus if there is a possible world, however unlike the actual world, in which one side of a criterion is true and the other side false, then that criterion is false.

Here is a thought experiment, depicting a logically possible scenario which many take to refute the body criterion. Sydney Shoemaker was the first to introduce this thought experiment (which we can call 'Brain Transplant') into the literature. He wrote:

It is now possible to transplant certain organs . . . [i]t is at least conceivable . . . that a human body could continue to function normally if its brain were replaced by one taken from another human body. . . . Two men, a Mr Brown and a Mr Robinson, had been operated on for brain tumours, and brain extractions had been performed on both of them. At the end of the operations, however, the assistant inadvertently put Brown's brain in Robinson's head, and Robinson's brain in Brown's head. One of these men immediately dies, but the other, the one with Robinson's head and Brown's brain, eventually regains consciousness. Let us call the latter 'Brownson'. . . . When asked his name he automatically replies 'Brown'. He recognizes Brown's wife and family. . . . And is able to describe in detail events in Brown's life . . . of Robinson's life he evidences no knowledge at all.¹⁰

Almost everyone agrees that the best description of this case is that Brown is Brownson. Few think: Brown dies and Robinson acquires a new brain and

a new psychology. So we may take the case of ‘Brain Transplant’ to refute the body criterion, and to refute animalism (the view that we are identical to human beings), since Brown is not the same human being as Brownson, yet he is the same person.

What further moral should we draw from this refutation of the body criterion? Well, it points towards the other two theories on offer. Perhaps Brown is Brownson because Brownson has Brown’s brain. Or perhaps Brown is Brownson because Brownson is psychologically continuous with Brown. But there is another thought experiment which raises problems for both the brain criterion and the psychological criterion: ‘Fission’.

Imagine that my brain is divided and each (equipollent) hemisphere is transplanted into two (brainless) bodies.¹¹ After the operation, two people wake up – call them by the proper names ‘Lefty’ and ‘Righty’ – each of whom is psychologically continuous with me (they both have my beliefs and character, and both seem to remember my past life). They are psychologically exactly similar to each other upon waking, but begin to differ thereafter.

I take ‘Fission’ to represent a logically possible scenario; it is a variation on ‘Brain Transplant’. How does it raise a problem for the brain criterion and the psychological criterion? The psychological criterion implies that I am Lefty and that I am Righty (since I am psychologically continuous with both). It then follows, by the transitivity of identity, that Lefty is Righty.¹² But this cannot be right. Lefty and Righty, though initially very similar, are two people not one.

Fission

Though not technically possible at present, the following scenario is at least logically possible. Surgeons divide my brain and transplant each hemisphere into two brainless bodies. Two people wake up – Lefty and Righty – who are psychologically just like me. Both have my character, sense of humour, right-wing leanings, etc., and both have apparent memories of my past. Some philosophers have thought that fission cases tell us something about the nature of personal identity and something about its importance. First, since I am neither Lefty nor Righty, but would continue to exist had there been only one survivor, it follows that personal identity consists in some continuity (physical or psychological) holding in a one–one or non-branching form. Second, my relation to Lefty and Righty contains everything that matters in normal cases of identity. Since I am not identical to Lefty or Righty, identity cannot be what matters. What matters is having a psychological continuer, irrespective of whether that continuer is oneself. Parfit accepts both these implications. Others accept the first implication but not the second. Defenders of the simple view reject both implications.

WHAT IS THIS THING CALLED METAPHYSICS?

It might be thought that there is no analogous problem for the brain criterion. That criterion doesn't imply that I am Lefty or that I am Righty, since neither Lefty nor Righty has my whole brain. But the brain criterion, if it is to be plausible, cannot require identity of the whole brain as a condition of personal identity. Some people have actually survived with half their brain destroyed. If such a person's remaining hemisphere was transplanted to a new body, we would consider this a continuation of that person's life. So, to have a chance of being true, the brain criterion must be understood along the following lines:

Revised Brain Criterion: A at t1 is the same person as B at t2 if and only if B has enough of A's brain to sustain psychological continuity.

'Fission' is a problem for the revised brain criterion. That criterion, like the psychological criterion, implies that I am Lefty and I am Righty (since Lefty and Righty have enough of my brain to support full psychological continuity).

Should we then conclude that the revised brain criterion and the psychological criterion are false? That would be too quick. A lot of ink has been spilled defending these criteria against the current objection. There are basically two ways in which a defence of these criteria can run. One way – the multiple-occupancy response – involves keeping the criteria as they are, but claiming that the case of fission has been misdescribed. The other way – the uniqueness response – involves modifying the criteria by inserting a non-branching clause in the analyses. Neither response is congenial to common sense.

THE COMPLEX VIEW DEFENDED

We can begin with the multiple-occupancy response, defended by Lewis, Noonan, and others.¹³ According to this response, my fission has been misdescribed. It does not involve three people (myself, Righty and Lefty), but only two (Lefty and Righty). Lefty and Righty both exist prior to fission, occupying the same body ('my' body). After fission they each occupy their own body.

However, it's not clear how this redescription of fission helps preserve the psychological criterion (which is the intent of Lewis and Noonan). If Lefty and Righty both exist prior to Fission, is it not true that Lefty (before fission) is psychologically continuous with Righty (after fission), and that Righty (before fission) is psychologically continuous with Lefty (after fission)? In which case, the psychological criterion implies that Lefty is Righty. But 'fission' was redescribed in order to avoid this consequence. How have matters been advanced?

Second, it hardly needs to be remarked that the multiple-occupancy account is metaphysically perverse. We are being asked to believe that two people occupy

a single body where, by all usual criteria, there is only one person present. Third, 'my' pre-fission utterances of 'I' will be ambiguous between Lefty and Righty. How then are we to account for the apparent unity of the 'I' thoughts associated with the pre-fission body? (My pre-fission 'I' thoughts do not feel ambiguous.) How could it be known which of Lefty and Righty was being referred to by some use of 'I'?

How do matters stand with the uniqueness response? This response has been endorsed by Parfit (in defence of the psychological criterion). He writes:

Williams has attacked [the psychological criterion] with the following argument. Identity is a one-one relation. So any criterion of identity must appeal to a relation which is logically one-one. Psychological continuity is not logically one-one. So it cannot provide a criterion. [This is a restatement of the present objection.] . . . Some writers have replied that it is enough if the relation appealed to is always in fact one-one. . . . I suggest a slightly different reply. Psychological continuity is a ground for speaking of identity when it is one-one.¹⁴

In other words, the suggestion is that we should analyse personal identity over time in terms of **non-branching** psychological continuity, a relation which is, by stipulation, logically one-one. Thus we should endorse:

Revised Psychological Criterion: A at t1 is the same person as B at t2 if and only if A and B are psychologically continuous and A's stream of mental life has not branched between t1 and t2.

According to this revised criterion, since branching has occurred in the case of fission, I am identical to neither Lefty nor Righty. Thus we avoid the unintuitive consequence that Lefty is Righty.

However, the revised psychological criterion has a counter-intuitive consequence. (I take it that a modified version of the revised brain criterion – *viz.*, A at t1 is the same person as B at t2 if and only if B has enough of A's brain to sustain psychological continuity and no one else at t2 has enough of A's brain to sustain psychological continuity – has this consequence too.¹⁵)

According to the revised psychological criterion, I am neither Lefty nor Righty, but had a nurse dropped and destroyed the left hemisphere during the operation, thus ensuring only one survivor, I would then have survived. If I am neither Lefty nor Righty in the fission world, but I am the survivor in the one-survivor world, then it seems that, prior to the fission operation, I have a strong reason to bribe the nurse to drop one of the hemispheres, thus ensuring my survival. But if, e.g., I have good reason to think that the right-hemisphere operation will be successful, should I really worry about whether another operation, in a different building, is successful or not? How can that possibly matter

WHAT IS THIS THING CALLED METAPHYSICS?

to me? Yet matter it does: it is the difference between life and death. Thus, on the revised psychological criterion, something that ought not to matter does matter.

However, Parfit has an ingenious reply to this objection. To start with, even in his earliest work (1971), Parfit was not happy describing fission as a case in which I am identical to neither Lefty nor Righty. Initially he thought that there was no good description of fission in terms of identity, and that, if fission occurred, 'we should need . . . to abandon the language of identity.'¹⁶ He argued against the description that I am neither Lefty nor Righty as follows:

We agreed that I could survive if my brain were successfully transplanted. And people have in fact survived with half their brains destroyed. It seems to follow that I could survive if half my brain were successfully transplanted and the other half were destroyed. But if this is so, how could I *not* survive if the other half were successfully transplanted? How could a double success be a failure?¹⁷

By the time he published his book *Reasons and Persons* (1984) Parfit was no longer persuaded by this argument, and conceded that 'I am neither Lefty nor Righty' is the best description of fission. Rightly so, since the 1971 argument is not compelling. It plays on an ambiguity in the word 'success'. In a purely medical sense, a procedure is successful if it goes as it should. In that sense, a double success is obviously not a failure. But if 'success' requires the survival of the original person, then a 'double success' may indeed be a failure. A success in the medical sense does not mean a success in the sense of personal survival.

However, there is another idea, present in both the 1971 article and the 1984 book, which is this: it is absurd for me to regard the prospect of fission as I would ordinary death. Unless this idea is pressed into service, just to say that I am neither Lefty nor Righty would be misleading. Parfit thinks that fission is as good as ordinary survival. My relation to Lefty, and my relation to Righty, each contain all that matters in ordinary survival. Since I am not identical to either Lefty or Righty (as we are now conceding), it follows that identity is not what matters. What does matter is psychological continuity and/or connectedness.

The thesis that identity is not what matters allows Parfit a reply to our objection to the revised psychological criterion. If identity does not matter, I have no reason to bribe the nurse to drop one of my hemispheres. I have no reason to bring about a one-survivor world in preference to a two-survivor world – even though I survive (continue to exist) only in the former world. As long as I have at least one survivor with whom I am fully psychologically continuous, it does not matter whether I have more than one.

DOES IDENTITY MATTER?

Clearly, then, Parfit needs the thesis that identity is not what matters in order to rebut the objection to the revised psychological criterion. But what exactly does the thesis amount to? There are really two theses to consider: the negative thesis that identity is not what matters and the positive thesis that the relation of psychological continuity and/or connectedness (what Parfit calls relation R) does matter.¹⁸

Suppose that I am about to divide tomorrow, and I know that Righty will suffer toothache. Parfit thinks that, in virtue of Righty's R-relatedness to me, I ought to have the very same concern for Righty's toothache as I would have if I were going to suffer toothache tomorrow. In this way, then, identity does not matter.

But what is the concern that I have for both Righty and myself? I have self-concern for my own future, and I plainly cannot have self-concern for Righty's future, since he is not me. So it seems that I cannot have the same kind of concern for both myself and Righty. Parfit will reply that this is a merely verbal point. We cannot call my concern for Righty 'self-concern', but it is essentially the same kind of concern that I have for myself.

This can be so only if *self-concern* is a composite concept built up out of the (supposedly) more basic components *identity* and *concern*, where the latter concept does not presuppose identity and is the kind of concern I can have for anyone to whom I am R-related. In this way, we can regard the self-concern I have for myself and the concern I have for Righty as essentially the same, differing only in their verbal description.

The crucial and currently undecided question is whether *self-concern* is a unitary concept, not reducible to more basic concepts, or whether it is a composite concept, built up precisely out of such concepts. It is only if *self-concern* turns out to be a composite concept that Parfit's defence of the revised psychological criterion can work.

The tenability of Parfit's reductionism turns on the very same kind of issue. According to reductionism, a person's psychological life can be completely described in impersonal terms (i.e., without reference to persons or personal identity). This may seem an implausible thesis even for simple mental states, such as pains and toothaches, which seem of their nature to require an owner or bearer. But there is a special problem with psychological states such as memory and intention which appear to have identity built into their content. Thus my memory of a past experience (say, tasting ice cream yesterday) not only requires a current bearer, but it seems to implicate me in its content: I remember that *I* tasted ice cream yesterday. (If it turned out that I have never tasted ice cream, you could perfectly well retort 'you don't remember tasting ice cream, you just think you do'.) How then can memory, a crucial feature of our psychological lives, be described impersonally?

WHAT IS THIS THING CALLED METAPHYSICS?

Parfit's reply is that memory is a composite concept built up out of *identity* and *quasi-memory* (*q-memory*). The latter concept is stipulated to be like memory in all phenomenological and causal respects, yet does not presuppose identity. Thus I can have q-memories of someone else's experiences, and what we call memories are just q-memories of one's own experiences. As with *self-concern*, the question is whether *memory* is indeed a composite concept, or whether it is a unitary concept not reducible to more basic conceptual atoms. (One worry, for example, is whether *q-memory* really can be understood independently of *memory*. If not, it will after all be identity-involving.) This debate, like that about *self-concern*, is still open.

THE SIMPLE VIEW

What conclusion should we draw from our previous discussion? It would be rash to conclude that the complex view is untenable. Nonetheless, it has been put under considerable pressure, and is very much on the defensive. Are there any other considerations that might tell against it, and thus push us towards the simple view?

