

METAPHYSICS:

The Big Questions

EDITED BY PETER VAN INWAGEN AND DEAN W. ZIMMERMAN



Philosophy: The Big Questions

Series Editor: James P. Sterba, University of Notre Dame, Indiana

Designed to elicit a philosophical response in the mind of the student, this distinctive series of anthologies provides essential classical and contemporary readings that serve to make the central questions of philosophy come alive for today's students. It presents complete coverage of the Anglo-American tradition of philosophy, as well as the kinds of questions and challenges that it confronts today, both from other cultural traditions and from theoretical movements such as feminism and postmodernism.

Aesthetics: the Big Questions

Edited by Carolyn Korsmeyer

Epistemology: the Big Questions

Edited by Linda Martin Alcoff

Ethics: the Big Questions

Edited by James P. Sterba

Metaphysics: the Big Questions

Edited by Peter van Inwagen and Dean W. Zimmerman

Philosophy of Language: the Big Questions

Edited by Andrea Nye

Philosophy of Religion: the Big Questions

Edited by Eleonore Stump and Michael J. Murray

Race, Class, Gender, and Sexuality: the Big Questions

Edited by Naomi Zack, Laurie Shrage, and Crispin Sartwell

Copyright © Blackwell Publishers Ltd., 1998

First published 1998

2 4 6 8 10 9 7 5 3 1

Blackwell Publishers Inc.
350 Main Street
Malden, Massachusetts 02148
USA

Blackwell Publishers Ltd
108 Cowley Road
Oxford OX4 1JF
UK

All rights reserved. Except for the quotation of short passages for the purposes of criticism and review, no part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the publisher.

Except in the United States of America, this book is sold subject to the condition that it shall not, by way of trade or otherwise, be lent, resold, hired out, or otherwise circulated without the publisher's prior consent in any form of binding or cover other than that in which it is published and without a similar condition including this condition being imposed on the subsequent purchaser.

Library of Congress Cataloguing-in-Publication Data

Metaphysics : the big questions / edited by Peter van Inwagen and Dean W. Zimmerman.

p. cm. — (Philosophy, the big questions ; 4)

Includes bibliographical references and index.

ISBN 0-631-20587-X (hardback). — ISBN 0-631-20588-8 (pbk.)

I. Metaphysics. I. van Inwagen, Peter. II. Zimmerman, Dean W. III. Series

BD111.M575 1998

110—dc21

98-11440

CIP

British Library Cataloguing in Publication Data

A CIP catalogue record for this book is available from the British Library.

Typeset in 10½ on 12½ pt Galliard
by Ace Filmsetting Ltd, Frome, Somerset
Printed in Great Britain by T.J. International, Padstow, Cornwall

This book is printed on acid-free paper

For Roderick M. Chisholm

CONTENTS

Preface	xi
Introduction: What is Metaphysics?	1
PART ONE WHAT ARE THE MOST GENERAL FEATURES OF THE WORLD? 15	
Introduction	17
What is the Relationship between an Individual and its Characteristics?	23
1 Universals and Resemblances: Chapter 1 of <i>Thinking and Experience</i> H. H. PRICE	23
2 The Elements of Being D. C. WILLIAMS	40
3 The Principle of Individuation: an Excerpt from <i>Human Knowledge, its Scope and Limits</i> BERTRAND RUSSELL	52
4 Distinct Indiscernibles and the Bundle Theory DEAN W. ZIMMERMAN	58
What is Time? What is Space? 67	67
5 Time: an Excerpt from <i>The Nature of Existence</i> J. McT. E. McTAGGART	67
6 McTaggart's Arguments against the Reality of Time: an Excerpt from <i>Examination of McTaggart's Philosophy</i> C. D. BROAD	74
7 The Notion of the Present A. N. PRIOR	80
8 The General Problem of Time and Change: an Excerpt from <i>Scientific Thought</i> C. D. BROAD	82
9 The Space-Time World: an Excerpt from <i>Philosophy and Scientific Realism</i> J. J. C. SMART	94

CONTENTS

10	Topis, Soris, Noris: an Excerpt from <i>The Existence of Space and Time</i> IAN HINCKFUSS	101
11	Some Free Thinking about Time A. N. PRIOR	104
12	The Fourth Dimension: an Excerpt from <i>The Ambidextrous Universe</i> MARTIN GARDNER	108
13	Incongruent Counterparts and Higher Dimensions JAMES VAN CLEVE	111
14	Achilles and the Tortoise MAX BLACK	120
15	A Contemporary Look at Zeno's Paradoxes: an Excerpt from <i>Space, Time, and Motion</i> WESLEY C. SALMON	129
16	Grasping the Infinite JOSÉ A. BENARDETE	149
17	The Paradoxes of Time Travel DAVID LEWIS	159
	How do Things Persist through Changes of Parts and Properties?	171
18	Of Confused Subjects which are Equivalent to Two Subjects: an Excerpt from <i>The Port-Royal Logic</i> ANTOINE ARNAULD AND PIERRE NICOLE	171
19	Identity through Time RODERICK M. CHISHOLM	173
20	Identity, Ostension, and Hypostasis W. V. O. QUINE	186
21	Identity: an Excerpt from <i>Quiddities</i> W. V. O. QUINE	188
22	In Defense of Stages: Postscript B to "Survival and Identity" DAVID LEWIS	190
23	Some Problems about Time PETER GEACH	192
24	The Problem of Temporary Intrinsics: an Excerpt from <i>On the Plurality of Worlds</i> DAVID LEWIS	204
25	Temporary Intrinsics and Presentism DEAN W. ZIMMERMAN	206
	How do Causes Bring about their Effects?	221
26	Constant Conjunction: an Excerpt from <i>A Treatise of Human Nature</i> DAVID HUME	221
27	Efficient Cause and Active Power: an Excerpt from <i>Essays on the Active Powers of the Human Mind</i> THOMAS REID	226

28	Psychological and Physical Causal Laws: an Excerpt from <i>The Analysis of Mind</i> BERTRAND RUSSELL	227
29	Causality: an Excerpt from <i>A Modern Introduction to Logic</i> L. SUSAN STEBBING	229
30	Causality and Determination G. E. M. ANSCOMBE	244
PART TWO WHAT IS OUR PLACE IN THE WORLD?		259
	Introduction	261
	How is the Appearance of a Thing Related to the Thing that Appears?	267
31	The Theory of Sensa: an Excerpt from <i>Scientific Thought</i> C. D. BROAD	267
32	Qualities: an Excerpt from <i>Consciousness and Causality</i> D. M. ARMSTRONG	272
33	The Status of Appearances: an Excerpt from <i>Theory of Knowledge</i> (1st edition) RODERICK M. CHISHOLM	281
	What is the Relation between Mind and Body?	291
34	Which Physical Thing Am I? An Excerpt from “Is There a Mind–Body Problem?” RODERICK M. CHISHOLM	291
35	Personal Identity: a Materialist Account SYDNEY SHOEMAKER	296
36	Divided Minds and the Nature of Persons DEREK PARFIT	310
37	Personal Identity: the Dualist Theory RICHARD SWINBURNE	317
38	The Puzzle of Conscious Experience DAVID J. CHALMERS	333
	Is it Possible for Us to Act Freely?	343
39	Free Will as Involving Determination and Inconceivable without It R. E. HOBART	343
40	Human Freedom and the Self RODERICK M. CHISHOLM	356
41	The Mystery of Metaphysical Freedom PETER VAN INWAGEN	365
42	The Agent as Cause TIMOTHY O’CONNOR	374

PART THREE IS THERE JUST ONE WORLD?	381
Introduction	383
43 Speaking of Objects	385
W. V. O. QUINE	
44 After Metaphysics, What?	388
HILARY PUTNAM	
45 Truth and Convention	392
HILARY PUTNAM	
46 Nonabsolute Existence and Conceptual Relativity: an Excerpt from “Putnam’s Pragmatic Realism”	399
ERNEST SOSA	
47 Addendum to “Nonabsolute Existence and Conceptual Relativity”: Objections and Replies	407
ERNEST SOSA	
PART FOUR WHY IS THERE A WORLD?	411
Introduction	413
Is There an Answer?	415
48 The Problem of Being: Chapter 3 of <i>Some Problems of Philosophy</i>	415
WILLIAM JAMES	
49 The Puzzle of Reality: Why does the Universe Exist?	418
DEREK PARFIT	
50 Response to Derek Parfit	427
RICHARD SWINBURNE	
Does the Answer Involve a Necessary Being?	431
51 The Cosmological Argument and the Principle of Sufficient Reason	431
WILLIAM L. ROWE	
52 The Ontological Argument: Chapters II–IV of the <i>Proslogion</i>	441
ST ANSELM	
53 Anselm’s Ontological Arguments	443
NORMAN MALCOLM	
PART FIVE IS METAPHYSICS POSSIBLE?	455
Introduction	457
54 The Rejection of Metaphysics: Chapter 1 of <i>Philosophy and Logical Syntax</i>	459
RUDOLF CARNAP	
55 Postmodernism, Feminism, and Metaphysics: an Excerpt from <i>Thinking Fragments</i>	469
JANE FLAX	
56 Metaphysics and Feminist Theory: Excerpts from “Feminist Metaphysics” and “Anti-Essentialism in Feminist Theory”	480
CHARLOTTE WITT	
Index	492

PREFACE

The problems of metaphysics are many. Some arise upon the least reflection about the world and our place in it. Others are less obvious, appearing as problems only to those willing to think very hard about highly abstract questions. The reader of this anthology will find philosophers grappling with metaphysical problems of both sorts – although we have deliberately decided to favor the less abstract, more immediately accessible problems, since this anthology is intended as an introduction to the subject. The essays and excerpts are largely free of unexplicated technical terminology and symbolism. And the topics covered complement those of a number of popular single-author introductions to metaphysics.

With the exception of the final group of essays, all the readings are made to fall under a series of questions about “the world.” We assume that the world includes everything that there is – that is, all that exists. The first and largest part, “What are the most general features of the world?,” includes readings on the problem of universals, the nature of particular things and the manner of their persistence through time, rival theories of the passage of time, absolute space and incongruent counterparts, causation, and a budget of paradoxes: McTaggart’s paradox, paradoxes of motion, of the infinite, of time travel, and of intrinsic change. The second, and second largest, part asks, “What is our place in the world?” Here are questions about the relation between the way things appear to us and the way they are (sense data, secondary qualities), personal identity (two forms of materialism, a version of Cartesian dualism, and Derek Parfit’s “Buddhism”), the nature of phenomenal experience, and free will. Part Three raises the question of “anti-realism”: Is there just one world, one complete inventory of what there is? Or does what there is vary from community to community or person to person? Part Four begins with reflection on whether there could be an answer to the question, “Why is there a world?” – that is, why is there something, rather than nothing? The part ends with two attempts to answer the question by appeal to a necessary being (the Deity of the cosmological and ontological arguments). The final part includes challenges to the very possibility of metaphysics from both positivist and postmodern perspectives.

Although most of the readings have appeared elsewhere, a few have been

PREFACE

written especially for this volume: Timothy O'Connor, "The Agent as Cause"; Ernest Sosa, "Addendum to 'Nonabsolute Existence and Conceptual Relativity': Objections and Replies"; Richard Swinburne, "Response to Derek Parfit"; James Van Cleve, "Incongruent Counterparts and Higher Dimensions"; Peter van Inwagen, "The Mystery of Metaphysical Freedom"; and Dean Zimmerman, "Temporary Intrinsics and Presentism." José Benardete's contribution, "Grasping the Infinite," includes a parable taken from his book *Infinity* (Oxford: Clarendon Press, 1964), but is otherwise new. And Zimmerman's "Distinct Indiscernibles and the Bundle Theory" is a considerably expanded version of a dialogue which originally appeared in *Mind*. The general introduction, "What Is Metaphysics?", is a substantive essay based largely on Peter van Inwagen's contribution to *Contemporary Metaphysics: a Reader*, edited by Stephen Laurence and Cynthia Macdonald (Oxford: Blackwell, 1998).

The introductions to the sections serve three purposes: (i) to indicate how the readings in the section are related to one another; (ii) to point out connections between these selections and readings in other parts of the anthology; and (iii) to suggest supplementary readings.

An undergraduate course in metaphysics could profitably use this text alongside a wide variety of books, including classics such as Descartes's *Meditations on First Philosophy*, Berkeley's *Three Dialogues between Hylas and Philonous*, and Bertrand Russell's *The Problems of Philosophy*. Here are some more recent introductory books that take up many of the questions addressed in this anthology and are, in general, appropriate for an undergraduate audience:

William R. Carter, *The Elements of Metaphysics* (Philadelphia, Penn.: Temple University Press, 1990).

Martin Gardner, *The Whys of a Philosophical Scrivener* (New York: Quill, 1983).

Although not strictly an introduction to philosophy or metaphysics, Gardner's "confessional" can be a useful text in introductory philosophy courses.

D. W. Hamlyn, *Metaphysics* (Cambridge, UK: Cambridge University Press, 1984).

William Hasker, *Metaphysics: Constructing a World View* (Downers Grove, Ill., and Leicester, UK: InterVarsity Press, 1983).

John F. Post, *Metaphysics: a Contemporary Introduction* (New York: Paragon House, 1991).

Quentin Smith and L. Nathan Oaklander, *Time, Change and Freedom: an Introduction to Metaphysics* (London: Routledge, 1995).

Richard Taylor, *Metaphysics*, 4th edn (Englewood Cliffs, N.J.: Prentice-Hall, 1992).

Peter van Inwagen, *Metaphysics* (Boulder, Col.: Westview Press, 1993).

The following anthologies and single-author texts may serve as companions to the volume for more advanced students (e.g., upper-level philosophy majors, or beginning graduate students):

- D. M. Armstrong, *Universals: an Opinionated Introduction* (Boulder, Col.: Westview Press, 1989).
- Bruce Aune, *Metaphysics: the Elements* (Minneapolis: University of Minnesota, 1985).
- José Benardete, *Metaphysics: the Logical Approach* (Oxford: Oxford University Press, 1989).
- Steven D. Hales (ed.), *Metaphysics: Contemporary Readings* (Belmont, Cal.: Wadsworth, 1998).
- Michael Jubien, *Contemporary Metaphysics* (Oxford: Blackwell, 1997).
- Stephen Laurence and Cynthia Macdonald (eds), *Contemporary Metaphysics: a Reader* (Oxford: Blackwell, 1998).
- Michael Loux, *Metaphysics: a Contemporary Introduction* (London: Routledge, 1997).