According to the simple view, the relation of personal identity is primitive and unanalysable. Butler held that the word 'same' is used in a 'strict and philosophical sense' when applied to persons, but in a 'loose and popular' sense when applied to bodies and other kinds of thing (e.g., artefacts). Similarly, Reid wrote that identity:

has no fixed nature when applied to bodies and very often questions about it are questions about words. But identity when applied to persons has no ambiguity and admits not of degrees or of more or less. It is the foundation of all rights and obligations and of all accountability, and the notion of it is fixed and precise.¹⁹

Thus, whereas it may be a verbal question whether to call this car the 'same car' as some earlier car, a matter which may be settled by stipulation, it is never a verbal question whether a person at one time is the same as some person at another time, and such a question can never be settled by stipulation.

In his 1971 article, Parfit began by claiming that we have a belief about the nature of personal identity and a belief about its importance.²⁰ Both beliefs, he thinks, are false. The belief about the nature of personal identity is that it is always an all-or-nothing matter. Either I will exist in some future situation or I won't. There can be no grey area. Parfit points out that we don't think this in the case of nations or machines. We don't think there has to be a 'yes or no' answer to 'Was England the same nation after 1066?' Nor does there always have to be such an answer in the case of persons.²¹

All versions of the complex view imply that there can be cases of indeterminacy in personal identity (i.e., cases where there is not a simple ‘yes or no’ answer).²² For example, consider the revised psychological criterion. If I undergo some very minor psychological change tomorrow, I am psychologically continuous with the person occupying my body tomorrow. This is a clear case of identity, according to the revised psychological criterion. If tomorrow I undergo massive psychological changes, involving total character change, implantation of new ‘memories’, etc., then I am not psychologically continuous with the person occupying my body tomorrow, and so I am definitely not identical to that person. But there are cases in between, where it is indeterminate whether I am psychologically continuous with the person tomorrow. In such cases, according to the revised psychological criterion, it is indeterminate whether I am identical with the later person. The question ‘Am I that person?’ has no ‘yes or no’ answer.

Parfit would, of course, welcome this consequence of the revised psychological criterion since it serves to undermine the common belief about the nature of personal identity, which is one of his targets. But although it is easy enough to make sense of indeterminacy in identity in the case of ships and cars, it is hard to make much sense of it in the case of persons, especially when we consider the matter from a first-person point of view. Bernard Williams was the first to press this point in his illuminating article ‘The Self and the Future’.²³

Williams points out that we have no model for anticipating, or emotionally responding to, indeterminacy in our own case. He writes:

To be told that a future situation is a borderline one for its being myself that is hurt, that it is conceptually undecidable whether it will be me or not, is something which, it seems, I can do nothing with; because, in particular, it seems to have no comprehensible representation in my expectations and the emotions that go with them.²⁴

An indeterminate or conceptually undecidable case is not like one where I am told that, e.g., one of us in the room will be shot tomorrow. In that case, I know what the two possibilities are: either I will be shot or I won’t. Nor is it like the case where I think that some ‘nameless horror’ will befall me: whatever the horror turns out to be, it will either befall me or not. But if I am told that someone tomorrow will suffer great pain, and then told that it is indeterminate whether I am that person, I have no idea how to react (fear? pity? ‘ambiguous fear?’). No cognitive or emotional response fits this case. There is ‘an obstinate bafflement to mirroring in my expectations a situation in which it is conceptually undecidable whether I occur.’²⁵

This problem for the complex view should not be underestimated. It may not refute that view, but it makes some of its deliverances hard to understand. In contrast, the simple view has no such problem. On that view, personal

WHAT IS THIS THING CALLED METAPHYSICS?

identity is always all or nothing and never comes in degrees. Unlike the identity of other things, there can never be situations in which a conceptual shadow is cast over our identity.²⁶

In addition, the simple view is not committed to either multiple occupancy or to the inclusion of a non-branching or no-competitors clause (and, hence, does not need to flirt with the view that ‘identity is not what matters’). According to the simple view, ‘Fission’, as described above, leaves open the question of identity. I could be Lefty, I could be Righty, or I could be neither. Any description in terms of physical and psychological continuities fails definitively to settle the question of who is who. Personal identity is always a ‘further fact’ over and above such continuities. Anyone who agreed to undergo fission would be taking a risk, but, as Williams observes of another puzzle case, ‘that there is room for the notion of a risk here is itself a major feature of the problem.’²⁷

It is sometimes complained that the simple view makes personal identity over time unknowable. This is not so. Certainly, on the simple view, our ordinary evidence of personal identity over time (the obtaining of various continuities) is no logical guarantee of identity. But we take ourselves to have knowledge of the external world and other minds even though our evidence does not logically guarantee the existence of an external world or of other minds. Why then should knowledge of our own and others’ identity be compromised by the logical defeasibility of our evidence?

If the simple view is true, what is a person? A person cannot be identical to his brain, or body, or any other biological entity (for then some version of the complex view would be true). On the simple view, a person’s existence may depend causally upon the existence and normal functioning of his brain, but there need be no metaphysical dependency. Such a view is certainly compatible with the thesis that persons are immaterial souls, but it is not obvious that it requires this thesis. The simple view also fits with the thesis that we are essentially or fundamentally persons.

CONCLUDING REMARKS

We have covered a lot of ground in this chapter. We distinguished numerical from qualitative senses of ‘identity’, and constitutive from evidential questions. We looked at the question of what it is to be a person, and then looked at various substantial accounts of what it is for a person to persist, all of which were found wanting. The revised psychological criterion may be the most plausible version of the complex view, reflecting best the concept of persons as mental beings, but its defence requires the

truth of the controversial thesis that identity is not what matters. Moreover, the possibility of indeterminacy in personal identity, which is a feature of all versions of the complex view, is deeply problematic. It is a virtue of the simple view that it avoids these difficulties, and it is a view of persons which deserves more attention than it has received in contemporary discussions.

STUDY QUESTIONS

- Is a person fundamentally a mental being?
- Does the example of 'Brain Transplant' refute the body criterion?
- Can we live with the consequences of the revised psychological criterion?
- What is the thesis that identity is not what matters?
- What are the costs of accepting the simple view?

ANNOTATED FURTHER READING

- H. Noonan, *Personal Identity*, second edition (London: Routledge) 2003. An excellent, if demanding, introduction to the history of the topic and to the contemporary debate.
- D. Parfit, *Reasons and Persons* (Oxford: Oxford University Press) 1984, Part III. This book has largely set the agenda for recent discussions of personal identity. Although not introductory, Parfit writes clearly and any interested reader should be able to follow Parfit's radical and revisionary train of thought.
- S. Shoemaker and R. Swinburne, *Personal Identity* (Oxford: Basil Blackwell) 1984. A readable debate between a leading defender of the complex view and a leading defender of the simple view.

INTERNET RESOURCES

- B. Garrett (1998, 2004) 'Personal identity', in E. Craig (ed.), *Routledge Encyclopedia of Philosophy*. Retrieved 31 May 2006 from <<http://www.rep.routledge.com/article/V024>>.
- E. T. Olson (2002) 'Personal Identity', *The Stanford Encyclopedia of Philosophy (Fall 2002 edition)*, Edward N. Zalta (ed.). Retrieved 31 May 2006 from <<http://plato.stanford.edu/archives/fall2002/entries/identity-personal>>.



REALISM AND ANTI-REALISM

• Introduction	138
• Delineating the realist/anti-realist dispute	140
• Case study I: character	142
• Case study II: the past	144
• Anti-realism: motivation and assessment	146
• Anti-realism disproved?	148
• Concluding remarks	150

INTRODUCTION

Realism about some subject matter (Fs, say) is normally understood as the view that Fs exist, and do so independently of us. ‘Independence’ is not a transparent notion, but we can say that Fs exist independently of us if any of the following three conditions hold:

- (i) If we hadn’t existed, Fs would still have existed.
- (ii) If we were to cease to exist, Fs would still exist.
- (iii) The nature of Fs is not determined simply by whatever we take their nature to be.¹

Intuitively, we are realists about the planets ([i]–[iii] are all true of the planets), but non-realists about the fashionable (none of [i]–[iii] is true of the fashionable: something is fashionable only because people treat it a certain way).

Opposition to realism about Fs can thus take either of two forms: the insistence that Fs do not exist or the insistence that, although Fs exist, they do not exist independently of us. Error theories (advanced, e.g., about moral discourse and arithmetic) and expressivism (advanced, e.g., about moral discourse) are examples of the first kind of non-realist

strategy.² Idealism is an example of the second kind of strategy. Bishop Berkeley thought that the planets existed, but since they are composed of ideas, they do not exist independently of minds (human or divine). In this chapter I want to look at a more recent example of the second kind of non-realist strategy, first introduced in the 1960s by Michael Dummett. Dummett calls his brand of non-realism ‘anti-realism’, and he can be seen as attempting to forge a new sense in which the existence of Fs might not be independent of us.

Dummett suggests that traditional disputes in metaphysics might be advanced, or even settled, if pursued within the framework of the debate between realism and anti-realism. Although this is a debate about meaning and truth, Dummett felt that it could cast light on traditional metaphysical disputes about the nature of mathematics, other minds, material objects and the reality of the past and future.

Drawing on two early papers, ‘Realism’ (1963) and ‘The Reality of the Past’ (1969), I will question whether Dummett’s realist/anti-realist framework really is a fruitful one for the furtherance of metaphysics.³ I will also look at a well-known argument which seems to show that anti-realism leads to contradiction.

Realism

For many, realism is the default view of the world. Mountains, continents and planets, for example, exist independently of us. Had we not existed, they would still have existed; if we were to cease to exist, they would still exist; and their nature is not fixed by whatever we take their nature to be. Mountains, continents and planets exist and do so in complete independence of us. Idealism involves a sharp rejection of such realism. According to the idealist, mountains, continents and planets are mind-dependent entities. For such entities, to be is to be perceived (*esse est percipi*) – either by a human mind or a divine one. According to the idealist, were there no minds, there would be no mountains, continents or planets. Dummett's anti-realism is intended to occupy a position midway between realism and idealism. The world is not independent of us, since statements about the external world cannot be unrecognizably true. But the existence of the planets, for example, does not depend on their being perceived by some mind. The planets would still have existed even if there had been no minds.

DELINEATING THE REALIST/ANTI-REALIST DISPUTE

The first task, of course, is to clarify what Dummett means by the terms 'realist' and 'anti-realist'. Dummett says that realism about Fs, in his sense, is not the doctrine that there are Fs (where Fs are entities of some disputed sort), nor anti-realism its denial. Thus the realist/nominalist dispute over the existence of universals is not an example of a realist/anti-realist dispute in Dummett's sense (see Chapter 3). Dummett's realist/anti-realist dispute is not concerned with the existence of some purported class of entities, but with features of certain classes of statements (statements about other minds, mathematics, the physical world, the past and future, etc.) Dummett is fond of quoting Kreisel's dictum that **Platonism** in mathematics is best understood, not as a doctrine about the existence of mathematical objects, but a thesis about the objectivity of mathematical statements. Realism, in Dummett's sense, is one expression of such objectivity.

Where the 'disputed class' refers to the class of statements in some area over which there is a realist/anti-realist dispute, Dummett characterizes realism as: 'the belief that statements of the disputed class possess an objective truth-value, independently of our means of knowing it: they are true or false in virtue of a reality existing independently of us.'⁴ In contrast:

The anti-realist opposes to this the view that statements of the disputed class are to be understood only by reference to the sort of things which

we count as evidence for a statement of that class. . . . [t]he meanings of those statements are directly tied to what we count as evidence for them, in such a way that a statement of the disputed class, if true at all, can be true only in virtue of something of which we could know and which should count as evidence for its truth.⁵

Hence:

The realist and anti-realist may agree that it is an objective matter whether, in the case of any given statement of the class, the criteria we use for judging such a statement to be true are satisfied: the difference between them lies in the fact that, for the anti-realist, the truth of the statement can only consist in the satisfaction of those criteria, whereas, for the realist, the statement can be true even though we have no means of recognizing it as true.⁶

Later he writes:

the dispute can arise only for classes of statements for which it is admitted on both sides that there may not exist evidence either for or against a given statement. It is this, therefore, which makes acceptance of the law of excluded middle for statements of a given class a crucial test for whether or not someone takes a realist view of statements of that class. The anti-realist cannot allow that the law of excluded middle is generally valid.⁷

Consequently, ‘The dispute . . . concerns the notion of truth appropriate for statements of the disputed class; and this means that it is a dispute concerning the kind of *meaning* which these statements have.’⁸

To each disputed class of statements there often corresponds a reductive class of statements. Thus, to the disputed class of statements about material objects there corresponds the reductive class of statements about sense experience. To statements about other minds there correspond statements about behaviour. In speaking of a ‘reductive class’, Dummett means to imply that statements of that class can be understood independently of statements of its corresponding disputed class. In such cases, the anti-realist view is that the truth of a statement in the disputed class consists in the truth of some statement, or set of statements, drawn from the reductive class.⁹

However, anti-realism should not be identified with reductionism: reductionism is neither necessary nor sufficient for anti-realism. It is not necessary since there is no reductive class for the disputed class of mathematical statements (*viz.*, those mathematical statements quantifying over an infinite domain).