Our lists of supplementary readings include suggestions for matching up chapters from all of these books with our selections. They also include many other books, articles, and (in a few cases) stories we think could be used to teach metaphysics to beginners.

We thank Marie Pannier for help with proofreading, the obtaining of copyright clearances, and other tasks. The dedication is offered in gratitude from two philosophers happy to have been on the “Chisholm Trail.”

PETER VAN INWAGEN
DEAN W. ZIMMERMAN

INTRODUCTION: WHAT IS METAPHYSICS? *

What is the Subject Matter of Metaphysics?

It is notoriously hard to provide a satisfactory account of what metaphysics is. Certain twentieth-century coinages like “metaphilosophy” and “metapsychology” encourage the impression that metaphysics is a study that somehow “goes beyond” physics. In reality, however, the Greek phrase “*ta meta ta phusica*,” from which our word “metaphysics” is derived, is the term that the early editors of Aristotle’s corpus used to refer to his book (from their point of view, his “books”) on what he called first philosophy. And this phrase means only “the ones [*sc.* books] that come after the ones about nature.” As is often the case, etymology is no guide to meaning.

The best approach to an understanding of what is meant by “metaphysics” is by way of the concepts of appearance and reality. It is a commonplace that the way things seem to be is often not the way they are, that the way things *apparently* are is often not the way they *really* are. The sun apparently moves across the sky – but not *really*. The moon seems larger when it is near the horizon – but its size never *really* changes. We might say that one is engaged in “metaphysics” if one is attempting to get behind all appearances and to describe things as they *really* are.

An example may be helpful. There are available two “interpretations” of quantum mechanics. One, the standard “Copenhagen” interpretation, implies that particles like electrons in passing from an emitter to a target do not in general follow “trajectories,” continuous paths through space. The other interpretation, the work of David Bohm, implies that particles do follow trajectories. It can be shown that these two interpretations are empirically equivalent: they

* Much of the “Introduction: What is Metaphysics?” originally appeared in Peter van Inwagen, “The Nature of Metaphysics: the State of the Art,” in Stephen Laurence and Cynthia Macdonald, eds, *Contemporary Metaphysics: a Reader* (Oxford: Blackwell, 1998). It also contains portions of Peter van Inwagen’s entry “Metaphysics,” in Adrian Hastings, ed., *The Oxford Companion to Christian Thought* (Oxford: Oxford University Press, forthcoming).

METAPHYSICS: THE BIG QUESTIONS

make the same predictions about the outcomes of all possible experiments. This fact has led many physicists and philosophers to say that the question as to which interpretation (if either) describes the way particles *really* behave is a metaphysical question. The two theories both “save the appearances” – or at least, if either fails to save the appearances the other will fail as well, and in the same way – but they are obviously not “the same.” Assuming that the concept “following a continuous path through space” is a coherent concept, particles must, as a matter of logic, either follow continuous paths through space – or not. Therefore, both interpretations cannot be correct descriptions of the real behavior of particles – although they could both be incorrect. It would be possible to argue that the question as to which (if either) is correct is impossible to answer and therefore idle. But that is not the same as saying it hasn’t got an answer. If it is a meaningful question it has an answer, and if “follows a continuous path through space” is a meaningful phrase, the question is meaningful. The anti-metaphysical attitude is typified by the scientist who is willing to make use of each of two logically inconsistent theories that make the same predictions about possible observations, and who deprecates as idle or meaningless the question of whether either theory describes things as they really are.

Let us try to describe the metaphysical impulse in a little more detail. If one is attempting to “get behind all appearances and describe things as they really are,” if one is “engaging in metaphysics,” then one is attempting to determine certain things with respect to certain statements (or assertions or propositions or theses), those statements that, if true, would be descriptions of the reality that lies behind all appearances, descriptions of things as they really are. (Primarily to determine which of them are true and which of them are false; but also, perhaps, to determine various other things about them, such as which ones it is reasonable to believe, and which ones are logically consistent with one another.)

Let us call this “reality that lies behind all appearances” simply Reality (with a capital). And let us call statements that, if true, would be descriptions of this Reality *metaphysical statements*. Which statements are, if true, descriptions of Reality? This is a difficult question, because simply from examining a speaker’s words it cannot be determined whether the speaker has made a metaphysical statement: one must also examine the context in which those words were spoken. It is necessary to do this because different “restrictions of intended reference” may be in force in different contexts, and this has the consequence that the same words can express different things in different contexts. If, for example, you look into your refrigerator and say sadly “There’s no beer,” you are not asserting that the existence of that beverage is a myth or an illusion. And this is because, in the context in which you are speaking, you and your audience know that your statement is intended to describe only the state of things inside your refrigerator.

Let us see how this point might apply in a case in which we want to know whether a speaker has made a metaphysical statement. The late Carl Sagan, in his television series *Cosmos*, made the following much-quoted statement: “The cosmos [i.e., the physical universe] is all there is or was or ever will be.” Was this a metaphysical statement? Well, that depends. In the context in which Sagan

made his statement, were there any restrictions of intended reference in force? Did Sagan, perhaps, intend his statement to apply only to physical things? Was he perhaps saying only that the cosmos was the totality of physical things, past, present, and future? In that case, his statement was not a metaphysical statement, but simply an explanation of the meaning of the word "cosmos." Or did Sagan perhaps make his statement in a context in which there were *no* restrictions of intended reference in force? Did he mean to say that *everything* – everything without qualification – was a part of the cosmos? (Notice that, on the "unrestricted" interpretation of his statement, the statement implies that there is no God – or anything else that is not physical; on the "restricted" interpretation, the statement has no such implication.) In that case, his statement can plausibly be described as a metaphysical statement – for if we learned that there was no God or anything else non-physical, would we not learn something about the reality that lay behind all appearances?

One requirement on a metaphysical statement, then, is that it be made with no restrictions of intended reference in force. A second, and closely related requirement is that the statement represent a serious attempt by the speaker to state the strict and literal truth. We often express ourselves carelessly or loosely or metaphorically. (Restriction of intended reference might be seen as a special case of not speaking "strictly.") We say things like "The sun is trying to come out," "The car doesn't want to start," "Time passes slowly when one is bored," and "Dark, angry clouds filled the sky." Since metaphysics is an attempt to get at how things really are, this requirement is not hard to understand. Those who say things like these do not mean to assert that the sun, the car, or the clouds are conscious beings, or that time can pass at different rates; and, therefore, at least some of the features of these statements do not represent an attempt to say how things really are. When doing metaphysics, such loose talk has to go: one must be willing to take responsibility for the strict and literal consequences of the words one has used to make a statement.

The metaphysician's aim, then, is to make assertions that strictly and literally describe reality and which can, with sufficient effort, be understood by anyone whose intellect is equal to the task. Metaphor may play a heuristic role in metaphysics – as in physics or economics or comparative linguistics – but metaphor must be banished from the metaphysician's "finished product." (The metaphysician may begin by calling space a receptacle or time the moving image of eternity, but at some point in the metaphysician's investigations these metaphors must be replaced with language that is meant to be taken literally. Or, at any rate, if a metaphysical work depends essentially on metaphor, it must be regarded as inherently incomplete, a sort of work in progress.) To say these things is not to say that the metaphysician (necessarily) regards the metaphorical as inferior to the literal; it is merely to demarcate what belongs properly to metaphysics.

May we then understand a metaphysical statement as one that (i) is made in a context in which no restrictions of intended reference are in force and (ii) is such that the speaker who makes it has made a serious effort to speak the strict and literal truth? This would not be satisfactory, for to call a statement "meta-

METAPHYSICS: THE BIG QUESTIONS

“physical” is to imply that it is a very general statement, and these two conditions include nothing that implies generality. An example may help us to understand the kind of generality that a statement must have, to be a metaphysical statement. Suppose Alice says, “All Greeks are mortal.” Let us assume that when she makes this statement, no restrictions of intended reference are in force: she means her statement to apply to *all* Greeks, and not only to the members of some “understood” special class of Greeks. And let us assume that this statement represents a serious effort on her part to speak the strict and literal truth. (Since the statement contains no figurative language and no “well, I hope you see what I mean” linguistic shortcuts, this assumption is reasonable enough.) Perhaps these two assumptions imply that her statement, if true, describes Reality, but it certainly does not describe very *much* of Reality. After all, it tells us only about Greeks; it tells us nothing about elephants or neutron stars or even non-Greek human beings. It is therefore not sufficiently general to count as a metaphysical statement.

What sort of statement would be “sufficiently general”? Might we say that to be sufficiently general to be a metaphysical statement, a statement must be about *everything*? This will not do, and for two reasons. First, *any* “all” statement is in one sense “about everything.” For example, the statement “All Greeks are mortal” is logically equivalent to “Everything is mortal if it is a Greek.” (Every elephant and every neutron star and every non-Greek immortal is mortal if it is Greek.) It is therefore not easy to say in any precise and useful way what it is for a statement to be “about everything.” Secondly, even if we ignore this difficulty and decide to rely on our intuitive sense of which statements are “about everything,” we shall run up against the fact that most philosophers would want to classify as “metaphysical” many statements that we would, speaking intuitively, say were not “about everything.” For example: Every event has a cause (this statement is, intuitively, not about everything, but only about events); Every physical thing is such that it might not have existed (. . . only about physical things); Any two objects that occupy space are spatially related to each other (. . . only about objects that occupy space). And there is a further problem: most philosophers would want to classify as metaphysical certain statements that are not “all” statements, not even ones that pertain to some special class of things like events or physical things: There is a God; Some things have no parts; There could be two things that had all the same properties.

Perhaps in the end, all we can say is this: some “categories” or “concepts” are sufficiently “general” that a statement will count as a “metaphysical statement” if – given that it is made in a context in which no restrictions of intended reference are in force, and given that the person who makes it is making a serious effort to say what is strictly and literally true – it employs only these categories. Among these categories are many that we have already used in our examples: “physical thing,” “spatial object,” “cause,” “event,” “part,” “property.” If we so define “metaphysical statement,” then the concept of a metaphysical statement will be open-ended and vague. It will be open-ended in that no final list of the categories that can occur in a “metaphysical statement” will be possible: we could try to make a complete list (we might go through all the historical texts

that were uncontroversially “metaphysical” and mark all the categories we came across that seemed to us to be “sufficiently general”), but, even if we had a list that satisfied us for the moment, we should have to admit that we might have to enlarge the list tomorrow. It is vague in that there will be borderline cases of “sufficiently general” categories: words such as “impenetrable,” “pain,” “straight,” and “surface” are possible examples of such borderline cases. But there will also be perfectly clear cases of categories that are not “sufficiently general”: “Greek,” “chair,” “elephant,” “neutron star,” “diminished-seventh chord,” “non-linear partial differential equation,” . . . (Words like “chair” and “elephant” can occur in a work on metaphysics, but only in examples meant to illustrate – or in counter examples meant to refute – theses whose statement requires only very general concepts.)

Where does this leave us? Let us suppose that Charles has made a certain statement. Let us suppose that when he made this statement, no restrictions of intended reference were in force. Let us suppose that Charles was willing to take responsibility for the strict and literal consequences of the words he has used to make the statement. And let us suppose that all the concepts or categories that Charles employed in making this statement were “sufficiently general.” (And let us suppose that his statement was not some logical truism like “Everything is either material or not material.”) Then, or so we would suggest, Charles has made a statement that, if true, describes Reality. That is, he has made a metaphysical statement. And if we try to decide whether his statement is true or false, reasonable or unreasonable, probable or improbable, consistent or inconsistent with various other metaphysical or non-metaphysical statements, then we are engaged in metaphysics.

Is Metaphysics Possible?

Is metaphysics, so conceived, possible? – that is, is it possible to “engage in metaphysics” in the above sense and to reach any interesting or important conclusions? Various philosophers have argued that metaphysics is impossible. The thesis that metaphysics is impossible comes in what might be called strong and weak forms. The strong form of the thesis is this: The goal is not there, since there is no Reality to be described; all the statements we have called metaphysical are false or meaningless. (And it is hard to see how all metaphysical statements could be simply false. If one metaphysician says that everything is material and another says that it is false that everything is material, then, if their statements are meaningful, one or the other of them must be true.) In this anthology, the strong form of the thesis finds expression in the selections by Rudolf Carnap, Hilary Putnam, and Jane Flax. The weak form of the thesis is this: The goal is there, but we human beings are unable to reach it, since the task of describing Reality is beyond our powers; metaphysical statements are meaningful, but we can never discover whether any metaphysical statement is true or false (or discover anything else interesting or important about the class of metaphysical statements).

METAPHYSICS: THE BIG QUESTIONS

Let us briefly examine an example of the strong form of the thesis that metaphysics is impossible. In the years between the world wars, the “logical positivists” (including Carnap and other members of the “Vienna Circle”) argued that the meaning of a statement consisted entirely in the predictions it made about possible experience. (This sort of view may be found in the first chapter of Carnap’s *Philosophy and Logical Syntax*, reprinted here as “The Rejection of Metaphysics.”) The logical positivists argued that metaphysical statements, statements that purported to describe Reality, made no predictions about experience. (The metaphysician asks, “Is time real, or are temporal phenomena mere appearances?” But our experiences would be the same – they would be just as they are – whether or not time was real. The metaphysician asks, “Are there universals, or is the appearance of there being attributes and relations a mere appearance, an illusion created by the way we think and speak?” But our experiences would be the same – like *this* – whether or not there were universals. And so, the logical positivists argued, for every metaphysical question. Metaphysical theses, being essentially attempts to *get behind* the way things appear to us, can make no predictions about the way things will appear to us.) Therefore, they argued, metaphysical statements are meaningless. Or, since “meaningless statement” is a contradiction in terms, the “statements” we classify as metaphysical are not really statements at all: they are things that look like statements but aren’t, rather as mannequins are things that look like human beings but aren’t.

But how does the logical positivist’s thesis fare by its own standards? Consider the statement,

The meaning of a statement consists entirely in the predictions it makes about possible experience.

Does this statement make any predictions about possible experiences? Could some observation show that this statement was true? Could some laboratory experiment show that it was false? It would seem not. It would seem that everything in the world would look the same – like *this* – whether or not this statement was true. And, therefore, if the statement is true it is meaningless; or, what is the same thing, if it is meaningful, it is false. Logical positivism would therefore seem to say of itself that it is false or meaningless; it would seem to be, as some philosophers say, “self-referentially incoherent.”

We have not the space to consider all the attempts that have been made to show that the idea of a reality that lies behind all appearances is in some sense defective. (Current exponents of “anti-realism” are only the latest example of such philosophers.) But, for what it is worth, we are convinced that all such attempts are victims of self-referential incoherency. The general case goes like this. Alfred the anti-metaphysician argues that any proposition that does not pass some test he specifies is in some sense defective (it is, say, self-contradictory or meaningless). And he argues that any metaphysical proposition must fail this test. But it invariably turns out that some proposition that is essential to Alfred’s anti-metaphysical argument itself fails to pass his test. Or so it seems to us that

it invariably turns out. The reader is warned, however, that most anti-metaphysicians will say that we are mistaken, and that their own anti-metaphysical arguments are not self-referentially incoherent. (The remainder will say that everyone who is anyone in philosophy knows that “the self-referential incoherency ploy” is without merit. This response has all the merits of a certain famous, if apocryphal, solicitor’s brief: “No case. Abuse plaintiff’s counsel.”)

What about the “weak form” of the thesis that metaphysics is impossible? Is the search for metaphysical truth a hopeless one, given the limitations of our intellects? Should one simply confess, “Such knowledge is too wonderful for me; it is high, I cannot attain unto it”? In our view, this question can be usefully discussed only in the context of a comprehensive and detailed examination of some actual and serious attempts at metaphysics.¹ But we believe that a close study of the sort of work typified by the contributions to this anthology reveals that a certain modest progress is attainable in metaphysics; those willing to put in some “honest toil” can at least hope to add to what we know about Reality. Progress in metaphysics often comes in disappointingly conditional packages, however; the most a metaphysician can usually be said to have conclusively shown is something of the form: “if such-and-such metaphysical thesis were true, then so-and-so would also have to be the case.”

In the remainder of this Introduction, we will attempt to give some content to the very abstract remarks we have made about the nature of metaphysics by examining a particular metaphysical problem.

A Metaphysical Problem: the Existence and Nature of Universals

One very important part of metaphysics has to do with what there is, with what exists. This part of metaphysics is called ontology. Ontology, that is, is that part of metaphysics that deals with metaphysical statements having general forms like “An *X* exists” and “There are *Ys*. ” (Here, it will be assumed that “there is (are)” and “exist(s)” mean essentially the same thing – that there is no important difference in meaning between “Horses exist” and “There are horses”. There are philosophers who deny this thesis: such philosophers exist.) In ontology, the second of our three requirements on a metaphysical statement is especially important – the requirement that the philosopher who makes a metaphysical statement be willing to take responsibility for the strict and literal consequences of the words used to make the statement. This is because we very frequently say things of the forms “An *X* exists” and “There are *Ys*” when we do not think there are *really* any *Xs* or *Ys*. An example will help to make what is meant by this clear.

Our friend Jan is an adherent of the metaphysical position known as materialism, the thesis that everything – everything without qualification – is material. We notice, however, that, despite her allegiance to materialism, Jan frequently says things that, when taken strictly and literally, are inconsistent with materialism. For example, just this morning we heard her say, “There’s a big hole in my

METAPHYSICS: THE BIG QUESTIONS

favorite blouse that wasn't there yesterday."² But no material object is a hole: material things are made of atoms, and nothing made of atoms is a hole; holes, so to speak, result from the *absence* of atoms. And yet Jan has said that there was one of them in her blouse. We point out to her that she has made a statement that is on the face of it inconsistent with materialism, and she replies:

It's true that I said there was a hole in my blouse, and that this statement, taken strictly and literally, implies that there is a hole; and it's true that a hole, if there really were such *things* as holes, wouldn't be a material thing. But I was speaking the language of everyday life; by the standards of metaphysics, I was speaking loosely. What I *could* have said, and what I would have said if I'd known that you were going to hold me responsible for the strict and literal consequences of my words, is that my blouse is *perforate*. The predicate "is perforate," when it is applied to a material object like my blouse, simply says something about the object's *shape*. If you perforate a coin, the resulting object will have a shape different from that of an imperforate, but otherwise identical, coin. When I say that a given material thing is perforate, this obviously does not imply that there is *another* thing, a thing not made of atoms, a thing called a "hole," that is "in" the material thing. The words "there's a hole in this thing" are just an idiomatic way of saying "this thing is perforate."

This speech provides an example of a philosophical tool, extremely important in ontology, called "paraphrase." Various idioms and expressions that are perfectly serviceable for everyday, practical purposes have metaphysically unwanted implications when they are interpreted strictly and literally – which is the way we are supposed to interpret a metaphysician's idioms and expressions. To find a *paraphrase* of a statement involving such "misleading" forms of words is to find a way of conveying what the statement is intended to convey that does not have the unwanted implications. (This is what we imagined Jan doing with the statement "There's a hole in my blouse.")

Metaphysicians have not spent a lot of time disputing about whether there really are holes. But they have spent a lot of time disputing about whether there really are so-called abstract things (such as properties, relations, propositions, and numbers). The medieval dispute about the reality of "universals" is an especially important example of this. This ancient dispute, or something very much like it, goes on today in several different forms. One of these "forms" is due to the work of the American philosopher W.V. Quine.³ We shall examine it. It will be our example of a way to approach a metaphysical problem.

A universal is, near enough, a property – such as humanity (the property that is "universal" to the members of the class of human beings and to the members of no more inclusive class), wisdom, the color blue, and widowhood. There are *apparently* properties. There is, for example, apparently such a thing as humanity. The members of the class of human beings, as the idiom has it, "have something in common," and what could this "something" be but the property "humanity"? It could certainly not be anything physical, for – Siamese twins excepted – no two human beings have any physical thing in common. And, of course, what goes for the class of human beings goes for the class of birds, the

class of white things, and the class of intermediate vector bosons: the members of each of these classes have something in common with one another, and what the members of a class have in common is a property – or so it appears. But there are metaphysicians who contend that this appearance is mere appearance and that *in reality* there are no properties. Other metaphysicians argue that in this case, at least, appearances are not misleading and that there really *are* properties. The metaphysicians who deny the real existence of properties are called nominalists and the metaphysicians who affirm the real existence of properties are called platonists.⁴ (Each of these terms could be objected to on historical grounds. But let us pass over these objections.)

How can the dispute between the nominalists and the platonists be resolved? Quine has proposed an answer to this question.⁵ Nominalists and platonists have different beliefs about what there is. How should one go about deciding what to believe about what there is? According to Quine, the problem of deciding what to believe about what there is is a very straightforward special case of the problem of deciding what to believe. (The problem of deciding what to believe is no trivial problem, to be sure, but it is a problem everyone is going to have somehow to come to terms with.) Let us look at the problem that is our present concern, the problem of what to believe about the existence of properties. If we want to decide whether to believe that there are properties – Quine tells us – we should examine the beliefs that we already have, and see whether any of them commits us to the existence of properties. If any does, then we have a reason to believe in the existence of properties: it is whatever reason we had for accepting the belief that commits us to the existence of properties – plus the general intellectual requirement that if one becomes aware that one's belief that *p* commits one to the further belief that *q*, then one should either believe that *q* or cease to believe that *p*.⁶ But let us consider an example. Suppose we find the following proposition among our beliefs:

Spiders share some of the anatomical features of insects.

A plausible case can be made for the thesis that this belief commits us to the existence of properties. We may observe, first, that it is very hard to see what an “anatomical feature” (such as “having an exoskeleton”) could be if it were not a property: “property,” “quality,” “characteristic,” “attribute,” and “feature” are all more or less synonyms. Does our belief that spiders share some of the anatomical features of insects therefore commit us to the existence of “anatomical features”? If we carefully examine the meaning of the sentence “Spiders share some of the anatomical features of insects,” we find that what it says is this:

There are anatomical features that insects have and spiders also have.

And it is a straightforward logical consequence of this proposition that there are anatomical features: If there are anatomical features that insects have and spiders also have, then there are anatomical features that insects have; if there are

METAPHYSICS: THE BIG QUESTIONS

anatomical features that insects have, then there are anatomical features – full stop.

Does this little argument show that anyone who believes that spiders share some of the anatomical features of insects is committed to platonism, to a belief in the existence of properties? How might a nominalist respond? Suppose we present this argument to Ned, a convinced nominalist (who believes, as most people do, that spiders share some of the anatomical features of insects). Assuming that Ned is unwilling simply to have inconsistent beliefs, there would seem to be four possible ways for him to respond to this argument:

- (1) He might become a platonist.
- (2) He might abandon his belief that spiders share many of the anatomical features of insects.
- (3) He might attempt to show that it does not after all follow from this belief, that there are anatomical features.
- (4) He might admit that his beliefs (his belief in nominalism and his belief that spiders share some of the anatomical features of insects) are apparently inconsistent, affirm his nominalistic faith that this inconsistency is apparent, not real, and confess that, although he is confident that there is some fault in our alleged demonstration that his belief about spiders and insects commits him to the existence of properties, he is at present unable to discover it.

Possibility (2) is not really very attractive. It is unattractive for at least two reasons. First, it seems to be a simple fact of biology that spiders share some of the anatomical features of insects. Secondly, there are many, many “simple facts” that could have been used as the premise of an essentially identical argument for the conclusion that there are properties. (For example: Elements in the same column in the Periodic Table tend to have many of the same chemical properties; Some of the most important characteristics of the nineteenth-century novel are rarely present in the twentieth-century novel.) Possibility (4) is always an option, but no philosopher is likely to embrace it except as a last resort. What Ned is likely to do is to try to avail himself of Possibility (3). He is likely to try to show that his belief about spiders and insects does not in fact commit him to platonism. What he will attempt to do in respect of this belief (and of all the others among his beliefs that apparently commit him to a belief in properties) is just what Jan did in respect of the belief that apparently committed her to a belief in holes: he will try to find a *paraphrase*, a sentence that (i) he could use in place of “Spiders share some of the anatomical features of insects” and (ii) does not even seem to have “There are anatomical features” as one of its logical consequences. If he can do this, then he will be in a position to argue that the commitment to the existence of properties that is apparently “carried by” his belief about spiders and insects is only apparent. And he will be in a position to

argue – no doubt further argument would be required to establish this – that the apparent existence of properties is mere appearance (an appearance that is due to the forms of words we use).

Is it possible to find such a paraphrase? (And to find paraphrases of all the other apparently true statements that seem to commit those who make them to the reality of properties?) This is a difficult and technical question. We record our conviction that it is at least very hard to do so.⁷ If Quine is right about “ontological commitment,” therefore, there is no easy way for anyone to be a consistent nominalist.

It must be emphasized that we have said almost nothing about the *nature* of “properties.” If what we have said so far is correct, some of the sentences we use to express certain very ordinary and non-metaphysical beliefs, sentences like “Spiders share some of the anatomical features of insects” and “Elements in the same column in the Periodic Table tend to have many of the same chemical properties,” define what we may call the “property role”: a property is whatever it is (beyond ordinary things like spiders and chemical elements) that using these sentences to express our beliefs carries *prima facie* commitment to. And if what we have said so far is correct, it is very hard to avoid the conclusion that objects of *some* sort play the property role. But philosophers who accepted this conclusion could differ fundamentally about the nature of the objects that play this role. Some philosophers think that the property role is played by things that are in some sense constituents of objects, that properties are in some very subtle and abstract sense of the word *parts* of the objects whose properties they are.⁸ Other philosophers (including at least one of the editors of this volume) think that this conception of properties is not so much false as meaningless and that the things that play the property role are in no sense parts or constituents of objects, but simply things that can be “said of” objects. According to this view of the nature of properties, the property “being white” is simply something that can be said truly of table salt and the Taj Mahal and cannot be said truly of copper sulfate or the Eiffel Tower. (But what kind of thing would *that* be? You may well ask.) There has perhaps been little progress since the Middle Ages in the attempt to say anything both informative and meaningful about the *nature* of universals, the nature of the things that play the property role. But it can be plausibly argued that even if we do not understand universals much better than the medieval philosophers, we now have a better understanding of the *problem* of universals. We now see that the best way to look at the debate between the nominalist and the platonist is as follows: the task of the nominalist is to establish the conclusion that our beliefs about ordinary things do not commit us to the thesis that anything plays the property role. The task of the platonist is to attempt to establish the conclusion that our beliefs about ordinary things do commit us to the existence of things that play the property role, and to attempt to give a plausible account of the nature of these things.

METAPHYSICS: THE BIG QUESTIONS

Notes

- 1 For an extremely interesting and sophisticated defense of the weak form of the thesis, see Colin McGinn, *Problems in Philosophy: the Limits of Inquiry* (Oxford: Blackwell, 1993).
- 2 This example is based on David and Stephanie Lewis, "Holes", *Australasian Journal of Philosophy*, 48 (1970), pp. 206–12.
- 3 For another, very different form, see David Armstrong, *Universals: an Opinionated Introduction* (Boulder, Col.: Westview Press, 1989).
- 4 The philosophers we are calling platonists are often called realists. We will avoid the terms "realist" and "realism," since they have several other meanings in metaphysics.
- 5 The issues that we are about to discuss are generally said to pertain to "ontological commitment," a term that is due to Quine. For Quine's views on ontological commitment, see his classic essay "On What There Is," reprinted in Laurence and Macdonald, eds, *Contemporary Metaphysics: a Reader*; and chapter vii, "Ontic Decision," of his *Word and Object* (Cambridge, Mass.: MIT Press, 1960).

Discussions of ontological commitment are generally rather technical. They are technical because they represent issues of ontological commitment as essentially related to "the existential quantifier," the symbol used in formal logic (it is most often a backwards "E") to express "there is" or "there exists." The tendency of philosophers to connect issues of ontological commitment with the existential quantifier is (in one way) entirely justified, and (in another) somewhat misleading. It is justified because any *technically fully adequate* formulation of Quine's theses on ontological commitment must involve the existential quantifier and the related device of "bound variables." It is misleading because it suggests that it is impossible to present an account of the essential philosophical points contained in these theses without at some point introducing the existential quantifier – and not simply the symbol, but the technical apparatus that governs its use in formal logic and the various philosophical disputes that have arisen concerning its "interpretation." And this is false: it is possible to give a useful introductory account of the philosophical points contained in Quine's various discussions of ontological commitment that contains no "existential apparatus" but the ordinary words and phrases – "there is," "exists" – for which the existential quantifier is the formal replacement. The discussion of "there is" and paraphrase in this introduction is an attempt at such an introductory account of these points.