WHAT IS THIS THING CALLED METAPHYSICS?

Evidence for a mathematical statement is proof or computation. Yet a proof is a proof *of* a mathematical statement. Hence the notion of a proof is not intelligible independently of the notion of a mathematical statement, and thus cannot constitute a reductive class for mathematical statements.

Similar considerations hold for statements about the future or the past. Part of the evidentiary basis for statements about the past is memory, yet a memory is always a memory of a past state of affairs, so cannot be characterized independently of the past. Part of the evidentiary basis for statements about the future is intention, yet an intention is always an intention to do something in the future, so cannot be characterized independently of the future.

Reductionism is not sufficient for anti-realism, since it is possible to take a realist view of statements of the reductive class. In such a case, the truth of a statement of the disputed class may be true in virtue of the truth of some statement or set of statements of the reductive class, but since we are realists about the reductive class, we must be realists about the disputed class.

We will examine the motivation for anti-realism in due course. But it may be useful to look in some detail at two realist/anti-realist case studies which Dummett offers – one about character, the other about the past – in order to get a sense of the dispute and of the various moves open to realist and anti-realist. Dummett describes the case of character as one in which ‘very few people would seriously adopt a realist attitude.’¹⁰ As we shall see, this is far from clear.

CASE STUDY I: CHARACTER

The character trait in question is bravery. In setting up the example, Dummett makes two simplifying assumptions. First, he assumes that there is no vagueness involved in the application of the predicate ‘brave’. Any response to a dangerous situation can be classified either as brave or as not brave. Second, he assumes that a single brave action is enough to justify the ascription of the character trait of bravery to an agent. Neither assumption is true, but Dummett thinks this no great matter, and simplifies his attempt to illustrate the essentials of the realist/anti-realist dispute.

Let us now ask of the imaginary, and recently deceased, Jones whether he was brave or not. Suppose Jones led a sheltered life at a university and never encountered danger. Thus we have no evidence for ‘Jones was brave’ and no evidence for ‘Jones was not brave’. In that case, says Dummett, ‘Jones was brave’ is true only if the counterfactual conditional ‘if Jones had been exposed to danger, he would have acted bravely’ is true. Similarly, ‘Jones was not brave’ is true only if the counterfactual conditional ‘if Jones had been exposed to danger, he would not have acted bravely’ is true.

The untested Jones

Professor Jones led a sheltered life in his university town, ensconced in his book-lined study, rarely venturing out to those parts of town where ruffians and footpads hold sway. Jones never encountered danger. He was never threatened or attacked himself, nor did he witness such outrages being committed against others. Nothing in his behaviour provides evidence for the truth of 'Jones was brave' or for the truth of 'Jones was not brave'. Nor are we in possession of any method that will enable us to decide the issue by a mechanical procedure in a finite time. Let us suppose, further, that no relevant evidence will turn up in the future. Then, says Dummett, it's not the case that 'Jones was brave' is either true or false. The realist, in contrast, says that 'Jones was brave' is a precise, unambiguous declarative sentence of English, true if Jones had the quality of braveness, false otherwise. Since Jones either possessed that quality or not, 'Jones was brave' is either true or false, even if we have no idea which. Our evidence for a statement is one thing, its truth or falsity another.

As Dummett says, we might have indirect evidence for one of these counterfactuals if bravery was associated with other character traits, and we had evidence that Jones possessed some of the accompanying traits. But let us suppose that we have no such auxiliary evidence. In that case, we have no ground for asserting either counterfactual.

Dummett then makes a further claim: a counterfactual conditional cannot be barely true. If a counterfactual conditional is true, it is true in virtue of the truth of some categorical (non-conditional) statement. As Dummett says, this principle is 'intuitively compelling'.¹¹ We can illustrate this principle in the case of a dispositional property such as solubility, which is best understood in terms of a counterfactual. A lump of sugar is soluble if, and only if, were it placed in water, it would dissolve. This counterfactual is true in virtue of some categorical fact about the chemical structure of sugar.

All of the above, we are assuming, is common ground between realist and anti-realist. The argument now proceeds in either of two directions. The anti-realist reasons that neither of our pair of counterfactuals about Jones is true, since there is no true categorical statement about Jones's actual behaviour that grounds either counterfactual. But if neither counterfactual is true, then 'Jones was brave' is neither true nor false. Hence, we have the distinctively anti-realist thought: 'Jones was brave' is neither true nor false because we have no evidence counting for or against the statement, nor any guarantee of ever acquiring such evidence.

In contrast, the realist concedes that the truth of a counterfactual statement must be grounded in the truth of a categorical statement, but holds that one

WHAT IS THIS THING CALLED METAPHYSICS?

of the counterfactuals is grounded in a true (but unknown) categorical statement about Jones's character. Thus, it may be that the counterfactual 'if Jones had been exposed to danger, he would have acted bravely' is true in virtue of the categorical truth (unknown to us) 'Jones was brave'.

The realist regards the counterfactual as true in virtue of some truth about Jones's character, whereas the anti-realist regards Jones's bravery (or lack of it) as grounded in some truth about his behaviour (hence, if there is no such truth about his behaviour, there is no fact of matter as to whether Jones was brave). The realist assumes that 'Jones was brave' is either true or false, even though we can't know which. The anti-realist assumes that since we can't know which, 'Jones was brave' is neither true nor false.

According to the realist, 'Jones was brave' must *be* either true or false, since the man's character – conceived of as an inner mechanism which determines his behaviour – must either have included the quality of bravery or lacked it.¹² However, Dummett adds, 'only a philosophically quite naïve person would adopt a realist view of statements about character.'¹³ Well, naïve or not, the realist view of the matter is intuitive. We think of psychological characteristics as lying behind and giving rise to behaviour. We behave as we do *because* we are in certain mental states (e.g., we raise our voice because we are angry). Our mental states are not to be identified with our behaviour, but rather cause our behaviour. Equally, we have no difficulty with the idea that, due to self-control or external circumstances, another person may have a certain psychological quality even though it never shows up in his behaviour. In contrast, anti-realism distorts our conception of others' mental states and presupposes a naïve and implausible **behaviourism**. What else but a commitment to behaviourism could motivate the anti-realist's restriction of the relevant counterfactual-grounding categorical statements to those about Jones's publicly observable behaviour?

CASE STUDY II: THE PAST

On one view of future contingent statements, e.g., 'there will be a sea battle tomorrow', such statements are deemed neither true nor false. However, philosophers who hold this view do so, not because of anti-realist considerations, but because they hold that the future is unreal. There are no future facts to make our statements about the future true or false now. Traditionally, a similar view has not been held about past-tensed statements since, on most views of time, the past is real. (See Chapter 5.) However, Dummett thinks that an anti-realist argument can be developed for both kinds of statement, though here we will discuss only statements about the past.

In 'Realism', Dummett briefly outlines how a realist/anti-realist dispute about the past can arise. There are presumably many statements about the past for

which we currently have no evidence for or against, and may never have any evidence (e.g., 'Caesar had eggs for breakfast on his thirtieth birthday'). The realist will insist that such statements, if not defective because of vagueness, ambiguity or reference failure, are either true or false. The anti-realist will insist that, in the absence of present or future evidence, such statements are neither true nor false.

In 'The Reality of the Past', Dummett suggests the following strategy for the realist. The realist can justify his understanding of statements about the past by appealing to truth-value links between present-tensed and past-tensed statements. If I say 'Garrett is in his office', this must have the same truth value as the statement made one year hence 'Garrett was in his office a year ago.' It is impossible that these statements have different truth values: hence the truth-value link. According to Dummett, the realist must say that it is 'from an understanding of the truth-value link, as exemplified in such a case, that we derive a grasp of what it is for a statement in the past tense . . . to be true.'¹⁴

Three points are noteworthy. First, Dummett assumes that, if we were realists about the past, we could be so only by appealing to truth-value links connecting present-tensed statements with past-tensed ones. But why should a realist accept this? To concede this is to concede a distinctively anti-realist thought: that we struggle to an understanding of past-tensed statements from evidence available now or in the future. For a realist, in contrast, 'Caesar had eggs for breakfast on his thirtieth birthday' is a perfectly clear, meaningful declarative sentence, intelligible in its own right to speakers of English. It does not stand in need of truth-value links to give it a determinate truth value.

A pattern is beginning to emerge: in the Jones example, Dummett assumed that statements about other minds must be grounded in statements about behaviour. Now he assumes that statements about the past must be grounded in statements about present and future evidence. Both assumptions distort our understanding of other minds and the past. No realist (or anyone else, for that matter) should accept them.

Second, anti-realism about the past is counter-intuitive. We take it that there is one real, actual past, and it either includes Caesar's having eggs for breakfast on his thirtieth birthday or it does not. How then can it fail to be either true or false that Caesar had eggs on his thirtieth birthday?

Third, there is a further oddity. We are supposing that there is no present or future evidence that might bear on 'Caesar had eggs for breakfast on his thirtieth birthday'. The anti-realist claims that the statement expressed by this sentence is neither true nor false. Suppose that we have evidence that Caesar had eggs for breakfast on his thirty-first birthday. Then we are entitled to think of this statement as either true or false (since we now have evidence for its truth). But is it not odd that two such similar statements should differ so markedly with respect to their possession of a determinate truth value?

WHAT IS THIS THING CALLED METAPHYSICS?

Anti-realism about mathematical statements is not implausible in this way. The anti-realist will allow that mathematical statements ranging over a finite domain, however large, can legitimately be thought determinate in truth value. Statements ranging over an infinite domain, for which we have neither proof nor counter-example, cannot be thought of as determinate in truth value. Thus, the anti-realist allows that ‘The first billion billion even numbers are the sum of two primes’ is either true or false, even if no one has actually verified or falsified it, since it could be verified or falsified in a finite time using a mechanical decision procedure. But Goldbach’s conjecture – every even number is the sum of two primes – cannot be thought of as either true or false since, at present, we have neither proof nor counter-example and no guarantee of ever acquiring either. Here, at least, there is a categorical difference between those statements to which an anti-realist is willing to grant determinate truth value, and those to which he is not. Anti-realism about the past allows no such demarcation.

ANTI-REALISM: MOTIVATION AND ASSESSMENT

As will have become evident from the foregoing, the anti-realist project is generated from considerations to do with meaning and truth. But what considerations exactly? In his earlier article, Dummett says little beyond the following: for the anti-realist ‘the meaning of a statement is intrinsically connected with that which we count as evidence for or against the statement.’¹⁵ He is a little more forthcoming in his later article. There the anti-realist:

maintains that the process by which we come to grasp the sense of statements of the disputed class, and the use which is subsequently made of these statements, are such that we could not derive from it any notion of what it would be for such a statement to be true independently of the sort of thing we have learned to recognize as establishing the truth of such statements. . . . In the very nature of the case, we could not possibly have come to understand what it would be for the statement to be true independently of that which we have learned to treat as establishing its truth: there simply was no means by which we could be shown this.¹⁶

It is hard to feel the force of these considerations. Consider again Goldbach’s conjecture that every even number is the sum of two primes. This is a conjecture for which we have, at present, neither proof nor counter-example. Since the even numbers form an infinite series, there is no guarantee that we will ever obtain either. It is thus a conjecture which is in dispute between realist and anti-realist. Some of Dummett’s remarks give the impression that the anti-

realist must hold the sentence ‘every even number is the sum of two primes’ to be meaningless. If meaning is tied to evidence, and we have no evidence for or against this sentence, then is it not meaningless? However, this is not Dummett’s considered opinion, which is just as well since the sentence obviously is meaningful.

Dummett’s complaint is not that the sentence is meaningless, but that we have no conception of ‘what it would be for such a statement to be true independently of the sort of thing we have learned to recognize as establishing the truth of such statements’ (in this case, a proof).¹⁷ In the absence of evidence for or against, and no guarantee of acquiring such evidence, we cannot think of the conjecture as determinately either true or false. In his earlier paper, Dummett expressed this by saying that we should regard a disputed statement, such as Goldbach’s conjecture, as neither true nor false. However, this is amended in the later paper to the claim that we are not entitled to assert that Goldbach’s conjecture is either true or false (where not being entitled to assert P is not the same as being entitled to deny P). This is a subtle but important shift. It would obviously be absurd to hold that Goldbach’s conjecture is neither true nor false simply because we have, at present, neither proof nor counter-example. We would then have shown Goldbach’s conjecture to be untrue simply because we lack a proof or a counter-example. Mathematical results are not that easy to obtain! Simply refusing to assert that Goldbach’s conjecture is either true or false does not have this absurd consequence.