- 6 Suppose we were to discover that some belief of ours – that Mars has two moons, let us say – committed us to the existence of properties. Should that discovery move us to question, or perhaps even to abandon, our belief that Mars had two moons? That would depend on whether we had, or thought we had, some reason to believe that there were no properties. If we did think we had some reason to believe that there were no properties, we should have to try to decide whether our reason for thinking that Mars had two moons (presumably we have one) was more or less compelling than our reason for thinking that there were no properties.
- 7 In one sense, Quine himself believes that the required paraphrase is possible. He believes that statements like our "spider-insect" statement can be understood in such a way that they commit those who make them to nothing other than *sets* – besides, of course, spiders and insects or whatever other "ordinary" objects the statements may mention. But these sets, it must be emphasized, are very far from being ordinary objects. The set of all spiders, for example, is not a spider or any other sort

- of physical object, and reference to “the set of spiders” cannot be dismissed as a mere linguistic device for referring to all spiders collectively: sets are *objects*. Sets are, in fact, from the point of view of those who call themselves nominalists, hardly more acceptable than properties, and, in present-day discussions of ontology, “nominalism” is generally taken to imply that there are no such objects as sets.
- 8 Two versions of the view are represented in this anthology: Bertrand Russell, “The Principle of Individuation,” and D. C. Williams, “The Elements of Being.” Russell’s version is criticized here by Zimmerman, in “Distinct Indiscernibles and the Bundle Theory.”

Suggestions for Further Reading

- Aune, Bruce, *Metaphysics: the Elements* (Minneapolis: University of Minnesota, 1985), chs 1 and 2: “What is Metaphysics?” and “Existence.”
- Benardete, José A., *Metaphysics: the Logical Approach* (Oxford: Oxford University Press, 1989), Part 1
- Burke, Michael, “Existence,” in Stephen Hales, ed., *Metaphysics: Contemporary Readings* (Belmont, Cal.: Wadsworth, 1998).
- Carter, William R., *The Elements of Metaphysics* (Philadelphia, Penn.: Temple University Press, 1990), ch. 1: “Metaphysics.”
- Gardner, Martin, *The Night is Large* (New York: St Martin’s Press, 1996), chs 32 and 43: “The Significance of Nothing” and “The Irrelevance of Everything.”
- Hamlyn, D. W., *Metaphysics*. (Cambridge, UK: Cambridge University Press, 1984), ch. 1: “Introduction.”
- Hasker, William, *Metaphysics Constructing a World View* (Downers Grove, Ill., and Leicester, UK: InterVarsity Press, 1983), ch. 1: “Introducing Metaphysics.”
- Jubien, Michael, *Contemporary Metaphysics* (Oxford: Blackwell, 1997), ch. 1: “Metaphysics.”
- Laurence, Stephen, and Cynthia Macdonald, *Contemporary Metaphysics: a Reader* (Oxford: Blackwell, 1998), Section 1: “Ontological Commitment and Methodology.”
- Loux, Michael, *Metaphysics: a Contemporary Introduction* (London: Routledge, 1997), Introduction.
- Post, John F., *Metaphysics: a Contemporary Introduction* (New York: Paragon House, 1991), ch. 1: “Is Metaphysics Possible?”
- Russell, Bertrand, *The Problems of Philosophy* (New York: H. Holt, 1912), ch. 15: “The Value of Philosophy.”
- van Inwagen, Peter, *Metaphysics* (Boulder, Col.: Westview Press, 1993), ch. 1: “Introduction.”

PART ONE

WHAT ARE THE MOST GENERAL FEATURES OF THE WORLD?

Introduction

What is the Relationship between an Individual and its Characteristics?

- 1 Universals and Resemblances: Chapter 1 of *Thinking and Experience*
H. H. PRICE
- 2 The Elements of Being
D. C. WILLIAMS
- 3 The Principle of Individuation: an Excerpt from *Human Knowledge, its Scope and Limits*
BERTRAND RUSSELL
- 4 Distinct Indiscernibles and the Bundle Theory
DEAN W. ZIMMERMAN

What is Time? What is Space?

- 5 Time: an Excerpt from *The Nature of Existence*
J. McT. E. McTAGGART
- 6 McTaggart's Arguments against the Reality of Time: an Excerpt from *Examination of McTaggart's Philosophy*
C. D. BROAD
- 7 The Notion of the Present
A. N. PRIOR
- 8 The General Problem of Time and Change: an Excerpt from *Scientific Thought*
C. D. BROAD
- 9 The Space-Time World: an Excerpt from *Philosophy and Scientific Realism*
J. J. C. SMART
- 10 Topis, Soris, Noris: an Excerpt from *The Existence of Space and Time*
IAN HINCKFUSS

- 11 Some Free Thinking about Time
A. N. PRIOR
- 12 The Fourth Dimension: an Excerpt from *The Ambidextrous Universe*
MARTIN GARDNER
- 13 Incongruent Counterparts and Higher Dimensions
JAMES VAN CLEVE
- 14 Achilles and the Tortoise
MAX BLACK
- 15 A Contemporary Look at Zeno's Paradoxes: an Excerpt from *Space, Time, and Motion*
WESLEY C. SALMON
- 16 Grasping the Infinite
JOSÉ A. BERNADETE
- 17 The Paradoxes of Time Travel
DAVID LEWIS

How do Things Persist through Changes of Parts and Properties?

- 18 Of Confused Subjects which are Equivalent to Two Subjects: an Excerpt from *The Port-Royal Logic*
ANTOINE ARNAULD AND PIERRE NICOLE
- 19 Identity through Time
RODERICK M. CHISHOLM
- 20 Identity, Ostension, and Hypostasis
W. V. O. QUINE
- 21 Identity: an Excerpt from *Quiddities*
W. V. O. QUINE
- 22 In Defense of Stages: Postscript B to "Survival and Identity"
DAVID LEWIS
- 23 Some Problems about Time
PETER GEACH
- 24 The Problem of Temporary Intrinsics: an Excerpt from *On the Plurality of Worlds*
DAVID LEWIS
- 25 Temporary Intrinsics and Presentism
DEAN W. ZIMMERMAN

How do Causes Bring about their Effects?

- 26 Constant Conjunction: an Excerpt from *A Treatise of Human Nature*
DAVID HUME
- 27 Efficient Cause and Active Power: an Excerpt from *Essays on the Active Powers of the Human Mind*
THOMAS REID
- 28 Psychological and Physical Causal Laws: an Excerpt from *The Analysis of Mind*
BERTRAND RUSSELL
- 29 Causality: an Excerpt from *A Modern Introduction to Logic*
L. SUSAN STEBBING
- 30 Causality and Determination
G. E. M. ANSCOMBE

Introduction

The readings assembled here concern some of the most pervasive and, ultimately, puzzling features of the world. One feature is just that the world contains things *and their features* – that is, their characteristics or properties. But how deep is this seeming dichotomy between individual things and the properties they exhibit? Most things – perhaps all things – are in space and time; but what is space and what is time? Is time just another dimension, in many ways like the spatial ones? Time seems quite different from space in some ways, most notably in having a built-in direction. But does the difference between forward and backward directions in time result from the direction of a process of “absolute becoming” – a process producing a distinction between past, present, and future with no spatial analogue? Are finite regions of space infinitely divisible; and if they were, would that render motion impossible? Are things that persist through time spread out in the temporal dimension in the same way they are spread out in the spatial dimensions? Are causal regularities anything more than uniformities that hold everywhere in space-time? Or is there some deeper relation that binds cause to effect?

A What is the Relationship between an Individual and its Characteristics?

Two closely related problems come up in these readings. The first is the status of characteristics or properties, attributes that can be shared by any number of individuals. When someone says that two things are exactly similar with respect to shape, this might be taken to imply that there is something – namely, a certain shape – that they have in common. Does this require that we believe in *shapes* in addition to the things that *have* shapes? If so, are these shapes in space and time located right where the things that have them are located? Or are they, strictly speaking, not themselves in space and time at all?

In “Universals and Resemblance,” H. H. Price contrasts a view he calls “the Aristotelian doctrine of *universalia in rebus*” with the “Philosophy of Ultimate Resemblances.” According to the former, the characteristics attributed to things are extra entities located in the same place as the characterized objects. When two co-existing objects have the same shape, there is something – a “universal” – that is present in two places at once. The Philosophy of Resemblances attempts to avoid commitment to such entities. D. C. Williams offers another way to do without universals, this time replacing them with sets of what he calls “tropes” – instances of properties, nonrepeatable things of which substances are said to be composed without remainder. In “The Principle of Individuation,” Bertrand Russell assumes the doctrine of *universalia in rebus*, and argues that substances may be composed, without remainder, of *these* sorts of universals. Zimmerman develops an objection to this sort of “bundle theory” of substance in a dialogue between a defender of the Russellian view and a critic.

The problem of universals receives extensive treatment in the general “Introduction: What is Metaphysics?,” above (see section 3, “A Metaphysical Problem:

METAPHYSICS: THE BIG QUESTIONS

the Existence and Nature of Universals"). The discussion there dovetails nicely with the readings from Price and Williams.

Suggestions for Further Reading

- Aune, Bruce, *Metaphysics: the Elements* (Minneapolis: University of Minnesota, 1985), chs 3 and 4: "Universals and Particulars" and "Linguistic Arguments for Abstracta."
- Armstrong, D. M., *Universals: an Opinionated Introduction* (Boulder, Col.: Westview Press, 1989).
- Benardete, José, *Metaphysics: the Logical Approach* (Oxford: Oxford University Press, 1989), part 2.
- Carter, William R., *The Elements of Metaphysics* (Philadelphia, Penn.: Temple University Press, 1990) ch. 4: "Substance."
- Hamlyn, D. W., *Metaphysics* (Cambridge, UK: Cambridge University Press, 1984), chs 4 and 5: "Substance" and "Particular and General".
- Jubien, Michael, *Contemporary Metaphysics* (Oxford: Blackwell, 1997), chs 2, 3, and 8: "Numbers," "Platonism," and "Modality."
- Loux, Michael, *Metaphysics: a Contemporary Introduction* (London: Routledge, 1997), chs 1, 2, and 3: "The Problem of Universals I – Metaphysical Realism," "The Problem of Universals II – Nominalism," and "Concrete Particulars I – Substrata, Bundles, and Substances."
- Russell, Bertrand, *The Problems of Philosophy* (New York: H. Holt, 1912), chs 9 and 10: "The World of Universals" and "On our Knowledge of Universals."
- van Inwagen, Peter, *Metaphysics* (Boulder, Col.: Westview Press, 1993), ch. 2: "Individuality."
- Laurence, Stephen and Cynthia Macdonald, *Contemporary Metaphysics: a Reader* (London: Routledge, 1997), Section 3: "Properties and Universals."

B What is Time? What is Space?

The readings under this heading fall into four groups. First there is McTaggart's paradox and Broad's criticism of it. One part of McTaggart's argument is supposed to refute the notion that there is an objective fact about what events are present, a fact that is constantly changing. Many contemporary philosophers defend this part of McTaggart's argument (see the works by D. H. Mellor and Paul Horwich listed below); but Broad's refutation of McTaggart still seems to us to be decisive.

The second group represents three different theories about the nature of the distinction between past, present, and future: (i) only what is present is real, (ii) present and past are both real, (iii) present, past, and future are equally real. The first essay in this group is A. N. Prior's "The Notion of the Present." The doctrine Prior advocates is sometimes called "presentism": the thesis that what exists is only what exists at present – events and individuals that are wholly past or future are strictly nonexistent. Prior's first essay is followed by C. D. Broad's defense of a "growing block" theory of time – both present *and past* events and things are real; the distinction between present and past is just the difference between what has just been added to the sum total of reality and what is buried deeper within the

four-dimensional block of space-time. J. J. C. Smart advocates the view of time that is probably most popular with contemporary philosophers: all events and things are equally real, spread out in a four-dimensional space-time block; and “the concepts of past, present, and future have significance relative only to human thought and utterance and do not apply to the universe as such.” Smart’s essay is followed by two criticisms of this view: Ian Hinckfuss argues, against Smart, that the fact that we use tensed language to pick out different times does not adequately account for the supposed “illusion” that time is passing or flowing. Prior’s lecture, “Some Free Thinking about Time,” includes another criticism of Smart (the “thank goodness that’s over” argument), discusses connections between logic and theories of time, and also offers a response to the contention that relativity implies a four-dimensional block universe.

The third part of this section consists of a pair of readings on the question whether the phenomenon of *enantiomorphism* supports the doctrine of absolute space. A right- and left-hand glove are a pair of *enantiomorphs*—objects alike in all geometrical properties, but differing in “shape” nonetheless. Martin Gardner explains how Kant came to the view that the difference between two such objects must be due to differences in their relations to a Newtonian absolute space in which they are embedded. Gardner criticizes Kant’s argument; part of his criticism depends upon the fact that, in a *four*-dimensional space, the left-hand glove could be “flipped over” and turned into a right-hand glove. James Van Cleve, in a paper appearing here for the first time, considers the implications of Kant’s argument in more detail. The second half of his paper (the “Dialogue on Higher Dimensions”) raises objections to the possibility of spaces with more than three dimensions.

The final four papers concern familiar paradoxes of space and time—the first three, paradoxes of the actually infinite; the last, paradoxes of time travel. Only one comment seems necessary here, and that concerns the role of José Benardete’s largely new essay. The question he considers is whether we can claim to have any notion of an infinitely large collection other than something like “a collection with more members than any human being could count.” In order to adopt Wesley Salmon’s response to Zeno’s paradoxes, we must obviously have a grasp of mathematical infinities that is more robust than this; but, as Benardete’s parable of the “Gumquats” makes evident, there are reasons to think such notions must remain forever beyond our ken. Benardete offers a “dispositionalist” response to Kripke’s Wittgensteinian skepticism about meaning, and concludes with a brief statement of the implications of his position for the infinite-task paradoxes.

There are quite a few connections between these readings and others in this anthology. Near the end of “The General Problem of Time and Change,” Broad tries to explain how talk about future things can be meaningful without implying that future things exist. This passage illustrates the use of paraphrase to avoid unwanted ontological commitments, a strategy discussed in section 3 of the introductory essay, “Introduction: What is Metaphysics?” The excerpts from Quine and Lewis in the next subsection present a four-dimensionalist metaphysics much like Smart’s. Prior’s defense of tense logic and his response to relativity in “Some Free Thinking . . .” resemble positions advocated by Peter Geach in “Some Prob-

METAPHYSICS: THE BIG QUESTIONS

lems about Time,” below (Geach’s paper also includes a brief discussion of time travel). Zimmerman’s “Temporary Intrinsics and Presentism” (below) continues exploration of the relations between “taking tense seriously” and presentism, including a brief defense of Prior’s “thank goodness that’s over” argument.

Suggestions for Further Reading

- Abbott, E. A., *Flatland: a Romance of Many Dimensions* (New York: Dover, 1952).
- Aune, Bruce, *Metaphysics: the Elements* (Minneapolis: University of Minnesota, 1985), ch. 6: “Worlds, Objects, and Structure.”
- Banchoff, Thomas S., *Beyond the Third Dimension: Geometry, Computer Graphics, and Higher Dimensions* (New York: Scientific American Library, 1990).
- Benardete, José A., *Infinity* (Oxford: Clarendon Press, 1964).
- Borges, Jorge Luis, “Avatars of the Tortoise,” included in *Labyrinths, Selected Stories and Other Writings* (New York: New Directions, 1964).
- Chisholm, Roderick M, and Dean W. Zimmerman, “Tense and Theology”, *Nous*, 31 (1997), pp. 262–5.
- Hamlyn, D. W., *Metaphysics* (Cambridge, UK: Cambridge University Press, 1984) ch. 7: “Space and Time.”
- Gardner, Martin, *The Night is Large* (New York: St Martin’s Press, 1996), ch. 7: “Can Time Stop? The Past Change?”
- , *Time Travel and other Mathematical Bewilderments* (New York: W. H. Freeman, 1988), ch. 1: “Time Travel.”
- , *The Unexpected Hanging and other Mathematical Diversions* (New York: Simon and Schuster, 1969), ch. 6: “The Church of the Fourth Dimension.”
- Horwich, Paul, *Asymmetries in Time* (Cambridge, Mass.: MIT, 1987), ch. 2.
- Lovecraft, H. P., *At the Mountains of Madness and other Novels* (Sauk City, Wisc.: Arkham House, 1964); includes the story “The Dreams in the Witch-House” – witches, demons, and aliens in the fourth dimension!
- Mellor, D. H., *Real Time* (Cambridge: Cambridge University Press, 1981), see ch. 6: “The Unreality of Tense,” for Mellor’s defense of McTaggart.
- Moore, A. W., *The Infinite* (London: Routledge, 1990).
- Quinton, Anthony, “Spaces and Times,” *Philosophy*, 37 (1962), pp. 130–47.
- Ray, Christopher, *Time, Space and Philosophy* (London: Routledge, 1991).
- Rucker, Rudy, *Geometry, Relativity and the Fourth Dimension* (New York: Dover, 1977).
- , *The Fourth Dimension: a Guided Tour of the Higher Universes* (Boston: Houghton Mifflin, 1984).
- , *Infinity and the Mind* (Boston, Mass.: Birkhauser, 1982).
- (ed.), *Mathenauts: Tales of Mathematical Wonder* (New York: Arbor House, 1987); includes the following stories involving the fourth dimension: Greg Bear, “Tangents”; Martin Gardner, “Left or Right?”; and Rudy Rucker, “Message Found in a Copy of *Flatland*.”
- Shoemaker, Sydney, “Time without Change”, *Journal of Philosophy*, 66 (1969), pp. 363–81; reprinted in his *Identity, Cause, and Mind* (Cambridge: Cambridge University Press, 1984).
- Smith, Quentin, and L. Nathan Oaklander, *Time, Change and Freedom: an Introduction to Metaphysics* (London: Routledge, 1995), Dialogues 3 and 6: “The Relational and Substantival Theories of Time” and “The Passage of Time,” and Appendix A: “Physical Time in Einstein’s Special Theory of Relativity.”

- Swinburne, Richard, *Space and Time* (London: Macmillan, 1968; 2nd edn 1981).
 Taylor, Richard, *Metaphysics* (Englewood Cliffs, N.J.: Prentice-Hall, 1992) chs. 7, 8, 9:
 “Space and Time,” “The Relativity of Time and Space,” “Temporal Passage.”
 van Fraassen, Bas, *An Introduction to the Philosophy of Space and Time* (New York: Random House, 1970).

C How do Things Persist through Changes of Parts and Properties?

To the question: How do things persist through changes of parts?, the authors of the *Port-Royal Logic* and Roderick Chisholm all answer: They don’t! The view defended in these selections is called “mereological essentialism,” the thesis that nothing can survive the gain or loss of any parts. A doctrine known as “the metaphysics of temporal parts” is advocated in the excerpts from W. V. O. Quine and David Lewis. Their position can easily allow for change of parts over time. But many philosophers, including Peter Geach, find the view hard to swallow. In “The Problem of Temporary Intrinsics,” Lewis argues that only the metaphysics of temporal parts is compatible with the fact that things change in their intrinsic properties. Zimmerman claims that those who “take tense seriously” should accept “Presentism” (see the discussion of Prior’s “The Notion of the Present,” above); that the presentist has (as Lewis admits) no problem with temporary intrinsics; and that Lewis is wrong to say that presentism “goes against what we all believe.”

Suggestions for Further Reading

- Armstrong, D. M., *Universals: an Opinionated Introduction* (Boulder, Col.: Westview Press, 1989), pp. 2–4.
- Aune, Bruce, *Metaphysics: the Elements* (Minneapolis: University of Minnesota, 1985), ch. 5: “Changing Things.”
- Carter, William R., *The Elements of Metaphysics* (Philadelphia, Penn.: Temple University Press, 1990), chs 5 and 6: “Parts and Wholes” and “Change.”
- Hales, Stephen D., *Metaphysics: Contemporary Readings* (Belmont, Cal.: Wadsworth, 1998), Section 10: “Mereology.”
- Jubien, Michael, *Contemporary Metaphysics* (Oxford: Blackwell, 1997), chs 4 and 9: “Identity” and “Things and Their Parts.”
- Loux, Michael, *Metaphysics: a Contemporary Introduction* (London: Routledge, 1997), ch. 6: “Concrete Particulars II – Persistence Through Time.”
- Smith, Quentin, and L. Nathan Oaklander, *Time, Change and Freedom: an Introduction to Metaphysics* (London: Routledge, 1995) Dialogue 5: “The Problem of Change.”
- Taylor, Richard, *Metaphysics* (Englewood Cliffs, N.J.: Prentice-Hall, 1992), ch. 7: “Space and Time.”
- van Inwagen, Peter, *Metaphysics* (Boulder, Col.: Westview Press, 1993), pp. 169–171.

D How do Causes Bring about their Effects?

David Hume sets the stage for contemporary debates about causation by asking

METAPHYSICS: THE BIG QUESTIONS

what more there could be to the notion besides spatiotemporal contiguity (causes never operate at a spatial distance, or over a temporal gap), temporal succession (causes precede their effects), and constant conjunction (resembling causes are always and everywhere followed by resembling effects). Thomas Reid rejects Hume's reduction of causation to such relations, and claims to have immediate acquaintance with active causal power to produce effects – but in only one case: that of his own actions. (Reid's position finds expression elsewhere in this anthology: see Roderick Chisholm's "Human Freedom and the Self," and Timothy O'Connor's "The Agent as Cause." In "The Mystery of Metaphysical Freedom," van Inwagen questions the explanatory significance of Reid's claim.)

Bertrand Russell, in the spirit of Hume, argues that there's nothing more to the relation of cause and effect than "nearly invariable antecedence." Near the end of the selection from her *A Modern Introduction to Logic*, Susan Stebbing responds to Russell's attempt to undermine more substantive accounts of the causal relation. But her main goal is to show how common-sense causal notions become refined by scientific knowledge, resulting in an analysis of causation in terms of (what has come to be called) "nomic subsumption": the instantiation of a causal law by a pair of events. In the final reading of this section, Elizabeth Anscombe argues (*contra* Stebbing, Russell, and Hume) that events can be causally related without their falling under anything like a causal law.

We have already noted the affinities among Reid, Chisholm, and O'Connor. Another important connection with topics discussed elsewhere in the volume stems from Stebbing's discussion of "immanent causation." In "Personal Identity: A Materialist Account," Sydney Shoemaker defends, for the case of persons, a version of what he calls "the causal continuity account" of identity over time. On such a theory, earlier and later stages of a persisting thing must be bound together by relations of immanent causation. The same sort of theory is put to work in David Lewis's "The Paradoxes of Time Travel."

Suggestions for Further Reading

- Armstrong, D. M., *Universals: an Opinionated Introduction* (Boulder, Col.: Westview Press, 1989), pp. 82–4.
- Aune, Bruce, *Metaphysics: the Elements* (Minneapolis: University of Minnesota, 1985), ch. 6: "Worlds, Objects, and Structure."
- Benardete, José, *Metaphysics: the Logical Approach* (Oxford: Oxford University Press, 1989), ch. 22: "Causation."
- Carter, William R., *The Elements of Metaphysics* (Philadelphia, Penn.: Temple University Press, 1990), ch. 9: "Causal Determinism."
- Hales, Steven D., *Metaphysics: Contemporary Readings* (Belmont, Cal.: Wadsworth, 1998), Section 7: "Concreta: Events," readings by Bennett, Kim, and Lombard.
- Laurence, Stephen, and Cynthia Macdonald, *Contemporary Metaphysics: a Reader* (Oxford: Blackwell, 1998), Section 5: "Events," readings by Lombard and Kim.
- Loux, Michael, *Metaphysics: A Contemporary Introduction* (London: Routledge, 1997), ch. 4: "Propositions and Their Neighbors."
- Taylor, Richard, *Metaphysics* (Englewood Cliffs, N.J.: Prentice-Hall, 1992), ch. 10: "Causation."

What is the Relationship between an Individual and its Characteristics?

1 Universals and Resemblances: Chapter 1 of *Thinking and Experience** ---

H. H. Price

When we consider the world around us, we cannot help noticing that there is a great deal of recurrence or repetition in it. The same colour recurs over and over again in ever so many things. Shapes repeat themselves likewise. Over and over again we notice oblong-shaped things, hollow things, bulgy things. Hoots, thuds, bangs, rustlings occur again and again.

There is another and very important sort of recurrence which we also notice. The same pattern or mode of arrangement is found over and over again in many sets of things, in many different pairs of things, or triads, or quartets, as the case may be. When A is above B, and C is above D, and E is above F, the above-and-below pattern or mode of arrangement recurs in three pairs of things, and in ever so many other pairs of things as well. Likewise we repeatedly notice one thing inside another, one preceding another, one thing between two others.

These recurrent features sometimes recur singly or separately. The same colour recurs in this tomato, that sunset sky, and this blushing face; there are few other features, if any, which repeat themselves in all three. But it is a noteworthy fact about the world that there are *conjoint* recurrences as well as separate ones. A whole group of features recurs again and again as a whole in many objects. Examine twenty dandelions, and you will find that they have many features in common; likewise fifty cats have very many features in common, or two hundred lumps of lead. In such cases as these there is conjoint recurrence of many different features. Again and again they recur together in a clump or block. This is how it comes about that many of the objects in the world group themselves together into Natural Kinds. A Natural Kind is a group of objects which have *many* (perhaps indefinitely many) features in common. From observing that an object has some of these features, we can infer with a good deal of probability that it has the rest.

These constant recurrences or repetitions, whether separate or conjoint ones, are what make the world a dull or stale or boring place. The same old features

* From H. H. Price, *Thinking and Experience* (London: Hutchinson's University Library, 1953).

keep turning up again and again. The best they can do is to present themselves occasionally in new combinations, as in the black swan or the duck-billed platypus. There is a certain *monotony* about the world. The extreme case of it is found where the same old feature repeats itself in all parts of a single object, as when something is red all over, or sticky all through, or a noise is uniformly shrill throughout its entire duration.

Nevertheless, this perpetual repetition, this dullness or staleness, is also immensely important, because it is what makes conceptual cognition possible. In a world of incessant novelty, where there was no recurrence at all and no tedious repetitions, no concepts could ever be acquired; and thinking, even of the crudest and most primitive kind, could never begin. For example, in such a world nothing would ever be recognizable. Or again, *in so far as* there is novelty in the world, non-recurrence, absence of repetition, so far the world cannot be thought about, but only experienced.

Hitherto I have been trying to use entirely untechnical language, so that we may not commit ourselves unawares to any philosophical theory. But it is at any rate not unnatural, it is not a *very* wild piece of theorizing, to introduce the words ‘quality’ and ‘relation’ for referring to those facts about the world to which I have been trying to draw the reader’s attention. A *quality*, we say, is a recurrent feature of the world which presents itself in individual objects or events taken singly. Redness or bulginess or squeakiness are examples. A *relation*, on the other hand, is a recurrent feature of the world which presents itself in complexes of objects or events, such as this beside that, this preceding that, or B between A and C. It is also convenient sometimes to speak of *relational properties*. If A precedes B, we may say that A has the relational property of preceding B, and B has the converse relational property of succeeding A.

One further remark may be made on the distinction between qualities and relations. I said just now that a quality presents itself in individual objects or events taken singly, and a relation in complexes of objects or events. But it must not be forgotten that an individual object or event usually (perhaps always) has an *internal* complexity. In its history there is a plurality of temporal phases, and often it has a plurality of spatial parts as well. And there are relations between these parts, or these phases, which it has. Such relations *within* an individual object or event are sometimes said to constitute the ‘structure’ of the object or event. For scientific purposes, and even for purposes of ordinary common sense prediction, what we most need to know about any object or process is its structure. And from this point of view the chief importance of qualities, such as colour or hardness or stickiness, is that they often enable us to infer the presence of a structure more minute than our unaided senses would reveal. It has often been maintained that sensible qualities are ‘subjective’. But subjective or not, they have a most important function. They give us a clue to what the minute structure of objects and events is. If a gas smells like rotten eggs, we can infer that it is sulphuretted hydrogen.

The terms ‘quality’ and ‘relation’ enable us to give a simple analysis of *change*. The notion of change has puzzled some philosophers greatly, ever since Heracleitus, or some disciple of his, remarked long ago that *πάντα ῥεῖ* ‘all things

flow'. Indeed, it has sometimes led them to suppose that this world *is* a world of perpetual novelty after all, and not the tedious or boring or repetitious world which it has to be, if conceptual cognition is to be possible. They have, therefore, concluded – rightly, from their premises – that all conceptual cognition is radically erroneous or illusory, a kind of systematic distortion of Reality; so that *whatever* we think, however intelligent or however stupid we may be, we are in error. On this view, only non-conceptual cognition – immediate experience or direct intuition – can be free from error.