What of the motivation for anti-realism about mathematics? Why think that our conception of what it is for Goldbach’s conjecture to be true depends on our possession of a proof of it or on our ability to recognize a proof if one were presented to us? Such evidential considerations seem irrelevant to questions of understanding. When I come to understand ‘every even number is the sum of two primes’, I take it, as I take any precise, unambiguous declarative sentence of English, to be making a clear and definite claim about reality. The sentence ‘every even number is the sum of two primes’ represents reality (in particular, that portion of reality consisting of the series of even numbers) to be a certain way. The sentence is true just if reality is that way; otherwise it is false. This captures exactly what is involved in my understanding of Goldbach’s conjecture, and it makes no reference to any evidence I may have for or against the conjecture. These remarks are the merest platitudes, yet they serve to cripple the anti-realist project at the outset. In which case, we are quite entitled to regard Goldbach’s conjecture as either true or false, even though we have no idea which truth value it has.¹⁸

Dummett doubtless conceived anti-realism to be an improvement on A. J. Ayer’s logical positivism.¹⁹ Logical positivism was a theory of (literal) sentence meaning. According to it, a declarative sentence was literally meaningful only if it was either analytic (true or false in virtue of meaning alone) or empirically verifiable. Ayer saw himself as part of the great British empiricist tradition

A. J. Ayer (1910–89)

Sir Alfred Jules Ayer was educated at Eton and Christ Church, Oxford. He became Grote Professor at University College, London, before taking up the Wykeham Chair of Logic at Oxford in 1959. A colourful and controversial figure, he had a wide circle of friends and divided much of his time between London and Oxford. He is undoubtedly the only philosopher to have made a record with Lauren Bacall. One of the leading British philosophers of the last century, fame came early in life with the publication of *Language, Truth and Logic* in 1936. The propositions of metaphysics were declared meaningless since neither analytic nor empirically verifiable. Propositions of logic and mathematics, indeed all *a priori* propositions, were counted analytic. Ethical sentences, such as 'torture is wrong', which are neither analytic nor empirically verifiable, were deemed meaningful by virtue of expressing attitudes or emotions, rather than by stating ethical facts. Though Ayer later modified the stark doctrines of his early work, he never relinquished his empiricist outlook.

stretching back to John Locke, and as a slayer of metaphysical dragons. Logical positivism was certainly an **empiricist** theory, and it was a theory that had implications for the meaningfulness of certain sentences of English. Thus consider the sentence 'everything is doubling in size'. This sentence is neither analytic nor empirically verifiable. Even in principle, we could have no sensory evidence for or against the sentence (since there is no neutral vantage point from which it could be assessed). Hence, logical positivism would have to declare this sentence meaningless. Yet the sentence clearly is meaningful. We know exactly what it says.

One supposed advantage of anti-realism was that it would avoid the consequence that a sentence such as 'everything is doubling in size' is meaningless. Rather, the point would be that we are not entitled to think of the sentence as determinately either true or false. But, on reflection, this is hardly less counter-intuitive. There is no reference failure or vagueness or ambiguity in the sentence. It is a declarative sentence of English, making an intelligible claim about reality. How can reality fail to be either the way it describes or some other way? In which case, we cannot but think of it as either true or false. It may be admirable that Dummett should struggle to keep alive the spirit of the great British empiricist tradition, but it is a lost cause nonetheless.²⁰

ANTI-REALISM DISPROVED?

The above discussion has not been kind to the anti-realist polemic. Unfortunately, things are going to get worse. In addition to the counter-intuitive claims

of anti-realism, and its lack of any plausible motivation, there is a powerful argument against the coherence of anti-realism. The argument was originally due to F. B. Fitch, and first published in 1963.²¹

It is a consequence of anti-realism that there are no unknowable truths. If there were unknowable truths, then a statement could be true, and hence determinate in truth value, even though there was no possibility of us ever coming to recognize it as true. It is precisely such independence of mind from reality that the anti-realist wishes to outlaw. Realism, in contrast, is quite consistent with the existence of unknowable truths.

Thus we can take it that the anti-realist is committed to the following principle:

(KP) For all statements p , if p is true, it is possible to know p .

In order to state Fitch’s argument, we need one further assumption. Although the anti-realist cannot countenance unknowable truths, he must allow that some truths are unknown. This is simply an expression of our lack of omniscience. Unlike God, we don’t know all the truths there are. Despite some ambiguous formulations, Dummett’s anti-realist is happy to allow that there are many truths for which we have no evidence (e.g., because no one has bothered to gather the evidence, or because it would take too long for humans to gather it). Thus, for the anti-realist, although there are no unknowable truths, there are many unknown truths. Let q be such an unknown truth.

Fitch’s proof proceeds as follows (where ‘K’ stands for ‘it is known that ...’ and ‘~’ stands for ‘not’):

- | | |
|-------------------------------------|--|
| (1) q and $\sim Kq$ | Assumption |
| (2) Possibly $K(q$ and $\sim Kq)$ | (1) [(KP)] |
| (3) Possibly $(Kq$ and $K\sim Kq)$ | (2) (distribution of knowledge over conjunction and possibility principle) |
| (4) Possibly $(Kq$ and $\sim Kq)$; | (3) (knowledge entails truth and possibility principle) |
| (5) (KP) is false; | (4) <i>reductio</i> |
| (6) Anti-realism is false. | (5) (Anti-realism entails [(KP)]) |

We begin with (1). Premise (1) is an assumption everyone accepts. It expresses our lack of omniscience. Premise (2) follows from (1) by (KP) (substituting ‘ q and $\sim Kq$ ’ for ‘ p ’). (2) says that it’s possible to know $(q$ and $\sim Kq)$. Premise (3) follows from (2) by the principle that knowledge distributes over conjunction, together with the possibility principle. The former says that anyone who knows a conjunction thereby knows each conjunct: from ‘ X knows A and B ’ we may infer ‘ X knows A ’ and ‘ X knows B ’. The latter says that from ‘Possibly R ’ and ‘ R entails S ’ we can infer ‘Possibly S ’.

WHAT IS THIS THING CALLED METAPHYSICS?

Premise (4) follows from (3) by the principle that knowledge implies truth: from 'X knows that A' we may infer the truth of 'A'. This principle is regarded as essential to knowledge, and is one crucial respect in which knowledge differs from 'non-factive' states such as belief (from 'X believes that p' we cannot infer 'p' – X may be wrong).²² The move from (3) to (4) also requires an application of the possibility principle.

But (4) is impossible; hence, some earlier premise or principle must be false. Premise (1) is unassailable; the possibility principle and the principles that knowledge distributes over conjunction and that knowledge implies truth seem undeniable; hence the source of the contradiction must be (KP). So (KP) is false, and since anti-realism entails (KP), anti-realism is false.

Although I am happy to treat Fitch's proof as a straightforward refutation of anti-realism, it has to be said that some philosophers regard Fitch's reasoning as paradoxical and attempt to disclose some fallacy in his argument.²³ To accept Fitch's proof is to accept that if some truths are unknown, then some truths are unknowable. By **contraposition**, this is equivalent to the principle that if all truths are knowable then all truths are known. This principle may be thought counter-intuitive, independently of the realist/anti-realist dispute.²⁴ However, there is no consensus on what is supposed to be wrong with Fitch's proof, and until a consensus emerges it is quite reasonable to treat Fitch's proof as a refutation of (KP) and, hence, as a refutation of anti-realism.

CONCLUDING REMARKS

Dummett's anti-realist project has not fared well in our discussion: it is counter-intuitive, inadequately motivated and open to refutation by Fitch's proof. This conclusion bodes ill for Dummett's wider project of attempting to place the theory of meaning at the centre of philosophy. Dummett tried to delineate a new sense in which the world (or some aspect of it) is not independent of us and our evidence-gathering abilities. In this he failed, but others may yet succeed.²⁵

STUDY QUESTIONS

- What is involved in being a realist about some area?
- How would you characterize Dummett's anti-realism?
- Is it plausible to think that 'Jones was brave' is neither true nor false?
- Can you think of any compelling motivation for anti-realism about some subject matter?
- How might an anti-realist reply to Fitch's proof?

ANNOTATED FURTHER READING

- A. J. Ayer, *Language, Truth and Logic* (Harmondsworth: Penguin) 1976, especially Ayer's Introduction and Chapter 1. Ayer's book is the classic statement of logical positivism, and that doctrine can usefully be seen as a (close) ancestor of anti-realism.
- M. A. E. Dummett, 'Realism' and 'The Reality of the Past', in his collection *Truth and Other Enigmas* (Cambridge, Mass.: Harvard University Press) 1978. Both articles are quite difficult, but 'Realism' is the more accessible of the two.
- D. Edgington, 'The Paradox of Knowability', *Mind*, Vol. 94, 1985. A useful if high-level discussion of Fitch's paradox.
- J. McDowell, 'Criteria, Defeasibility and Knowledge', in J. Dancy (ed.), *Perceptual Knowledge* (Oxford: Oxford University Press) 1988. A nice critique of the epistemological underpinnings of anti-realism. Clearly written, but demanding.
- C. Wright, *Realism, Meaning and Truth*, second edition (Oxford: Basil Blackwell) 1993. The articles in this collection are all concerned with the realist/anti-realist debate but too difficult for the novice. However, Wright's Introduction is indeed a genuine introduction to the debate and can be recommended to the enthusiastic tyro.

INTERNET RESOURCES

- B. Brogaard and J. Salerno (2004) 'Fitch's Paradox of Knowability', *The Stanford Encyclopedia of Philosophy* (Summer 2004 edition), Edward N. Zalta (ed.). Retrieved 31 May 2006 from <<http://plato.stanford.edu/archives/sum2004/entries/fitch-paradox>>.
- E. Craig (1998) 'Realism and Antirealism', in E. Craig (ed.), *Routledge Encyclopedia of Philosophy*. Retrieved 31 May 2006 from <<http://www.rep.routledge.com/article/N049>>.
- A. Miller (2005) 'Realism', *The Stanford Encyclopedia of Philosophy* (Fall 2005 edition), Edward N. Zalta (ed.). Retrieved 31 May 2006 from <<http://plato.stanford.edu/archives/fall2005/entries/realism>>.
- B. Taylor (1998, 2004) 'Dummett, Michael Anthony Eardley', in E. Craig (ed.), *Routledge Encyclopedia of Philosophy*. Retrieved 31 May 2006 from <<http://www.rep.routledge.com/article/DD083>>.

GLOSSARY

a priori

Designates the way in which a sentence or statement is known. A sentence is known *a priori* just if understanding the sentence alone suffices to know its truth-value. 'All bachelors are men' is known *a priori*; 'water boils at 100 degrees celsius' is not. Though related, the notion of the *a priori* should be distinguished from the notions of necessity and analyticity (truth in virtue of meaning). Knowledge which is not *a priori* is *a posteriori* or empirical.

abstract

This word has many different meanings, but two are important for our purposes. When philosophers describe universals or numbers as abstract objects, they mean abstract in the sense of not in space or time. When trope theorists describe tropes as abstract particulars, they mean abstract in the sense of 'fine' or 'diffuse'. A billiard ball's redness trope is not thought to exist outside space and time.

accidental/essential

This distinction derives from Aristotle. The properties of an object can be divided into those that are essential and those that are accidental. If F is an accidental property of x, then, though x is F, it might not have been F. In contrast, if G is an essential property of x, then x could not but have been G. In possible worlds talk, if F is an accidental property of x, there are possible worlds in which x exists but is not F; if G is an essential property of x, there are no worlds in which x exists but is not G. W. V. Quine famously tried to discredit the distinction, calling it 'invidious'. (See Quine's paper 'Reference and Modality' in his *From a Logical Point of View* [New York: Harper & Row] 1963.)

analytic

A sentence is said to be analytic when it is true (or false) in virtue of meaning alone. Examples include 'all bachelors are men', 'all spinsters are women', 'all triangles

GLOSSARY

have three sides'. A non-analytic or synthetic sentence is true (or false) in virtue of its meaning and the worldly facts. Examples include 'all bachelors wear trousers', 'all spinsters are miserable', 'triangles are my favourite geometric object'. In his famous 1952 paper 'Two Dogmas of Empiricism', the Harvard philosopher W. V. Quine questioned the pedigree and significance of the analytic/synthetic distinction. (The paper is reprinted in his collection *From a Logical Point of View* [New York: Harper & Row] 1963.)

antecedent

In a conditional of the form 'if P then Q', P is the antecedent (and Q is the consequent).

behaviourism

There are many varieties of behaviourism, but the key idea is that the mind can be reduced to (identified with) behaviour or dispositions to behaviour, where behaviour is understood in non-mentalistic terms. However, such an identification violates a common sense datum, *viz.*, that mental states and events cause behaviour and dispositions to behaviour. If my itch causes me to scratch, it cannot be identified with that scratching (since nothing causes itself). In addition, behaviourism misdescribes the way we know of our own mental states. I know that I have an itch by having it, not by observing my behaviour.

causal loops

Causal loops are an exotic possibility. If travel into the past is possible, a man might travel into the past, impregnate his mother, and so be his own father. The man's existence would form a causal loop. Each event on the loop is caused by another event in the loop, yet the existence of the entire loop is uncaused.

conditional

A conditional is any sentence of the form 'if P then Q', where P is the antecedent and Q the consequent.

contingent

A sentence is contingent if it is true in some possible circumstances (or possible worlds) and false in others. Thus 'it rained in Edinburgh on 1 January 2006' is contingent: it is true, but it might have been false. A sentence is non-contingent or necessary if it is either true in every possible circumstance or false in every possible circumstance. Some also want to talk of contingent beings (such as you and me) and necessary beings (such as God or the number 2).

contraposition

Any conditional is logically equivalent to its contrapositive. That is, 'if p then q' is equivalent to 'if not-q then not-p'. (Note that 'if p then q' is not equivalent to 'if not-p then not-q'.)

counterfactual conditional

A counterfactual conditional is a conditional with a false antecedent which states what would have been the case had the antecedent been true. Thus, I may not throw a brick at the window, but we can still truly say: if I had thrown a brick at the window, the window would have smashed. We all use and understand counterfactuals, but there is much dispute about their underlying logic.

empiricist

The classic empiricist philosophers were Locke, Berkeley and Hume. The central tenet of empiricism is that substantial or worldly knowledge can be gained only *via* the five senses. Reason may indeed yield knowledge (e.g., of the *a priori* truths of logic and arithmetic), but such knowledge is not substantial. Rationalists hold that, to the contrary, reason can yield knowledge of the world. An obvious example of a rationalist argument is the ontological argument that attempts to prove the existence of God by reason alone.

fundamental

We are concerned with 'fundamental' as it occurs in sentences of the form 'X is more fundamental to Z than Y'. The comparative form 'more fundamental' is unusual. In his discussion, McTaggart does not mean 'essential' by 'fundamental'. He seems to mean 'essential and complete'. Thus even if X and Y are both essential to Z, X is more fundamental if the X-facts exhaust the Z-facts, but the Y-facts do not. In this sense, McTaggart thinks that the A series is more fundamental to time than the B series.

general causal claims

These are causal claims relating types of event or occurrence. Thus 'HIV causes AIDS' is an example of a general causal claim. I may utter this sentence without having any particular person in mind.

if and only if

A sentence of the form 'P if and only if Q' is equivalent to the conjunction 'if P then Q and if Q then P' and is therefore true only when P and Q have the same truth value.

indeterministic

On some theories, quantum phenomena are indeterministic. That is, what happens at one time at the quantum level may not be fully determined by what happened at previous times. A certain outcome may only be likely but not determined. Such indeterminism is thought to be a consequence of Werner Heisenberg's uncertainty principle.