These conclusions are so queer that we suspect something is wrong with the premises. We can now see what it is. The notion of Change, as Plato pointed out, has itself to be analysed in terms of the notions of Quality and Relation. In qualitative change, as when an apple changes from being green to being red, an object has quality q_1 at one time and a different quality q_2 at a later time. In relational change, an object A has a relation R_1 to another object B at one time, and a different relation R_2 to B at a later time. At 12 o'clock, for example, it is six inches away from B, at 12.5 it is a mile away from B; at one time the relation it has to B is the relation 'hotter than', at another the relation 'as hot as', at another the relation 'cooler than' . . .

We may now sum up the results of this ontological discussion so far by introducing another technical term, again not so *very* technical, the term 'characteristic'. Characteristics, we say, are of at least two different types, qualities and relations. What has been said so far then comes to this: there are *recurrent characteristics* in the world, which repeat themselves over and over again in many different contexts. Is it not just an obvious fact about the world, something we cannot help noticing whether we like it or not, that there *are* recurrent characteristics? Now these recurrent characteristics have been called by some philosophers *universals*. And the line of thought we have been pursuing leads very naturally to the traditional Aristotelian doctrine of *universalia in rebus*, 'universals in things'. (To provide for universals of relation, 'things' must be understood to cover complexes as well as individuals. The *res* which the universal 'beside' is in is not this, nor that, but this-and-that.)

I do not propose to discuss the Platonic doctrine of *universalia ante rem*, 'universals anterior to (or independent of) things'. This is not because I think it uninteresting or unimportant, but merely because it is more remote from common sense and our ordinary everyday habits of thought than the Aristotelian theory of *universalia in rebus*. It is a sufficiently difficult task in these days to convince people that there is any sense in talking of universals at all, even in the mild and moderate Aristotelian way.

The doctrine of *universalia in rebus* may, of course, be mistaken, or gravely misleading. There certainly are objections to it, as we shall find presently. But I cannot see that it is in the least absurd or silly, as the most approved thinkers nowadays seem to suppose. Nor can I see that it arises entirely from erroneous views about language, as the same thinkers seem to suppose; for example, from the superstition that all words are names, from which it would follow that general or abstract words must be names of general or abstract entities. On the contrary, this philosophy seems to me to be the result, and the very natural

result, of certain *ontological* reflections. It seems to me to arise from reflections about the world; from consideration of what things are, and not – or certainly not merely – from consideration of the way we talk about them. On the contrary, it could be argued that we talk in the way we do, using general terms and abstract terms, because of what we find the world to be; because we find or notice *recurrences* in it.

Let us now consider how the doctrine of *universalia in rebus* might mislead us, although it arises in this natural and plausible way from the ontological considerations which we have been discussing. One danger of it obviously is that universals may be regarded as a sort of *things* or entities, over and above the objects or situations in which they recur. We may indeed emphasize the word ‘in’. We may insist that universals are *in* things, and not apart from them as the doctrine of *universalia ante rem* maintains. But is the danger of supposing that they are themselves things or quasi-things entirely removed? Does it not arise over again as soon as we reflect upon the implications of the word ‘in’ itself?

If it is our profession to be misled – as, of course, it *is* the profession of philosophers – we shall be liable to suppose that redness is in the tomato somewhat as juice is in it, or as a weevil is in it. And if so, what can be meant by saying that redness is recurrent? How can it be *in* thousands of other tomatoes as well, or hundreds of post-boxes, or dozens of blushing faces? It does not make sense to say that a weevil is in many places at once. Again, when the tomato begins to decay and turns brown, where has the redness gone to, which used to be in it? (The weevil has gone somewhere else; you will find him in the potato basket.) Likewise, where has the brownness come from?

If we prefer to say that the tomato *has* redness, rather than ‘redness is in it’, we shall again mislead these literal-minded persons, and in the same sort of way. Does the tomato *have* redness as Jones *has* a watch? If so, how can millions of other things have it too?

I confess that I do not think much of these difficulties. The meaning of ‘in’ and ‘have’ in this context can be quite easily exhibited by examples, just as their literal meaning can, when we say that there is a weevil in the tomato, or I have a watch. Surely we all know quite well what is being referred to when two things are said to *have* the same colour? And is it really so very difficult to recognize what is meant by saying that the same colour is *in* them both? It is true no doubt that the words ‘in’ and ‘have’ are here being used in a metaphorical sense, though not, I think, extravagantly metaphorical. But we must use metaphorical words, or else invent new and technical terms (which are themselves usually metaphorical words taken from a dead language, Greek or Latin). Our ordinary language exists for practical ends, and it has to be ‘stretched’ in one way or other if we are to use it for purposes of philosophical analysis. And if our metaphors can be cashed quite easily by examples, as these can, no harm whatever is done.

It could, however, be argued that the terminology of ‘characteristics’, which was current in the last philosophical epoch, some twenty years ago, is better than the more ancient terminology of ‘universals’. A characteristic is pretty ob-

viously a characteristic *of* something or other, and cannot easily be supposed to be an independent entity, like the weevil. Nor can we be easily misled into supposing that when something ‘has’ a characteristic, i.e. is characterized by it, this is at all analogous to the having of a watch. . . .

Henceforth, the Aristotelian theory of *universalia in rebus* will be called ‘the Philosophy of Universals’ for short. If our argument so far has been correct, the Philosophy of Universals is drawing our attention to certain important facts about the world. Yet it is at the same time proposing an analysis of those facts. We cannot dispute the facts, nor can we dispute their fundamental importance. We cannot deny that something which may be called ‘the recurrence of characteristics’ is genuinely there. We must also admit that if it were not there, conceptual cognition could not exist. If the world were not like this, if there were no recurrence in it, it could be neither thought about nor spoken about. We could never have acquired any concepts; and even if we had them innately (without needing to acquire them) they could never have been applied to anything.

But though we cannot dispute the facts, nor their importance, we may, nevertheless, have doubts about the analysis of them which the Philosophy of Universals proposes. At any rate, another and quite different analysis of them appears to be possible. It is the analysis offered by what one may call the Philosophy of Ultimate Resemblances. (Henceforth I shall call this ‘the Philosophy of Resemblances’ for short.) This is the analysis which most contemporary philosophers accept, so far as they consider the *ontological* side of the Problem of Universals at all. It is also accepted by Conceptualists, like Locke. The Philosophy of Resemblances is more complicated than the Philosophy of Universals, and more difficult to formulate. It involves one in long and cumbrous circumlocutions. Yet it claims, not unplausibly, that it keeps closer to the facts which have to be analysed. The unkind way of putting this, the one its critics prefer, is to say that it is ‘more naturalistic’. Let us now consider the Philosophy of Resemblances in more detail.

When we say that a characteristic, e.g. whiteness, *recurs*, that it presents itself over and over again, that it characterizes ever so many numerically different objects, what we say is admittedly in some sense true. But would it not be clearer, and closer to the facts, if we said that all these objects resemble each other in a certain way? Is not this the rock-bottom fact to which the Philosophy of Universals is drawing our attention, when it uses this rather inflated language of ‘recurrent characteristics’? The Philosophy of Universals of course agrees that all the objects characterized by whiteness do resemble one another. But according to it, resemblance is always derivative, and is just a *consequence* of the fact that the very same characteristic – whiteness, in this case – characterizes all these objects. To use more traditional language, it says that when A resembles B, this is *because* they are both instances of the same universal.

Now this is all very well where the resemblance is exact, but what are we to say when it is not? Let us consider the following series of examples: a patch of freshly fallen snow; a bit of chalk; a piece of paper which has been used for wrapping the meat in; the handkerchief with which I have been dusting a rather

dirty mantelpiece; a full evening dress bow-tie which has been left lying about for several years on the floor. All these, we say, are white objects. But are they exactly alike in their colour, if white may be counted as a colour for the purpose of this discussion? Clearly they are not. They are, of course, more or less alike. In fact there is a very considerable degree of colour-likeness between them. But certainly they are not exactly alike in colour. And yet if the very same characteristic, whiteness, is present in them all (as the Philosophy of Universals, apparently, says it is) ought it not to follow that they are exactly alike in colour?

To make quite clear what the point at issue is, we shall have to distinguish, rather pedantically perhaps, between *exact* resemblance in this or that respect and *total* or *complete* resemblance. To put it in another way, resemblance has two dimensions of variation. It may vary in intensity; it may also vary in extent. For example, a piece of writing paper and an envelope, before one has written on either of them, may be exactly alike in colour, and perhaps also in texture. These likenesses between them have the maximum degree of intensity. But the two objects are not completely or totally alike. For one thing, they are unlike in shape. Moreover, the envelope is stuck together with gum and has gum on its flap, while the piece of writing paper has no gum on it. It might perhaps be thought that two envelopes from the same batch *are* completely alike; and certainly they come nearer to it than the envelope and the piece of notepaper. All the same, there is unlikeness in respect of place. At any given time, envelope A is in one place and envelope B is in a different place. On the Relational Theory of Space, this is equivalent to saying that at any given time A and B are related in unlike ways to something else, e.g. the North Pole, or Greenwich Observatory.

According to Leibniz's Principle of the Identity of Indiscernibles, complete or total likeness is an ideal limit which can never quite be reached, though some pairs of objects (the two envelopes, for example) come closer to it than others. For if *per impossible* two objects were completely alike, place and date included, there would no longer be two objects, but only one. Whether Leibniz's Principle is correct, has been much disputed. But we need not concern ourselves with this dispute. It is sufficient to point out that if there were two objects which resembled each other completely, in date and place as well as in all other ways, and this complete resemblance continued throughout the whole of the histories of both, there could not possibly be any evidence for believing there were two of them. So in this discussion we need not concern ourselves any more with complete or total resemblance, though it is of course an important fact about resemblances that they vary in extent, as well as in degree of intensity.

What does concern us is intensity of resemblance. The maximum intensity of it is what I called 'exact resemblance in this or that respect'. Now some people appear to think that even this is an ideal limit. They seem to think that no two objects are ever *exactly* alike even in one way (e.g. colour, or shape) though, of course, many objects are closely alike in one way or in several. I do not see what evidence we could have for believing such a sweeping negative generalization. It is true that sometimes, when we thought at first that there was an exact likeness in one or more respects between two objects, we may find on more

careful examination that there was not. We may have thought that two twins were exactly alike in the conformation of their faces. We look more closely, and find that John's nose is a millimetre longer than William's. But still, there are many cases where there is no discoverable inexactness in a resemblance. We often find that two pennies are indistinguishable in shape, or two postage stamps indistinguishable in colour. And we should not confine ourselves to cases where two or more objects are being compared. There is such a thing as monotony or uniformity within a single object. For example, a certain patch of sky is blue, and the same shade of blue, all over. It is monotonously ultramarine. In other words, all its discernible parts are exactly like each other in colour; at any rate, we can discover no unlikeness of colour between them. Again, there is often no discoverable unlikeness of pitch between two successive phases of the same sound. Will it be said that such monotony is only apparent, not real? But what ground could we have for thinking that no entity is ever really 'monotonous' in this sense, not even in the smallest part of its extent, or throughout the smallest phase of its duration? Thus there is no good ground for maintaining that resemblance of maximum intensity never occurs at all, still less for maintaining that it never *could* occur. Nevertheless, it is not so very common for two objects to be exactly alike even in one way, though monotony within a single object or event is more frequent. What we most usually find in two or more objects which are said to be 'alike' is *close* resemblance in one respect or in several.

We can now return to the controversy between the Philosophy of Resemblances and the Philosophy of Universals. It is argued that if the Philosophy of Universals were right, exact resemblance in one or several respects (resemblance of maximum intensity) ought to be much more common than it is; indeed, that *inexact* resemblance in a given respect, say colour or shape, ought not to exist at all. Of course, there could still be incomplete or partial resemblance, resemblance between two objects in one respect or in several, and lack of resemblance in others. But whenever two objects do resemble each other in a certain respect, it would appear that the resemblance ought to be exact (of maximum intensity), if the Philosophy of Universals were right; either it should be exact, or else it should not exist at all. The Philosophy of Universals tells us that resemblance is derivative, not ultimate; that when two objects resemble each other in a given respect, it is because the very same universal is present in them both. This seems to leave no room for *inexact* resemblance.

Now if we consider the various white objects I mentioned before – the whole series of them, from the freshly fallen snow to the unwashed bow-tie – how can anyone maintain that the very same characteristic, whiteness, recurs in all of them? Clearly it does not. If it did, they must be exactly alike in their colour; and quite certainly they are not. If we are to use the language of universals or characteristics, shall we not have to say that each of the objects in this series, from the snow to the unwashed tie, is characterized by a *different* characteristic, or is an instance of a *different* universal? In this case, then, the resemblance seems to be ultimate and underivative, *not* dependent on the presence of a single universal in all these objects, although they certainly do resemble each other.

Let us consider another example. Two pennies may be exactly alike in their shape. If so, one may plausibly say that the very same characteristic, roundness, is present in both of them, and that their resemblance is dependent on this. But what about a penny and a sixpence? They certainly *are* alike in shape; but not exactly, because the sixpence has a milled edge and the penny a smooth one. So here again, it would seem, there is no *single* characteristic present in them both, upon which the resemblance could be dependent. This resemblance again seems to be ultimate and underivative.

Thus the Philosophy of Universals, when it makes all resemblance derivative, appears to forget that resemblances have degrees of intensity. Resemblance is treated as if it were degreeless, either present in its maximum degree or else not present at all. In practice, the Philosopher of Universals concentrates his attention on *close* resemblances, and averts his attention from the awkward circumstance that few of them are exact; and resemblances of a lower degree than this (small or moderate ones, not intense enough to be called ‘close’) are just neglected altogether. But is it not a glaringly obvious fact that resemblances do differ in degree or intensity?

That being so, shall we not be inclined to *reverse* this alleged dependence-relation between ‘being alike’ and ‘being characterized by’? Surely we shall be inclined to say that it is resemblance which is more fundamental than characterization, rather than the other way round. We shall, of course, be willing to go on using terms like ‘characteristic’ and ‘characterized by’; they are part of ordinary language, and everyone has a sufficient understanding of them. But we shall define ‘characteristic’ in terms of resemblance, and not conversely. Where a number of objects do happen to resemble each other exactly in one respect or three or fifteen, there, and in consequence, we shall be quite willing to say that they have one, or three, or fifteen ‘characteristics in common’. But in other cases, where the resemblance is less than exact, we shall not be willing to say this. We shall just say that they resemble each other in such and such a degree, and stop there. In a given set of objects there is whatever degree of resemblance there is. Let us be content to take the facts as we find them.

Turning for a moment to the epistemological side of the matter, surely it is obvious that the applicability of concepts does *not* require an exact resemblance in the objects which a concept applies to? Of course there does have to be a considerable degree of resemblance between all the objects which ‘satisfy’ a given concept. As we say, there has to be a sufficient likeness between them, e.g. between all the objects to which the concept White applies. What degree of likeness is sufficient, and where the borderline comes between something which falls just within the concept’s sphere of application and something else which just falls outside it, is often difficult to decide. For instance, one may wonder whether the *very* dirty bow-tie is white at all. Indeed, it is difficult to see how such a question can be definitely answered, at least in the case of whiteness and many other familiar concepts. The right way to tackle it, perhaps, is to refuse to answer it as it stands. Perhaps we should rather say that a concept may be ‘satisfied’ in many different degrees; or, in more commonsensical language, that there are good instances and bad instances, better and worse ones, and some so

bad that it is arbitrary whether one counts them as instances or not. Thus the piece of chalk is a *better* instance of whiteness than the rather dirty handkerchief is. The patch of freshly fallen snow is a better instance still, perhaps a perfect one. We may give it the mark a (+). Then $a\beta$ is about the right mark for the piece of chalk, and we will give the unwashed bow-tie γ =, to denote that it is just on the borderline between ‘pass’ and ‘failure.’

It is not easy to see how the doctrine of *universalia in rebus* can make any room for this important and familiar notion of degrees of instantiation. But there is plenty of room for it in Conceptualism, which is the epistemological counterpart of the ontological Philosophy of Resemblances. We must add, in fairness, that there is also plenty of room for it in the Platonic doctrine of *universalia ante rem*. Indeed Plato, or perhaps Socrates, was the first philosopher who noticed that there are degrees of instantiation. This is one of the points, and a good point, which Conceptualism and Platonic Realism have in common.¹

In the last few pages, I have been discussing the difficulties which the Resemblance Philosophers find in the Philosophy of Universals. But the Philosophy of Resemblances has its difficulties too. The most important ones are concerned with resemblance itself. I shall discuss two of them, and the solutions proposed for them. The first arises from the phrase ‘resemblance in respect of . . .’.

It is obvious that we must distinguish between *different* resemblances. Objects resemble each other in different respects, as well as in different degrees. Red objects resemble each other in one respect, round objects in another respect. The members of a natural kind, for instance cats or oak trees, resemble each other in many respects at once. Thus it would be much too vague if we said that red objects, for example, are just a set of objects which resemble one another, or sufficiently resemble each other. That would not distinguish them from blue objects, or round objects, or any other class of objects one cares to name. We must specify what resemblance it is. Red objects are those which resemble each other ‘in a certain respect’. But in *what* respect? And now it looks as if we should have to introduce universals again. Our first answer would probably be that they resemble each other in respect of colour; and this looks very like saying that they are all instances of the universal Colouredness. That is bad enough; but we shall be driven to go farther, because we have not yet said enough to distinguish red objects from blue ones or green ones. Can we stop short of saying that red objects are just those objects which resemble each other in respect of *redness*? And here we seem to be admitting the very point which the Philosophy of Universals is so anxious to maintain; namely that the resemblance between these objects is after all derivative, dependent upon the presence of a single universal, Redness, in them all. To generalize the argument: whenever we say that A, B and C resemble each other in a certain respect, we shall be asked ‘In *what* respect?’ And how can we answer, except by saying ‘in respect of being instances of the universal φ ’ or ‘in respect of being characterized by the characteristic φ ’? We may try to get round the difficulty by saying that they resemble each other in a certain *way* (avoiding the word ‘respect’), or

that there is a certain *sort* of resemblance between them. But when we are asked to specify in *what* way they resemble each other, or what sort of resemblance there is between them, surely we shall still have to answer by mentioning such and such a universal or characteristic? 'The way in which red objects resemble each other is that all of them are instances of the universal Redness, or all of them are characterized by the characteristic Redness.'

This is one of the classical objections to the Philosophy of Resemblances. The argument purports to show that resemblance is not after all ultimate or underivative, but is dependent on the presence of a universal or characteristic which is common to the things which resemble each other. There is something about this objection which arouses our suspicions. It comes perilously near to the tautology 'red things are the things which are red'. The Resemblance philosophers were not undertaking to deny this tautology. They do not deny that *x* is red entails *x* is red. They are only concerned to offer an analysis of *x* is red itself.

Let us now consider the answer they might make to this celebrated objection. Roughly, it consists in substituting 'resemblance *towards* . . .' for 'resemblance in respect of . . .'. Resemblance towards what? Towards certain standard objects, or *exemplars* as I shall call them – certain standard red objects, or standard round objects, or whatever it may be.

It is agreed by both parties that there is a *class* of red objects. The question is, what sort of a structure does a class have? That is where the two philosophies differ. According to the Philosophy of Universals, a class is so to speak a promiscuous or equalitarian assemblage. All its members have, as it were, the same status in it. All of them are instances of the same universal, and no more can be said. But in the Philosophy of Resemblances a class has a more complex structure than this; not equalitarian, but aristocratic. Every class has, as it were, a nucleus, an inner ring of key-members, consisting of a small group of standard objects or exemplars. The exemplars for the class of red things might be a certain tomato, a certain brick and a certain British post-box. Let us call them A, B and C for short. Then a red object is any object which resembles A, B and C as closely as they resemble one another. The resemblance between the exemplars need not itself be a very close one, though it is of course pretty close in the example just given. What is required is only that every other member of the class should resemble the class-exemplars *as* closely as they resemble one another. Thus the exemplars for a class might be a summer sky, a lemon, a post-box, and a lawn. These do resemble one another, though not very closely. Accordingly there is a class consisting of everything which resembles these four entities *as* closely as they resemble each other. It is the class of coloured things, whereas the previous one was the class of red things.

It may be thought that there is still a difficulty about the resemblance between the exemplar objects themselves. In *what respect* do the tomato, the brick and the post-box resemble each other? Surely this question still arises, even though it does not arise about the other members of the class? And how can one answer it, except by saying that these three objects resemble each other in respect of being red, or of being characterized by redness?

But this assumes that we know beforehand what ‘being red’ is, or what ‘being characterized by redness’ amounts to. And this begs the question against the Resemblance Philosophy. The Resemblance Philosophers maintain that our knowledge of what it is for something to be red just consists in a capacity to compare any particular object X with certain standard objects, and thereby to discover whether X does or does not resemble these standard objects as closely as they resemble each other. It does not make sense to speak of comparing the standard objects *with themselves*, or to ask whether *they* resemble one another as closely as they do resemble one another. Yet that is just what we should be trying to do, if we tried to say ‘in what respect’ they are alike. To say that *they* are red, or are characterized by redness, would not be an informative statement, but a tautology.

This objection does however draw our attention to an important point. According to the Philosophy of Resemblances, there cannot be a class unless there are exemplar objects to hold the class together. Nevertheless, the same class may have *alternative* sets of exemplars. The class of red things, we said, consists of everything which resembles the post-box, the tomato and the brick as closely as they resemble each other. It could equally be said to consist of everything which resembles a certain bit of sealing wax, a certain blushing face and a certain sunset sky as closely as *they* resemble each other. In that case, it does make sense to ask whether the post-box, the tomato and the brick are red, or are characterized by redness. And the answer ‘Yes, they are’ is now no longer tautologous. We are no longer trying, absurdly, to compare them with themselves. We are comparing them with three other things, and discovering that they do all have a sufficient degree of resemblance to these other things. But because there are (within limits) alternative sets of standard objects for the same class, we are led to suppose, erroneously, that a class can exist without any standard objects at all. This or that set of standard objects can be deposed from its privileged position without destroying the unity of the class; and we then suppose, by a process of illegitimate generalization, that the class would still remain what it is if privilege were abolished altogether. There must be *a* set of standard objects for each class, though within limits it does not matter which set of objects have this status.

Thus in the Philosophy of Resemblances, as well as the Philosophy of Universals, there does after all have to be something which holds a class together, if one may put it so. Where the two philosophies differ is, in their view of what that something is. In the Philosophy of Universals, what holds a class together is a universal, something of a different ontological type from the members. In the Philosophy of Resemblances there is no question of different ontological types. There are just particular objects, and there is nothing non-particular which is ‘in’ them, in the way that a universal is supposed to be ‘in’ the particulars which are its instances. What holds the class together is a set of nuclear or standard members. Anything which has a sufficient degree of resemblance to these is thereby a member of the class; and ‘resembling them sufficiently’ means ‘resembling them as closely as they resemble each other’.

Again, to turn for a moment to epistemological considerations, it is their

relationship to the standard objects or exemplars which enables all these objects to satisfy the same concept, e.g. the concept Red, and likewise enables the same word or other symbol, e.g. the word 'red', to apply to them all. But this is to anticipate. The Philosophy of Resemblances is an *ontological* doctrine, though it may be used as the starting point for certain epistemological theories (Conceptualism, Imagism and Nominalism), just as the Philosophy of Universals may be used as the starting point of a Realist epistemology. If the Philosophy of Resemblances is true at all, it might still have been true even if there had been no thinkers and no speakers. As it happens, there are thinkers and speakers too. But there may be many classes in the world, which do exist (because the requisite resemblances do happen to be there) although no mind happens to have formed the corresponding class-concepts, and no speaker has acquired the habit of using the corresponding class-symbols. Thus there is nothing subjectivist or anthropocentric about the Philosophy of Resemblances. It denies that there are universals *in rebus*, but it asserts that there are resemblances *inter res*. Certain objects really are as like the objects A, B and C as these are to one another, whether anyone notices the fact or not. Known or not, spoken of or not, the relationship is there; just as in the Philosophy of Universals objects are instances of universals whether they are known to be so or not. In this respect, both these philosophies are equally 'realistic'.

We must now turn to the second of the classical objections against the Philosophy of Resemblances, an objection so familiar that one might almost call it notorious. It is concerned with resemblance itself. Surely resemblance is itself a universal, present in many pairs or groups of ressemblant objects? It is of course, a universal of relation. The instances of it are not individual objects taken singly, but complexes, and each of these complexes is composed of two objects or more. In their attempt to 'get rid of universals', the Philosophers of Resemblance seem to concentrate their attention on universals of *quality* (e.g. redness, colour, shape) and say little or nothing about universals of relation. Hence they have failed to notice that resemblance itself is one of them. But if we are obliged to admit that resemblance at any rate is a genuine universal, a relation which does literally recur in many different situations or complexes, what ground have we for denying that there are other *universalia in rebus* as well?

It may seem audacious to question this formidable argument, which has convinced many illustrious men. But is it as strong as it looks? The Resemblance philosophers might very well reply that it begs the question at issue, that it just assumes what it purports to prove. For after all, what reason is given for the crucial assertion that resemblance is a universal? Apparently none. It is not enough just to say 'surely resemblance at any rate is a universal'. Could any reason be given? We might perhaps try to find one by starting from the linguistic side of the matter. The word 'resemblance', we might say, is an *abstract* word, like the words 'redness' and 'proximity'; therefore it must stand for a universal or characteristic (a relational one, of course). But if this is the argument, it seems to beg the question. For if one does start from a linguistic point of view, the very question at issue is whether abstract words and general words do stand for

universals. And if the argument is to be cogent, it ought to be an argument about the noun ‘resemblance’ in particular, or about the verb ‘to resemble’ in particular. We ought to be shown that it is somehow peculiarly obvious that *this* word at any rate (or this pair of words) stands for a universal, even though it may be less obvious that other general words do.

Perhaps it will be said, the peculiar obviousness consists in this, that even the people who try to get rid of universals have to use *this* general word at least, or equivalent general words such as ‘similar’, ‘like’. True enough, one cannot speak in a language consisting entirely of proper names and demonstratives. One cannot say anything at all without using some general words. As an observation about the nature of language, this is perfectly indisputable. But the question is, what are its implications? Does it follow that because we must use general words, there are, therefore, general somethings *in rerum natura* which they stand for? That is just the point at issue. One cannot just assume that the answer is ‘Yes.’ Of course the Philosophy of Resemblances admits that we do use general words, and cannot avoid using them if we are to speak at all. It does not at all deny the fact. But it does deny the conclusion which the Philosophy of Universals draws from it – namely that because we use general words, there must be *general somethings* (universals) which they mean. Has anything been done to show that this denial is mistaken? Nothing. The Philosophy of Universals has just repeated over again the principle which has to be proved, the principle that every general word stands for a universal; adding – what is obvious – that *if* this principle is true, the word ‘resemblance’ is an illustration of it. Of course. But *is* the principle true?

If the Philosopher of Resemblances is asked to explain how the general word ‘resemblance’ is used, or what kind of meaning it has, he will presumably point out that there are resemblances of *different orders*. Two cats, A and B, resemble each other, and two sounds, C and D, also resemble each other. These are first-order resemblances. But it is also true that the two-cat situation resembles the two-sound situation, and resembles many other situations too. This is a second-order resemblance. The A-B situation and the C-D situation really are alike, though the constituents of the one are unlike the constituents of the other. In virtue of this second-order likeness (a likeness *between* likeness-situations) we may apply the same general word to both of them; and the word we happen to use for this purpose is the word ‘resemblance’, in a second-order sense. There is nothing wrong or unintelligible in the notion of second-order resemblance. Or if it be said that there is, we can reply with the *tu quoque* argument that universality must itself be a universal. When it is said that ‘cathood is a universal’ the word ‘universal’ is itself a general word, just as ‘cat’ is when we say ‘Pussy is a cat.’ So according to the to the Philosophy of Universals, there must be a universal called ‘universality’. And if it is a universal, universality must accordingly be an instance of itself. But this is a contradiction. For according to this Philosophy, anything which is an *instance* of a universal is *ipso facto* a particular, and not a universal. To get out of this difficulty, the Philosophy of Universals must introduce the notion of ‘different orders’ too. The word ‘universal’, it has to say, stands for a second-order universal, whereas ‘green’ or ‘cat’ or ‘in’ stand

for first-order ones. This is equivalent to saying that the expression ‘a universal’, or the propositional function ‘ ϕ is a universal’, can occur only in a metalanguage.