GLOSSARY

indexical

A word is an indexical just in case its reference is determined (in part) by the context of its utterance. Thus, an utterance of 'I' is indexical since its reference is determined by the identity of its utterer; an utterance of 'here' is indexical since its reference is determined by the location of its utterer; an utterance of 'now' is indexical since its reference is determined by the time of utterance. 'I', 'here', and 'now' are referring terms, yet have the curious feature of immunity to reference-failure or misreference.

indicative conditional

An indicative conditional is different from a counterfactual. It merely states what will happen or has happened given the truth of some antecedent condition. Thus 'if Bill comes to the party then there will be a scene' and 'if there are footprints in the sand then Fred was here last night' are both indicative conditionals.

laws of nature

A standard example of a law of nature is: all metals expand when heated. But what are laws of nature? Some philosophers (the Humeans) think that laws of nature are simply well-established regularities. Others (anti-Humeans) think that laws involve some kind of necessity which explains the observed regularities. A genuine law tells us what *must* occur.

Leibniz's law

The law of logic which states that if A is identical to B then every property of A is a property of B and *vice versa*. It is sometimes stated without reference to properties as: if A is B then anything true of A is true of B and *vice versa*. This law must be distinguished from the principle of the identity of indiscernibles: if A and B share all their properties, then A is B.

logical structure

Modern interest in the logical structure of natural language sentences was stimulated by the work of Frege, Wittgenstein and Russell in the late nineteenth and early twentieth centuries. Some thought that understanding the logical structure of language would reveal the structure of the world. But even if that hope is abandoned, there is still reason to be interested in questions of logical structure. Getting clear about the logic of our language may help us avoid an over-inflated ontology (see Russell's critique of Meinong) or may help us explain inferences that are otherwise puzzling (e.g., why 'Bill ran quickly' implies 'Bill ran').

logically necessary

For all propositions P and Q, P is said to be logically necessary for Q just in case it is logically impossible for Q to be true and P false (that is, just in case Q entails P). If P is necessary for Q, Q is sufficient for P. It is logically impossible for Q to

be true and P false just if the conjunction (Q and not-P) is contradictory. Theists who advocate the 'necessary evils' response to the problem of evil hold that a conjunction such as (benevolence without suffering) is contradictory.

material conditional

The phrase 'material conditional' is a technical term of logic, represented by ' \supset '. A sentence ' $P \supset Q$ ' is false just in case P is true and Q is false, otherwise it is true. In other words, ' $P \supset Q$ ' is equivalent to 'not-P or Q'. It is uncontroversial that counterfactual conditionals are not material. (If they were, all counterfactuals would be true, which they're not.) Some people think that indicative conditionals are material conditionals. But this is controversial. Is an indicative conditional of ordinary English true whenever its antecedent is false? Is 'if $2 + 2 = 5$ then I'm a Chinaman' true?

McTaggart changes

These are changes that objects undergo simply in virtue of the passing of time (e.g., being one hour older). Shoemaker wants to exclude these changes from consideration, otherwise time without change would, by definition, be impossible. In our world McTaggart changes do no causal work, so on one understanding of 'change', they don't count as genuine changes.

mental and/or bodily continuity

Some philosophers think that the identity of a person over time can be understood in terms of bodily continuity, while others prefer to understand personal identity in terms of psychological continuity. On the bodily view, I continue to exist just as long as my living human body continues to exist. On the psychological view, I continue to exist just as long as my stream of mental life persists (my beliefs, memories, hopes, fears, etc.). Defenders of the psychological view hold that their view best fits with our concept of a person (that is, the concept of a certain kind of psychological being).

modal

Pertaining to possibility and necessity. Modal sentences are those of the form: possibly P, necessarily P, A might have been F, A is necessarily G, B can't be G, etc. Modal claims have different strengths depending on the modality in question. Thus 'I can't lift that car' refers to a physical impossibility (i.e., my lifting the car is incompatible with facts about my physique and the laws of nature). 'I can't lift and not lift that car' refers to a logical impossibility (my simultaneously lifting and not lifting the car is incompatible with the laws of logic). There are other modalities too (e.g., legal: 'you can't park there'). Note also that we can distinguish *de dicto* from *de re* modal sentences. In the *de dicto* sentence 'necessarily $2 + 2 = 4$ ', necessity is predicated of a sentence or proposition. In the *de re* sentence 'Socrates is necessarily human', necessary or essential humanity is predicated of a non-propositional object, Socrates.

GLOSSARY

necessary and sufficient conditions

The aim of conceptual analysis has traditionally been the production of necessary and sufficient conditions for the application of some concept. The most famous example is Plato's analysis of knowledge as justified true belief (*Theaetetus* 201c–210d). It was assumed that the concepts of belief, truth, and justification could be understood independently of the concept of knowledge. The analysis thus purports to provide a reductive analysis of the concept of knowledge.

necessary being

God is traditionally conceived to be a necessary being. That is, it is impossible that he not exist. In possible world talk, he exists in every possible world. Some philosophers, e.g., Bertrand Russell, have questioned the coherence of the phrase 'necessary being', but on dubious grounds.

no action at a temporal distance

Temporal analogue of 'no action at a spatial distance'. To give up the temporal principle is to hold that A at t1 can bring about B at t2 without there being any event after t1 and before t2 sufficient to bring about B. Just as many believe that action at a spatial distance is possible, Shoemaker thinks that action at a temporal distance is possible too.

non-branching

A relation branches when it holds between one thing and two or more later things. The fission of an amoeba is a real-life example of physical continuity holding in a branching form. Psychological continuity can also hold in a branching form, as the case of 'Fission' shows. Since one thing cannot be identical to two things, some criteria of identity stipulate that identity obtains only when there is no branching of the relevant sort.

numerical identity

This sense of 'identity' is expressed in sentences such as 'water is H₂O', 'Superman is Clark Kent', '2 is the positive square root of 4'. Each of these sentences concerns just one entity, variously described or named. Numerical identity conforms to Leibniz's law. That is, if 'A is B' expresses numerical identity, then whatever is true of A is true of B and *vice versa*.

omnipotent

God is held to be all-powerful or omnipotent. That is, God is able to bring about any logically possible state of affairs. Thus, it does not tell against God's omnipotence that he cannot make my table be simultaneously round and square, since this is not a logically possible state of affairs. Some have seen a paradox here: can God create a stone so heavy that he cannot lift it? Say 'yes', and God seems to lack omnipotence; say 'no', and again God seems not to be omnipotent.

But this is a pseudo-paradox. There is no contradiction in the idea of one of us creating a stone so heavy that we cannot lift it, but the description of this task, as a task for an omnipotent being, is contradictory, and so logically impossible. Hence, it doesn't count against God's omnipotence that he cannot create a stone so heavy that he cannot lift it.

omniscience

God is traditionally held to be all-knowing. That is, for any true proposition *p*, God knows that *p*. Some have questioned whether this can be correct. If (as some believe) God is outside of time, can he know that it is now 4 p.m.? Can such a truth not be grasped only by a being in time? Or again I know that I'm tired. Can God grasp this truth, or is it graspable only by me? The answers to these questions may await developments in the philosophy of language.

Platonism

Plato held that the forms (the true realities) exist outside space and time. To be a Platonist about some range of entities is thus to hold that those entities exist outside space and time. (As noted in Chapter 9, Dummett advocates a slightly different take on the idea behind Platonism.)

projection

Projectivism about some property *F* is the view that we mistakenly take *F* to be a property of things in the external world, when in fact it is generated by our own minds and projected onto the world. Thus, we may describe a situation as fearful, but it is so only because we react to it in a certain way. Another example: some philosophers think that the world itself can be vague, but this seems a projection from our (vague) concepts to a (precise) reality. Hume is the classic source for projectivism. He wrote of 'the mind's propensity to spread itself upon external objects' (*Treatise*, 1.3.14 para 24).

qualitative identity

This sense of 'identity' is expressed in sentences such as 'they are identical twins' and 'we both drive the same car', etc. The latter sentence concerns two cars not one, both made by the same manufacturer. Qualitative identity does not conform to Leibniz's law: if *A* and *B* are identical twins, it is not the case that whatever is true of *A* is true of *B* and *vice versa*. For example, the twins differ in the exact time of their births and in their subsequent spatial paths.

quantifiers

These are words which tell us what proportion or quantity of things have a certain property. Thus all of the following are answers to the question 'How many *F*s are *G*s?': all *F*s are *G*s; some *F*s are *G*s; most *F*s are *G*s; many *F*s are *G*s; a few *F*s are *G*s; at least twenty-four *F*s are *G*s; no *F*s are *G*s, etc. The development of

GLOSSARY

quantificational logic by Frege in the nineteenth century represented a major advance over previous systems of logic.

reduction

In the debate about time, what does it mean to say that tensed truths (formulated in A series vocabulary) can be reduced to tenseless ones (formulated in B series vocabulary), or that tensed facts can be reduced to tenseless ones? Though the B theorist may admit tensed truths and facts alongside tenseless truths and facts, he cannot regard them as having equal status. For the B theorist, reality is tenseless: its correct description does not require tensed truths. Thus, although it is true that Hitler's death is past, and a fact that it is so, this is because any contemporary utterances of 'Hitler's death is past' are later than Hitler's death. On the B theory, tenseless truths are ultimate and irreducible; tensed truths are dependent and eliminable.

reductive

Words such as 'reductive' and 'reductionism' are philosophers' terms of art, which have many different meanings. Parfit's idea is that we are reductionists about Fs if we think that reality can be completely described without reference to Fs. Many are reductionists, in this sense, about social entities, such as nations or committees. Thus although there are truths about committees (e.g., 'the committee unanimously voted to appoint Smith'), a description of reality that doesn't refer to committees, but only to the actions of individual members, may leave no truth out. Truths about committees will be accounted for by truths about individuals. Parfit thinks that we ought to be reductionists about persons. Truths about persons can be accounted for by truths about bodies and mental states.

singular causal claims

These are causal claims relating particular, unrepeatable events. Thus 'Mary's frenzied attack caused Bill's death' is an example of a singular causal claim.

substance

This notion, as used here, derives from Aristotle. Concrete entities, and especially biological entities (Plato, Socrates, that horse, this tree, etc.) are the 'primary substances' for Aristotle. If the world had contained no primary substances, it would have contained nothing. 'Secondary substances' such as horseness and redness are dependent substances. If there were no horses or red things, horseness and redness would not have existed. This helps explain Aristotle's disagreement with Plato over the nature of universals.

token-reflexive expressions

The phrase 'token reflexive' was coined by the German philosopher of physics, Hans Reichenbach. Also called 'indexicals', these are words (such as 'I', 'you',

'here', 'there', 'now', 'today', 'yesterday', etc.) whose reference, when uttered, is determined by contextual factors. The contextual factors are standardly people ('I', 'you'), place ('here', 'there'), and time ('now', 'today', 'yesterday'). In the case of 'I'-utterances, the utterer is the relevant contextual factor; in the case of 'here'-utterances, the place of utterance is the relevant contextual factor, and so on. This explains why 'I' in my mouth refers to me, but 'I' in your mouth refers to you. Vary the context and the object of reference varies.

truth-value links

Truth-value links occur whenever two languages or ways of speaking are systematically related. For example, first-person and third-person utterances are systematically related. 'I am bald' said by person X is true if and only if 'X is bald', uttered by anyone, is true. Or again, 'today it is raining', uttered on day 1, is true if and only if 'yesterday it was raining', uttered on day 2, is true. Note that the equivalence in truth value is necessary. It is not a coincidence that the pairs of utterances have the same truth value.

'yes or no' answer

There are some questions which do not receive a 'yes or no' answer. For example, if we ask 'Is he bald?' of some man who has some hair on his head, but not much, that question may receive no answer. An important feature of questions which lack a 'yes or no' answer is that no further factual information would enable us to answer them. The lack of any definite answer is the result of our concepts not the world. Parfit thinks that some questions of personal identity lack a 'yes or no' answer.

NOTES

1 GOD

- 1 Anselm *Proslogion II* in *St Anselm's Proslogion*, M. Charlesworth (ed.) (Oxford: Oxford University Press) 1965, p. 116. Translation due to W. Mann 'The Ontological Presuppositions of the Ontological Argument', *Review of Metaphysics*, 26, 1972 pp. 260–1.
- 2 Gaunilo, 'On Behalf of the Fool' in *St. Anselm's Proslogion*, pp. 163–5.
- 3 'Must God Exist?' in *Philosophy in the Open* (Milton Keynes: Open University Press) 1978, p. 115.
- 4 *Ibid.*, pp. 119–120.
- 5 *Ibid.*, p. 115.
- 6 See S. Kripke, *Naming and Necessity* (Oxford: Basil Blackwell) 1980.
- 7 This fallacy should be distinguished from the fallacy of division, *viz.*, inferring that each member of a totality has a property because the totality has it.
- 8 The conclusion that the cause of the universe lies outside the universe does not by itself imply that the cause of the universe is a necessary being. However, proponents of the cosmological argument typically assume that anything existing outside the universe is a necessary being.
- 9 See the discussion of David Lewis' modal realism in Chapter 2.
- 10 This is not a contradiction. That something has always existed implies nothing about its modal status (i.e., whether it is necessary or contingent). Contingent, eternal objects are indeed hard to find, but the category is a coherent one.
- 11 W. Paley, *Natural Theology*, ed. F. Ferre (New York: Bobbs-Merrill Co. Inc.) 1963, pp. 1–2.
- 12 *Ibid.*, p. 2.
- 13 D. Hume, *Dialogues Concerning Natural Religion*, ed. N. Kemp Smith (London: Nelson) p. 168.
- 14 *Ibid.*, p. 167.
- 15 It is sometimes suggested that, though a man may suffer on Earth, he has an immortal soul and will enjoy infinite happiness in the afterlife. But is this really a reply to the argument from evil? How does it justify earthly suffering?

NOTES

- 16 Here I am indebted to J. L. Mackie's excellent article 'Evil and Omnipotence', *Mind*, Vol. 64, 1955, pp. 200–12.
- 17 The modal realist would not accept this way of putting it. (See Chapter 2.)

2 EXISTENCE

- 1 I here ignore the question of whether past and future objects exist. See Chapter 5.
- 2 D. Lewis, *On the Plurality of Worlds* (Oxford: Basil Blackwell) 1986, pp. 2–3.
- 3 This is a source of potential criticism. Lewis excludes, by definition, the possibility that a single world might contain two or more spatio-temporally unconnected regions. It might be thought that this issue should not be settled by *fiat*.
- 4 *Ibid.*, p. 73.
- 5 D. Parfit, 'The Puzzle of Reality: Why Does the Universe Exist?', reprinted in P. van Inwagen and D. Zimmerman (eds) *Metaphysics: The Big Questions* (Oxford: Basil Blackwell) 1998, p. 419. (Parfit's article is followed by a useful reply from R. Swinburne, pp. 427–9.)
- 6 Parfit, *op. cit.*, pp. 419–20.
- 7 S. Kripke, *Naming and Necessity* (Oxford: Basil Blackwell) 1980, p. 45, n. 13.
- 8 C. McGinn, *Logical Properties* (Oxford: Oxford University Press) 2000, p. 16.
- 9 B. Russell, 'Descriptions' in his *Introduction to Mathematical Philosophy* (New York: Simon & Schuster) 1961, p. 169.
- 10 Russell's article is reprinted in his *Logic and Knowledge* (London: Unwin Hyman Ltd) 1956, pp. 39–57.
- 11 Hence, for Russell, 'reference-failure' is an oxymoron. If a grammatical singular term can be meaningful in the absence of its object, then it is not a referring term. For this reason, Russell concluded that definite descriptions and ordinary proper names were not referring terms.
- 12 See Kripke, *op. cit.*, especially Lectures I and II.
- 13 On this view, existence is a property, but of concepts not objects. Thus the concept 'George Bush' has the property of being instantiated.
- 14 A. J. Ayer, *Language, Truth and Logic* (Harmondsworth: Penguin) 1976, p. 58.
- 15 Or, at least, a problem for anyone who rejects the view that proper names such as 'Superman' are disguised descriptions. On the description theory of names, criticized by Kripke, 'Superman does not exist' is rendered as ' $\sim\exists xFx$ ', where 'F' is some description associated with 'Superman'. Such a rendering is neither contradictory nor requires the non-existent for its truth.
- 16 In addition, the quantifier view seems better placed to handle general existential sentences, such as 'tigers exist'. This sentence is rendered as ' $\exists x(x \text{ is a tiger})$ ', and utterances of 'tigers exist' express the same statement however many tigers exist. But how does the property view understand 'tigers exist'? Exactly what is being said to possess the property of existence? Presumably either the totality of tigers, or each tiger individually. Either way, this seems to imply that an utterance of 'tigers exist' would have expressed a different statement had fewer tigers existed.

3 UNIVERSALS AND PARTICULARS

- 1 Universals are often distinguished from particulars by their ability to ‘be wholly present in two or more places at the same time’. But this can’t be quite right if there are particulars that exist outside space and time. On some views, numbers are outside space and time, and they instantiate universals: e.g., each even number instantiates the universal *evenness*. But *evenness* is not wholly present in two places at the same time, since numbers do not exist in space or time. This consideration shows that the concepts ‘multiple instantiation’ and ‘multiple location’ are not the same.
- 2 If it is mysterious how a particular instantiates a universal, it is even more mysterious how a lower-order universal instantiates a higher-order universal (as *redness* instantiates *colour*). Is there a single notion of instantiation in both cases?
- 3 Talk of states of affairs is not a *façon de parler*. A traditional property realist conceives of the state of affairs of A’s being F as an addition to ontology over and above the existence of A and F. More is required for A’s being F than just A and F: a world might contain A and F, yet not contain A’s being F, since A is not F in that world, though other things are. And, for the same reason, the more required is not another entity, e.g., corresponding to the ‘is’ of instantiation. (See D. M. Armstrong, *Truth and Truthmakers* (Cambridge: Cambridge University Press) 2004, chapter 4.)
- 4 B. Russell, *Human Knowledge, its Scope and Limits* (New York: Simon & Schuster) 1948. Reprinted in *Metaphysics: The Big Questions*, ed. P. Van Inwagen and D. Zimmerman (Oxford: Basil Blackwell) 1999, p. 54.
- 5 *Ibid.*, p. 57.
- 6 B. Russell, *The Problems of Philosophy* (Oxford: Oxford University Press) 1978, p. 55.
- 7 Here ‘abstract’ does not mean ‘outside space and time’ but ‘fine’, ‘subtle’ and ‘diffuse’. See D. C. Williams, ‘On the Elements of Being: 1’ in D. H. Mellor and A. Oliver (eds), *Properties* (Oxford: Oxford University Press) 1997, p. 115. See also K. Campbell, *Abstract Particulars* (Oxford: Basil Blackwell) 1990.
- 8 Williams, *op. cit.*, p. 116.

4 CAUSATION

- 1 See, e.g., ‘Causal Relations’ reprinted in his *Essays on Actions and Events* (Oxford: Oxford University Press) 1985, pp. 149–62.
- 2 D. H. Mellor, *The Facts of Causation* (London: Routledge) 1995.
- 3 *Ibid.*, p. 132.
- 4 A relation R is reflexive just if: necessarily for any object a, aRa. A relation is symmetric just if: necessarily for any objects a and b, if aRb then bRa. A relation is transitive just if: necessarily for any objects a, b and c, if aRb and bRc then aRc. The relation ‘is as tall as’, for example, is reflexive, symmetric and transitive. Necessarily: any object is as tall as itself; if a is as tall as b, b is as tall as a; and if a is as tall as b and b is as tall as c then a is as tall as c. The relation ‘loves’ is, alas, neither reflexive, symmetric nor transitive.

NOTES

- 5 The example is due to Ned Hall. See 'Causation and the Price of Transitivity', *Journal of Philosophy*, Vol. 97, 2000, pp. 198–222.
- 6 J. S. Mill, *A System of Logic* (New York: Harper & Brothers) 1846, p. 198.
- 7 D. Lewis, 'Causation', reprinted in his *Philosophical Papers* Vol. II (Oxford: Oxford University Press) 1986, p. 162. Lewis's view here fits with his counterfactual theory of causation: wherever there is counterfactual dependence there is causation.
- 8 D. Hume, *A Treatise Concerning Human Nature*, in A. J. Ayer and R. Winch (eds), *British Empirical Philosophers* (London: Routledge) 1965, p. 359.
- 9 *Ibid.*, p. 359.
- 10 D. Hume, *An Enquiry Concerning Human Understanding*, ed. L. A. Selby-Bigge (Oxford: Oxford University Press) 1962, p. 74.
- 11 *Ibid.*, pp. 74–5.
- 12 *Ibid.*, p. 75.
- 13 *Ibid.*, p. 76.
- 14 T. Reid, *Essays on the Active Powers of the Human Mind*, ed. Baruch A. Brody (Cambridge, Mass.: MIT Press) 1969, p. 334.
- 15 See G. E. M. Anscombe 'Causality and Determination' and C. J. Ducasse 'On the Nature and the Observability of the Causal Relation', both reprinted in E. Sosa and M. Tooley (eds), *Causation* (Oxford: Oxford University Press) 1993, pp. 88–104 and pp. 125–36, respectively.
- 16 Anscombe, *op. cit.*, p. 92.
- 17 This objection is due to R. Taylor. See his *Metaphysics* (Englewood Cliffs, NJ: Prentice Hall) 1992, Chapter 10.
- 18 Hume, *Enquiry*, p. 76.
- 19 See D. Lewis, *op. cit.*
- 20 Overdetermination is to be distinguished from a case where two people together produce a certain effect, but where the effect would not have happened if either one of the people had not acted as they did. In such a case, we understand how D and E together cause F, even though D didn't cause F and E didn't cause F. D didn't cause F because D wasn't enough, by itself, to cause F. But that is precisely what one cannot say in a case of overdetermination.

5 TIME: THE FUNDAMENTAL ISSUE

- 1 J. M. E. McTaggart, 'Time Is Not Real', reprinted in R. C. Hoy and N. Oaklander (eds), *Metaphysics: Classic and Contemporary Readings* (Belmont, Calif.: Wadsworth) 1991, pp. 43–4.
- 2 *Ibid.*, p. 45.
- 3 *Ibid.*, p. 45.
- 4 *Ibid.*, p. 45.
- 5 Quoted in McTaggart, *ibid.*, p. 45.
- 6 B. Russell, *Principles of Mathematics* (Cambridge: Cambridge University Press) second edition, Allen & Unwin, 1937, section 442.
- 7 McTaggart, *op. cit.*, p. 46.
- 8 *Ibid.*, p. 46.
- 9 *Ibid.*, p. 48.

- 10 Ibid., p. 48.
- 11 Ibid., p. 49.
- 12 M. Dummett, 'A Defence of McTaggart's Proof of the Unreality of Time', in his *Truth and Other Enigmas* (Cambridge, Mass.: Harvard University Press) 1978, pp. 351–8.
- 13 Ibid., p. 356.
- 14 Ibid., p. 357.
- 15 P. Horwich, *Asymmetries in Time: Problems in the Philosophy of Science* (Cambridge, Mass.: MIT Press) 1987, Chapter 2.
- 16 Ibid., p. 22.
- 17 Ibid., pp. 22–3.
- 18 Dummett, op. cit., p. 354.
- 19 McTaggart outlines and criticizes Broad's view: McTaggart op. cit., pp. 49–51.
- 20 Broad's view seems preferable to presentism. Both views agree that the future is unreal, but presentism holds that the past is unreal too. But what does it mean to hold that the past is unreal? It cannot simply amount to the denial of the thesis that past events are happening now, since no one holds that thesis. It must then be the denial of the thesis that the past was real. But this is counter-intuitive, since we agree that, e.g., Socrates did drink hemlock, the truth of which presupposes that past objects were real.
- 21 Does any theory which denies the reality of the future imply the falsity of determinism (see Chapter 7)? Also, aside from presentism, the other versions of the A theory (i.e., those versions – Broad's and McTaggart's – which maintain the reality of the past) seem compatible with the claim that, e.g., it is now 3008 and we are living in the past. Is it not absurd that a theory of time should leave this open as a possibility?
- 22 Relativity theory aside, the absoluteness of simultaneity should never have been considered a necessary *a priori* truth. Michael Dummett supplies a nice example: 'Imagine that there were a permanent fog separating the Old World from the New, and that no regularity were detectable in the time taken to travel from one to the other: sometimes only a few days would have passed in the other hemisphere after a traveller's last visit there, sometimes many years. It would then be senseless to ask what had been happening in one half of the Earth at the very time that something happened in the other; yet each half would have its own history', M. Dummett *Truth and the Past* (New York: Columbia University Press) 2004, p. 86. Compare Dummett's discussion with Anthony Quinton's in 'Spaces and Times' reprinted in R. Le Poidevin and M. MacBeath (eds), *The Philosophy of Time* (Oxford: Oxford University Press) 1993, pp. 203–21. Note that such relativity seems compatible with the idea that each hemisphere has its own *now* (i.e., such relativity does not straightforwardly contradict the A theory).

6 TIME: THREE PUZZLES

- 1 A. N. Prior, 'Thank Goodness That's Over', *Philosophy*, Vol. 34, 1959, p. 17.
- 2 H. Mellor, *Real Time* (Cambridge: Cambridge University Press) 1985, p. 48.
- 3 See D. Parfit, *Reasons and Persons* (Oxford: Oxford University Press) 1986, Chapter 8.

NOTES

- 4 This is not to ask whether there could be a possible world in which time passes and nothing ever happens. We are asking whether, in a world in which many things happen, there could be a period (however small) during which nothing happens.
- 5 S. Shoemaker, 'Time without Change', reprinted in R. Le Poidevin and M. Macbeath (eds), *The Philosophy of Time* (Oxford: Oxford University Press) 1993, pp. 63–79.
- 6 Aristotle, *Physics*, Book 4, Chapter 11, 218b.
- 7 Shoemaker, *op. cit.*, p. 67.
- 8 Assuming that the law governing this universe is not that the freezes occur as noted except that 'all three regions skip a freeze every fifty-nine years' (Shoemaker, *op. cit.*, p. 72). But it seems justifiable to believe in the simpler exceptionless law.
- 9 Shoemaker, *op. cit.*, p. 75.
- 10 D. Lewis, 'The Paradoxes of Time Travel', *American Philosophical Quarterly*, Vol. 13, No. 1, 1976, pp. 145–52.
- 11 *Ibid.*, p. 145.
- 12 *Ibid.*, p. 146. There are even more exotic possibilities. 'An infinite amount of personal time can be squeezed into two minutes of external time. During the first minute, the psuedo-immortal lives the first day of his life. During the next half-minute, the psuedo-immortal lives the second day. During the following quarter-minute, a third day passes. Since there are infinitely many junctures in this sequence, the psuedo-immortal will enjoy infinitely many personal days' (R. Sorenson, 'The Cheated God: Death and Personal Time', *Analysis*, Vol. 65, No. 2, 2005, pp. 122–3).
- 13 Lewis, *op. cit.*, pp. 146–7.
- 14 A case which might present a problem is that of an instantaneous time traveller, e.g., someone who travels back to 1900 but takes no amount of personal time to do so. Such a person's 'world line' is discontinuous rather than continuous. Still, we might be able to justify the claim that the person who left in 2006 is the same as the person who arrived in 1900, rather than a mere duplicate, by appeal to causal factors. If the mental and physical characteristics of the 1900 person were caused by his 2006 states, then talk of the same person may be justified.
- 15 *Ibid.*, p. 148.
- 16 *Ibid.*, pp. 148–9.
- 17 *Ibid.*, p. 149. For some interesting reflections on causal loops, see R. Hanley, 'No End in Sight: Causal Loops in Philosophy, Physics and Fiction', *Synthese*, Vol. 141, 2004, pp. 123–52.
- 18 *Ibid.*, p. 149.
- 19 *Ibid.*, p. 149.
- 20 *Ibid.*, p. 150.

7 FREE WILL

- 1 D. Lewis, 'The Paradoxes of Time Travel', *American Philosophical Quarterly*, Vol. 13, No. 1, 1976, p. 151.
- 2 For example, E. Craig, 'Fatalism' in E. Craig (ed.) *Routledge Encyclopedia of Philosophy* (London: Routledge) 1998.
- 3 M. Dummett, 'Bringing about the Past', in his *Truth and Other Enigmas* (Cambridge, Mass.: Harvard University Press) 1978, p. 339.

- 4 Ibid., p. 338. This argument is intended to show the irrationality of retrospective prayer even if God exists and backwards causation is possible.
- 5 So-called after Aristotle's famous discussion in *De interpretatione*, ed. C. W. A. Witaker (Oxford: Oxford University Press) 2002, Book 9, sections 18a–19b.
- 6 In presupposing the 'open future', the Aristotelian solution presupposes the falsity of determinism.
- 7 A. J. Ayer, 'Fatalism', in his *The Concept of a Person* (London: Macmillan) 1963, pp. 252–3.
- 8 S. Cahn, 'Fatalism', in J. Kim and E. Sosa (eds), *A Companion to Metaphysics* (Oxford: Basil Blackwell) 1995, p. 169.
- 9 Some think the most pressing problem concerns the conditions for moral responsibility. On one view, we first decide what it is for an agent to count as morally responsible and then fashion a notion of free will to fit. In his landmark essay 'Freedom and Resentment' (in *Freedom and Resentment and Other Essays*, London: Methuen, 1974), P. F. Strawson focuses on our practice of holding people responsible, and related practices, concluding that such practices draw on no external condition (such as determinism, or indeterminism) for their legitimacy.
- 10 H. Frankfurt, 'Moral Responsibility and Alternate Possibilities', *Journal of Philosophy*, Vol. 66, 1969, pp. 829–39.
- 11 Of course, this is not the end of the story. It can be questioned whether the compatibilist conception of free will really is a conception of free will. The compatibilist concedes that my present beliefs and desires were caused (determined) by previous events, which in turn were caused by earlier events, and so on, stretching back to events before I was born. How then can I think of 'my' beliefs and desires as my own?
- 12 If free will is impossible, it is not compatible with anything (*a fortiori*, it is not compatible with determinism).
- 13 G. Strawson, *Freedom and Belief* (Oxford: Oxford University Press) 1986. See also his 'The Impossibility of Moral Responsibility', *Philosophical Studies*, Vol. 75, 1994, pp. 5–24.
- 14 J. G. Fichte, *The Vocation of Man*, ed. R. Chisholm (New York: The Liberal Arts Press) 1956, p. 27.
- 15 Notice that the truth or falsity of determinism is irrelevant to this argument.

8 PERSONAL IDENTITY

- 1 See J. Butler, 'Of Personal Identity', reprinted in J. Perry (ed.), *Personal Identity* (Los Angeles, Calif.: University of California Press) 1975; T. Reid, *Essays on the Intellectual Powers of Man*, ed. A. Woozley (London: Macmillan) 1941; and R. Swinburne, 'Personal Identity: The Dualist Theory', in S. Shoemaker and R. Swinburne, *Personal Identity* (Oxford: Basil Blackwell) 1984.
- 2 See D. Parfit, 'Personal Identity', reprinted in J. Glover (ed.), *The Philosophy of Mind* (Oxford: Oxford University Press), 1976, pp. 142–63; and *Reasons and Persons* (Oxford: Oxford University Press) Part III.
- 3 J. Locke, *Essay Concerning Human Understanding*, ed. W. Carroll (Bristol: Thoemmes) 1990, II, xxvii, 9.

NOTES

- 4 See, e.g., R. Descartes, *Meditations on First Philosophy*, ed. J. Cottingham (Cambridge: Cambridge University Press) 1996.
- 5 David Hume, *A Treatise of Human Nature* (1739), Part IV, Section IV: Of Personal Identity. Reprinted in A. J. Ayer and R. Winch (eds) *British Empirical Philosophers* (London: Routledge & Kegan Paul) 1965, p. 490.
- 6 Roughly, F is a substance sortal if it tells us what something fundamentally or essentially is. If x falls under substance sortal F, then x is fundamentally F. In contrast, G is a phase sortal if x falls, or might fall, under G for part but not all of its existence. Thus, 'teacher', 'philosopher', and 'unicyclist' are uncontroversial examples of phase sortals. Relatively uncontroversial examples of substance sortals are 'desk', 'car', 'ship', 'tree', 'dog', etc. See D. Wiggins, *Sameness and Substance* (Oxford: Basil Blackwell) 1980, pp. 62–6.
- 7 For a defence of this view, see P. Snowdon, 'Persons, Animals and Ourselves', in *The Person and the Human Mind*, ed. C. Gill (Oxford: Clarendon Press) 1990, and E. Olson, *The Human Animal* (Oxford: Oxford University Press) 1997.
- 8 That is, e.g., X at t1 is the same brain as Y at t2 if and only if a continuous path can be traced through space and time, from X at t1 to Y at t2.
- 9 That is, if A is psychologically continuous with B, and B is psychologically continuous with C, then A is psychologically continuous with C. But A may be strongly psychologically connected with B, and B with C, yet A not be strongly psychologically connected with C (e.g., memories fade over time).
- 10 S. Shoemaker, *Self-Knowledge and Self-Identity* (Ithaca, NY: Cornell University Press) 1963, pp. 23–4.
- 11 Most of us do not have equipollent or functionally equivalent hemispheres – but we might have had. The brain and psychological criteria are intended to apply to all persons, including those with equipollent hemispheres.
- 12 The transitivity of identity states that if $X = Y$ and $Y = Z$ then $X = Z$.
- 13 See D. Lewis, 'Survival and Identity', reprinted in *The Identities of Persons*, ed. A. O. Rorty (Los Angeles, Calif.: University of California Press) 1969, pp. 17–41; and H. Noonan, *Personal Identity* (London: Routledge) 2003, pp. 139–43.
- 14 Parfit, 'Personal Identity', op. cit., p. 149.
- 15 There are other objections to (all versions of) the brain criterion. As stated above, the concept of a person is the concept of a relatively sophisticated kind of mental being. Hence, the prospects do not look bright for any theory that understands personal identity over time in terms of the continued existence of an entity which is not essentially mental (such as a body, brain, or human being). Second, the only reason the brain is singled out as essential to personal identity is because it supports psychological continuity. In which case, its importance is entirely derivative and cannot be the essence of personal identity.
- 16 Parfit, 'Personal Identity', op. cit., p. 149.
- 17 Ibid., p. 144.
- 18 Parfit takes this combination of views to have significant implications for morality and rationality. For example, he thinks that if identity is not what matters, then the fact of the 'separateness of persons' is less deep. This, in turn, is taken to imply that less weight should be assigned to distributive principles. We should aim to maximize the net sum of benefits over burdens, whatever their distribution. Again, if psychological connectedness is part of what matters, a person may deserve less

- punishment the less he is psychologically connected to his earlier criminal self (e.g., if he has truly reformed). In the case of rationality, Parfit thinks that if identity is not what matters, then the self-interest principle (according to which it is 'especially rational to act in our own best interests') 'has no force' ('Personal Identity', op. cit., p. 161).
- 19 T. Reid, *Essays on the Intellectual Powers of Man*, ed. A. Woozley (London: Macmillan) 1941, quoted in Noonan, op. cit., p. 16.
 - 20 See also his discussion with Godfrey Vesey, 'Brain Transplants and Personal Identity', in *Philosophy in the Open* (Milton Keynes: Open University Press) 1978.
 - 21 Parfit, 'Personal Identity', op. cit., p. 146. His example, in the 1971 article, of a case where a question of personal identity does not receive a 'yes or no' answer is that of fission. But, as we have seen, this may not have been the best example for that purpose. Indeed, fission plays a much more important role in undermining the belief about the importance of personal identity.
 - 22 Note that the indeterminacy is in the truth value of sentences or statements of personal identity. I take Gareth Evans ('Can There Be Vague Objects?' *Analysis*, Vol. 38, 1978, p. 208), to have shown that the identity relation itself is never a source of vagueness or indeterminacy. Any vagueness in an identity sentence is due to vagueness in one or both of its singular terms (i.e., it is vague which object the term picks out).
 - 23 B. Williams, 'The Self and the Future' in his collection *Problems of the Self* (Cambridge: Cambridge University Press) 1982, pp. 46–64, especially pp. 58–63.
 - 24 *Ibid.*, p. 58.
 - 25 *Ibid.*, p. 61.
 - 26 The phrase 'conceptual shadow' is due to Williams, *ibid.*, p. 60.
 - 27 *Ibid.*, p. 63.

9 REALISM AND ANTI-REALISM

- 1 Condition (ii) allows us to be realists about trains, planes and automobiles. Although they would not have existed, had we not existed, were we to cease to exist, they would still exist. Condition (iii) allows us to be realists about mental states. Although it is true that, had I not existed, my mental states would not have existed, and true that were I to cease to exist, my mental states would cease to exist, the nature of my mental states is not fixed by whatever I take their nature to be. Since we may be mistaken about the nature of our mental states, realism about mental states is still an option.
- 2 According to an error theory about Fs, there are no Fs and, hence, statements about Fs are false (in error). According to expressivism about Fs, declarative sentences containing the term 'F' are not used to make statements, but have some other function (e.g., expressing attitudes or emotions). The expressivist typically agrees with the error theorist that there are no Fs.
- 3 M. A. E. Dummett, 'Realism' and 'The Reality of the Past', in his collection *Truth and Other Enigmas* (Cambridge, Mass.: Harvard University Press) 1978, pp. 145–65 and 358–74.
- 4 'Realism', op. cit., p. 146.
- 5 *Ibid.*, p. 146.
- 6 *Ibid.*, p. 147.

NOTES

- 7 Ibid., p. 155. The law of excluded middle states that, for all P, either P or not-P. This law is distinct from, though related to, the principle of bivalence (the principle that all statements in a given area are true or false). It is the latter which is the anti-realist's primary target.
- 8 Ibid., p. 146.
- 9 Although Dummett's aim is to resolve issues in metaphysics, the anti-realist typically counts a class of statements as 'disputed' on epistemic grounds. It is because our epistemic access to, e.g., material objects, other minds, and the past, is supposedly 'indirect' that statements about such domains are held to be problematic (in contrast, e.g., to statements about sense data, behaviour and present traces). In addition, Dummett assumes that statements about sense experience can be understood without reference to material objects, and that statements about human behaviour can be understood without reference to mental states. Both assumptions are controversial, and, arguably, false. See J. Cook, 'Human Beings', in P. Winch (ed.), *Studies in the Philosophy of Wittgenstein* (New York: Routledge) 1969, pp. 117–51; and J. McDowell, 'Criteria, Defeasibility and Knowledge', in J. Dancy (ed.), *Perceptual Knowledge* (Oxford: Oxford University Press) 1988, pp. 209–19.
- 10 'Realism', op. cit., p. 148.
- 11 Ibid., p. 148.
- 12 Ibid., p. 150.
- 13 Ibid., p. 150.
- 14 'The Reality of the Past' op. cit., p. 363.
- 15 'Realism', op. cit., p. 162.
- 16 'The Reality of the Past', op. cit., p. 362.
- 17 Ibid., p. 362.
- 18 In criticizing the case for anti-realism, I am not presupposing the principle of bivalence. Many sentences (e.g., vague or ambiguous ones) are neither true nor false. What is at issue is the anti-realist case for rejecting bivalence.
- 19 See A. J. Ayer, *Language, Truth and Logic* (Harmondsworth: Penguin) 1976.
- 20 There is of course much more to be said. I can do no better than to direct the interested reader to three papers by Alex Miller, which reveal in detail the flaws and fallacies in the case for anti-realism. See his 'The Significance of Semantic Realism', *Synthese*, Vol. 136, 2003, pp. 191–217; 'What is the Manifestation Argument?', *Pacific Philosophical Quarterly*, Vol. 83, 2002, pp. 352–83; and 'What is the Acquisition Argument?' in A. Barber (ed.), *Epistemology of Language* (Oxford: Oxford University Press) 2003, pp. 459–94.
- 21 F. Fitch, 'A Logical Analysis of Some Value Concepts', *The Journal of Symbolic Logic*, Vol. 28, 1963, pp. 135–42.
- 22 A state Φ is factive if and only if ' Φ ' implies 'p'.
- 23 See, e.g., D. Edgington, 'The Paradox of Knowability', *Mind*, Vol. 94, 1985, pp. 557–68.
- 24 For example, could God not have created us with incredibly powerful minds, capable of grasping every truth about the universe, consistently with our freely choosing not to investigate every issue? This apparent possibility is ruled out by Fitch's proof.
- 25 For example, Crispin Wright has recently explored new ways of opposing the realist's assumption that some subject matter is constituted independently of us. See, e.g., his *Truth and Objectivity* (Cambridge, Mass.: Harvard University Press) 1992.

INDEX

- a priori* 2, 15, 60, 82, 103, 112, 148
A theory 72–82, 87–90, 94, 99
abstract 38–9, 46, 55
analyticity 8–9
animalism 124, 127
Anscombe, E. 62–3, 66
anti-realism 138–9; case studies 142–6;
definition 140–2; motivation 146–8
Aquinas, St Thomas 4, 7–8, 39
argument from design *see* design
argument
Aristotle 8, 10, 37–9, 41, 91–2, 108–9
asymmetric relations 56
Ayer, A.J. 31–2, 110, 147–8
- B theory 72–6, 79–82, 88–90, 94, 99
backwards causation 56, 60, 62, 82,
88, 96
behaviourism 144
Berkeley, G. 7, 54, 139
Bradley, F.C. 72
Brentano, F. 28
Broad, D.C. 81, 99
bundle theory 42–3, 48
Butler, J. 122–3, 132
- Cahn, S. 110–11
Campbell, K. 46
causation 53–61, 82, 93; counterfactual
theory 64–5; free will 112, 114–15;
loops 56, 88, 96–7; necessity 60–1,
66; observability 62–3; regularity
theory 61–4; simple theory 65–7;
simultaneous 65; singular 54–5
cause, definition 61
change 74, 77, 79, 81, 88; definition
75–6, 91; time 90–4, 97–9
Clarke, S. 123
class nominalism 37, 44–5
compatibilism 105, 111, 114–15
composition fallacy 10
compresence 43
concept nominalism 44
conditionals 58, 64, 94, 108–9, 142–3
contiguity 59–60
contingency 2, 7–11
Copleston, F. 7–10
cosmological argument 2, 7–11
cosmology 10, 39, 92
counterfactual conditionals 25, 58, 61,
64–6, 94, 109, 142–4
counterpart theory 26
creation 2–3, 11–12, 24
criteria of identity 122
- Darwin, C. 12–13
Davidson, D. 55
Descartes, R. 4, 124–5
descriptions, theory of 29–31, 33
design argument 3, 11–12
determinism 15, 103, 105, 111–16
dualism 113, 125

INDEX

- Ducasse, C. 62
Dummett, M.A.E. 78–80, 82, 107,
139–40, 142–3, 145–7, 150
- empiricism 2–3, 7, 54, 58, 82, 103, 112,
147–8
empty names 31–3, 40
error theories 138
events 55–65, 72–3, 75–7; time 81–2, 94
evil 2–3, 13–16
evolutionary theory 12–13, 24
excluded middle 141
existence 21–2, 31–3; modal realism
25–7; non-existent objects 27–31;
world 23–5
external relations 38
- facts 55, 76, 79–81, 89–90
fatalism 15, 103–10, 116
Fichte, J.G. 115
fission 127–30, 134
Fitch, F.B. 149–50
forms, theory of 38–9, 41
Frankfurt, H. 114–15
free will 15, 103, 105, 110–16; coherence
of 115–16; determinism 111–12;
fatalism 104–10; incompatibilism
112–15; theology 110–11
Frege, G. 78
future 72–5, 77–82, 88–90; anti-realism
144; free will 104–7, 108, 110; time
travel 95–7, 99
- Gaunilo 6–7
Geach, P.T. 63
general terms 40
generalization 57, 62, 67, 90, 106
God 2–3, 24–5, 97, 103–4; argument
from evil 13–16; cosmological
argument 7–11; free will 110–12;
ontological argument 4–7;
teleological argument 11–13
Goldbach's conjecture 146–7
grandfather paradox 97–9
- hard determinists 111
Hegel, G.W.F. 72
- Hobbes, T. 111–12
Horwich, P. 78–80
Hume, D. 7, 11–12, 53–4, 56; causation
58–62, 64–6; free will 111–12, 115;
personal identity 124–5; time 91
- idealism 29, 72, 139–40
identity 30, 95, 121, 131–2; criteria 122
impressions 59
incompatibilism 111–16
indeterminism 62, 103, 105, 133
indexicals 82
indicative conditionals 109
individual substances 39, 42
infinite regress 41–2, 47, 116
instantiation 39, 41–2, 45, 47–8
intention 131, 142
intentionality 28
internal relations 38
- Kant, I. 4, 16
knowledge 66, 81–2, 110–11, 134,
149–50
Kreisel, G. 140
Kripke, S. 9, 26, 31
- laws of nature 54, 57, 67, 94, 105,
112–13
Leibniz's Law 121
Leverrier, J. 32
Lewis, D. 22–4, 27, 33; causation 56–7,
64; free will 106; personal identity
128; time 87, 94–8
libertarianism 105, 111, 113
Locke, J. 7, 54, 123–4, 148
logical positivism 147–8
- McGinn, C. 27, 31
McTaggart, J.M.E. 29, 71–5, 81–2;
change 91, 93; paradox 76–9
many-worlds hypothesis 25
material conditionals 109
meaning 21, 28, 31, 33; anti-realism 139,
141, 146–8, 150; universals 39–40
Meinong, A. 27–8, 30–1
Mellor, H. 55, 89–90
memory 125–6, 131–3, 142

- Mill, J.S. 56
 modal realism 22–7, 33
 Moore, G.E. 29
 moral responsibility 15, 103, 105,
 112–16, 123
 multiple-existence theory 26
 multiple-occupancy response 128, 134

 natural selection 13
 nature *see* laws of nature
 necessary being 2, 8–9
 necessary evils 14
 necessity 58–61, 65–7, ; free will 110,
 112; non-analytic 9
 negative existentials 32
 Newton, I. 12, 112
 nominalism 37, 43–8, 140
 non-analytic necessities 9
 non-branching continuity 129, 134
 non-existent objects 21, 27–33
 non-reductive analysis 122
 non-symmetric relations 56
 Noonan, H. 128
 numerical identity 40–1, 121

 objects 37, 39, 43, 46–8; causation 55,
 58, 61, 64; change 75
 observer-independent description 79–80
 omnipotence 2, 13–14, 16
 omniscience 2, 13, 16, 110, 149
 ontological argument 2, 4–9
 overdetermination 65

 Paley, W. 11–13
 Parfit, D. 24–5, 90, 122–5, 127, 129–33
 parody arguments 7
 part-whole relation 47–8
 particulars 37; bundle theory 42–3;
 nominalism 43–8; property realism
 38–42
 past 72–5, 77–82, 88–90; anti-realism
 139–40, 142, 144–6; free will 104–8,
 110–11; time travel 95–9
 persistence 122, 125–8
 persons ; complex view 125–31; definition
 123–5; identity of 121–3; persistence
 125–8; simple view 132–4

 phase sortals 124
 Plato 37–41
 Platonism 140
 possibility principle 149–50
 possible worlds 22–3, 25–7, 94, 111,
 126
 pre-emption cases 65
 predicate nominalism 37, 44
 predication 28–31, 48, 142
 present 75, 77–80, 88–90; presentism 81,
 99
 Prior, A.N. 87–9, 94, 99
 private language 78
 projectivism 59, 61
 properties 37, 39–48, 75–6, 80
 property realism 31–3, 37–42, 46–8
 propositions 75–6, 89, 104, 148

 qualitative identity 41, 134
 quantifiers 26, 30–3
 quantum theory 22, 62, 112
 quasi-memory 132

 rationalism 7
 realism 94, 138–50; case studies 142–6;
 definition 140; dispute 140–2
 reductionism 124–5, 131, 141–2
 reductive analysis 122, 125–6
 reference 21, 28–33, 40; frames of 82;
 theory 39–40
 regress *see* infinite regress
 regularity theory 58–9, 61–6, 112, 115
 Reid, T. 62, 122, 132
 reification 6
 relations 37–8, 46–7, 58; asymmetric 56;
 identity 129; reflexive 56; symmetric
 56; time 74, 80, 89, 94–5; transitive
 56
 resemblance nominalism 37, 45
 Rousseau, J.-J. 54
 Russell, B. 7–10, 28–31, 33; existence 33;
 paradox 29; time 72, 74–6;
 universals 37, 42–3, 45, 48

 St Anselm 4–8
 self-concern 131–2
 self-consciousness 123–4

INDEX

- self-determination 115–16
- self-interest theory 123
- self-predication 41
- self-sufficiency 9–10
- Shoemaker, S. 74, 87, 91–4, 99, 126
- singular terms 29–30
- singularism 62–4
- Smith, A. 54
- Socrates 38–41
- something/nothing 23–4, 27
- souls 38, 72, 124, 134
- space *see* spatio-temporality
- special-relativity theory 82
- Stoics 11
- Stout, G.F. 46
- Strawson, G. 59, 103, 111, 115–16
- Strawson, P.F. 115
- subject-predicate 42–3
- substance 37, 42–3, 47; sortals 124
- Swinburne, R. 122

- teleological argument 2–3, 11–13
- tensed/tenseless facts 88–90

- theological fatalism 103–4, 110–11
- theology 4, 11, 39, 123
- third man argument 41
- time: A/B theories 72–81; travel 87–8, 94–9; without change 90–4
- token-reflexive expressions 80
- transitive relations 56
- trope theory 37, 46–8
- truth-value links 72–3, 145

- universalism 62–3
- universals 37–42, 140; bundle theory 42–3; nominalism 43–8; property realism 38–42
- universe 2, 8, 10–12; existence 21, 24
- unknowable truths 149

- watchmaker argument 11–12
- Wells, H.G. 88
- Wiggins, D. 124
- Williams, B. 129, 133–4
- Williams, D. 37, 46
- Wittgenstein, L. 29, 63, 78