This suggests another way in which the Philosophy of Resemblances might reply to the objection that ‘resemblance is itself a universal’. The objection assumes that resemblance is just one relation among others: a relation of the same type as ‘on top of’, or ‘near to’, or ‘side by side with’. But according to the Philosophy of Resemblances, resemblance is not just one relation among others. Indeed, according to this philosophy, it would be misleading to call it ‘a relation’ at all. It is too fundamental to be called so. For what we *ordinarily* call ‘relations’ (as well as what we call ‘qualities’) are themselves founded upon or analysable into resemblances. For example, the relation ‘being inside of’ is founded upon the resemblance between the Jonah–whale complex, the room–house complex, the match–matchbox complex, etc. Moreover, the Philosophy of Universals itself does not really hold that resemblance is just one relation among others, and in pretending that it does, it is abandoning one of its own fundamental principles; indeed it is abandoning the very one which this argument (‘resemblance is itself a universal’) is ultimately intended to establish, the principle, namely, that all resemblance is derivative. In the Philosophy of Universals itself, resemblance has a status quite different from relations like ‘side by side with’ or ‘on top of’. Resembling is connected with *being an instance of* . . . in a way that ordinary relations are not. When A resembles B and C, this is supposed to be a direct consequence of the fact that A, B and C are all instances of the same universal; and not only when A, B and C are individual objects (in which case the universal is a universal of quality) but also when they are complexes, so that the universal they are instances of is a relational one, such as ‘being inside of’. If resemblance, in the Philosophy of Universals, is to be called a relation at all, it is a relation of a very special sort, quite different from anything to which the word ‘relation’ is *ordinarily* applied. We should have to say that it is a ‘formal’ or ‘metaphysical’ relation (as opposed to a ‘natural’ or empirical one) just as the relation of instantiation is, if that can be called a relation at all.

So much for the reply the Philosophy of Resemblances might make to this celebrated argument that ‘resemblance is itself a universal’. First, it might be objected that the argument begs the question, by just assuming (what it ought to prove) that because ‘resemblance’ is admittedly a general word, it must stand for a universal. Secondly, the argument overlooks the fact that there are resemblances of different orders. Thirdly, it treats resemblance as one relation among others, parallel in principle to ‘side by side with’ or ‘on top of’, whereas the Philosophy of Resemblances maintains that it is too fundamental to be called a relation at all, in the ordinary sense of the word ‘relation’. Fourthly, the Philosophy of Universals itself admits, in its own way, that resemblance does *not* have the same status as other relations, in spite of maintaining in this argument that it has.

Thus the Philosophy of Resemblances has an answer to these two classical objections, the one about ‘resemblance in respect of’ and the one we have just

discussed 'that resemblance is itself a universal'. But the Philosophy of Universals also has an answer to the objection about inexact resemblances, and to the complaint that it ignores the different degrees of intensity which resemblances may have. We must consider this answer if we are to do justice to both parties.

The first step is to distinguish between *determinable* and *determinate* characteristics. Universals or characteristics, it is said, have different degrees of determinateness. The adjectives 'determinable' and 'determinate' are too fundamental to be defined. But their meaning can be illustrated. Thus the characteristic of being coloured is a determinable, and the characteristic of being red is a determinate of it. Being red is again a sub-determinable, and has under it the determinates being scarlet, being brick-red, being cherry-red, etc. Likewise, being a mammal is a determinable characteristic, a highly complex one this time. There are many different ways of being a mammal. Being a dog, being a whale, being a man are some of the determinates of this determinable.

Now whenever two objects resemble each other with less than the maximum intensity (i.e. whenever they have what was called an 'inexact' resemblance) we can always say that the same *determinable* characteristic characterizes them both, though not the same determinate one. Two objects may each have a different shade of red. A is scarlet, and B is brick-red. They resemble each other fairly closely, but by no means exactly. That is because redness itself is a determinable characteristic, a sub-determinable falling under the higher determinable colouredness. The two objects do have this determinable characteristic in common, though each of them has a different determinate form of it. So we can still maintain that this resemblance, though inexact, is a derivative, dependent on the presence of the same determinable universal in both objects.

Let us apply these considerations to the two examples given earlier: (1) the various white objects; (2) the penny and the sixpence. It may now be maintained that all my different white objects – from the freshly fallen snow at one end of the series to the unwashed bow-tie at the other – do have a *determinable* characteristic in common; though 'whitish', rather than 'white', would be the appropriate word for it. 'White' might be taken to mean pure white. And pure white is only one determinate of the determinable *whitish*. We certainly should not say that all the objects in this series are pure white. At the most, only the freshly fallen snow is pure white, but not the piece of chalk, or the rather messy bit of paper, or the rather dirty handkerchief, or the very dirty unwashed bow-tie. But we should admit that all of them are 'whitish'.

Let us now consider my other example, the penny and the sixpence, which resemble each other in shape, but inexactly. The penny with its smooth edge and the sixpence with its milled (slightly serrated) edge have different determinate shapes. How is it, then, that they do still resemble each other in shape, though inexactly, and both would be called 'round coins' in ordinary speech? Because the same *determinable* shape – we might more appropriately call it 'roundish' – characterizes both of them; and it characterizes many other things as well, e.g. slightly buckled bicycle wheels, cogwheels with not too large teeth, which resemble each other a good deal less closely than the penny and the sixpence do.

By this expedient the Philosophy of Universals is able to maintain its thesis that all resemblances, inexact ones too, are derivative, and not ultimate, as the Philosophy of Resemblances would have them. Inexact resemblance, we are invited to say, depends upon or is derived from the presence of the same *determinable* characteristic in a number of objects; exact resemblance (resemblance of maximum intensity) depends upon their being characterized by the same *determinate* characteristic.

Perhaps this will also enable us to dispense with the notion of ‘degrees of instantiation’ which was mentioned earlier. It was not easy to see what could be meant in the Philosophy of *universalia in rebus* by saying that one object is a *better* instance of so-and-so than another, though this notion fits well enough into the Platonic theory of *universalia ante rem*, and into Conceptualism too. Perhaps it could now be suggested that the determinates of some determinables, e.g. ‘whitish’, ‘roundish’, are serially ordered. Thus the various determinates of whitishness which characterize the patch of snow, the piece of chalk, the paper, etc., may be arranged in a series beginning with ‘pure white’. After this comes ‘nearly pure white’ (the colour the piece of chalk has), then ‘farther from pure white’ and then ‘farther still from pure white’, until we come to a characteristic which is as far from pure whiteness as it can be without ceasing to be a determinate of whitishness at all. The system of marking ($a+$, a , $a-$, $\beta+$, etc.) which we suggested for indicating the ‘goodness’ or ‘badness’ of instances can still be used: only it is now applied not to the objects themselves, but to the determinate characteristics by which they are respectively characterized.

Thus this objection to the Philosophy of Universals, that it can make no room for inexact resemblances (resemblances of less than the maximum intensity), turns out after all to be indecisive, although it looked so convincing at first sight. The facts to which this argument draws our attention are of course perfectly genuine, and important too. It is, for example, an important fact about language that most of our general words apply to sets of objects which inexactly resemble one another; and it is an important fact about thinking, that the various objects which ‘satisfy’ a given concept, e.g. the concept of Crow, do not have to be exactly alike. Nevertheless, this argument does not at all refute the Philosophy of Universals, as it is often supposed to do. All it does is to point out what was lacking in our first rough-and-ready formulation of that philosophy. Certainly the Philosophy of Universals would be quite unworkable *without* the distinction between determinable and determinate universals. The doctrine that universals or characteristics have different degrees of determinacy is an indispensable part of it. But the distinction between determinables and determinates is perfectly consistent with the contention that there are recurrent characteristics in the world, and with the accompanying doctrine that resemblances are derivative, not ultimate. Indeed, it could be argued, the fact that recurrent characteristics do differ in their degree of determinateness is just as obvious as the fact of recurrence itself.

Finally, it is worth repeating that the phrases ‘inexact resemblance’, ‘not exactly alike’ are sometimes used in another way, to mean *incomplete* or *partial* resemblance. If A and B are closely alike in a large number of respects, but

unlike or not closely alike in one or two, we sometimes say that they are very like each other but not exactly like each other. For example, within the same species of bird we often find that there are slight differences of size or colouring between two individual specimens, although they also resemble each other closely in very many ways. It is obvious that if the phrase ‘inexact resemblance’ is used in *this* sense, the Philosophy of Universals has no difficulty at all about inexact resemblances. We merely have to say that many universals are common to the two birds, or recur in both of them; and consequently the two individuals resemble each other in a great many respects. We then add that bird A is also an instance of a certain universal φ , while bird B is not an instance of this, but of a certain other universal ψ ; and consequently there is a respect in which they are *not* alike. (It may be found, of course, and in this example it almost certainly will be, that though φ and ψ are different determinate universals, they are determinates of the same determinable universal, say ‘mottled’.) It must not be forgotten that every individual object is an instance of several universals at once, and often of very many at once. When we compare it with another object, we may easily find that some universals are common to both of them, and other universals are not. It would be a strange misunderstanding of the Philosophy of Universals to suppose that in this philosophy every particular is held to be an instance of only *one* universal. When we say that something is a cat, we are saying that it is an instance of many universals conjointly, and not just of one.

Our discussion has been long and complicated. What conclusion shall we draw from it? It would seem that there is nothing to choose between these two philosophies, the Philosophy of Universals or characteristics (*universalia in rebus*)² on the one hand, and the Philosophy of Ultimate Resemblances on the other. At any rate, it would seem that there is nothing to choose between them so long as they are considered as purely ontological doctrines, which is the way we have been considering them in this chapter. Both seem to cover the facts, though only when both are stated with sufficient care. Moreover, they both cover the *same* facts. This strongly suggests that they are two different (systematically different) terminologies, two systematically different ways of saying the same thing. It does not follow that both alike are just pieces of solemn and elaborate trifling. On the contrary, the thing which they both say is of the first importance, and we do need a way of saying it. The efforts which each party has made to provide us with a systematic terminology for saying it have not been a waste of time. For if there were no recurrent characteristics, *or* no resemblances between different objects – whichever way you choose to put it – there could be no conceptual cognition, and no use of general symbols either.

Now if there is only a (systematic) difference of terminology between these two philosophies, it is well to be familiar with both. Each of them may have its misleading features; and when we are in danger of being misled by the one, we may save ourselves by changing over to the other. . . .

Notes

- 1 In Christian Platonism, where Plato's transcendent 'forms' become concepts in the mind of God, the differences between Platonic Realism and Conceptualism are still further diminished, though they do not disappear altogether.
 - 2 It may be worthwhile to remind the reader that the phrase 'the Philosophy of Universals', as it has been used in this chapter, is *not* intended to cover the Platonic doctrine of *universalia ante rem*.
-

2 The Elements of Being*

D. C. Williams

First philosophy, according to the traditional schedule, is analytic ontology, examining the traits necessary to whatever is, in this or any other possible world. Its cardinal problem is that of substance and attribute, or at any rate something cognate with this in that family of ideas which contains also subsistence and inherence, subject and predicate, particular and universal, singular and general, individual and class, and matter and form. It is the question how a thing can be an instance of many properties while a property may inhere in many instances, the question how everything is a case of a kind, a this-such, an essence endowed with existence, an existent differentiated by essence, and so forth. Concerned with what it means to be a thing or a kind at all, it is in some wise prior to and independent of the other great branch of metaphysics, speculative cosmology: what kinds of things are there, what stuff are they made of, how are they strung together?

Although "analytic ontology" is not much practiced as a unit under that name today, its problems, and especially the problem of subsistence and inherence, are as much alive in the latest manifestoes of the logical analysts, who pretend to believe neither in substances nor in universals, as they were in the counsels of Athens and of Paris. Nothing is clear until that topic is clear, and in this essay I hope to do something to clarify it in terms of a theory or schema which over a good many years I have found so serviceable that it may well be true.

Metaphysics is the thoroughly empirical science. Every item of experience must be evidence for or against any hypothesis of speculative cosmology, and every experienced object must be an exemplar and test case for the categories of analytic ontology. Technically, therefore, one example ought for our present theme to be as good as another. The more dignified examples, however,

* From "On the Elements of Being," originally published in the *Review of Metaphysics*, vol. 7 (1953), reprinted with permission.

darkened with a patina of tradition and partisanship, while some frivolous ones are peculiarly perspicuous. Let us therefore imagine three lollipops, made by a candy man who buys sticks from a big supplier and molds candy knobs on them. Lollipop No. 1 has a red, round, peppermint head, No. 2 a brown, round, chocolate head, No. 3 a red, square, peppermint head. The circumstance here which mainly provokes theories of subsistence and inherence is similarity with difference: each lollipop is partially similar to each other and partially different from it. If we can give a good account of this circumstance in this affair we shall have the instrument to expose the anatomy of everything, from an electron or an apple to archangels and the World All.

My chief proposal to that end may be put, to begin with, as nothing more tremendous than that we admit literally and seriously that to say that *a* is partially similar to *b* is to say that a part of *a* is wholly or completely similar to a part of *b*. This is a truism when we construe it with respect to ordinary concrete parts, for example, the sticks in the lollipops. On physical grounds, to be sure, it is not likely that any three solid objects, not even three sticks turned out by mass industry, are exactly similar, but they often look as if they were, and we can intelligibly stipulate for our argument that our exemplary sticks do exactly resemble each other through and through. To say then that each of the lollipops is partially similar to each other, that is, with respect to stick, is to say that there is a stick in each which is perfectly similar to the stick in every other, even though each stick remains as particular and distinct an individual as the whole lollipop. We would seldom give a proper name to a lollipop and still more seldom to the stick in one, but we might easily do so – “Heraplem” for lollipop No. 1, for example, “Paraplete” for its stick, “Boanerp” for No. 2, and “Merrinel” for its stick. Heraplem and Boanerp then are partially similar because Paraplete and Merrinel are perfectly similar.

What now of the rest of each lollipop and what of their more subtle similarities, of color, shape, and flavor? My proposal is that we treat them in exactly the same way. Since we cannot find more parts of the usual gross sort, like the stick, to be wholly similar from lollipop to lollipop, let us discriminate subtler and thinner or more diffuse parts till we find some of these which *are* wholly similar. This odd-sounding assignment, of course, is no more than we are accustomed to do, easily and without noticing.

Just as we can distinguish in the lollipops, Heraplem and Boanerp, the gross parts called “sticks,” namely, Paraplete and Merrinel, so we can distinguish in each lollipop a finer part which we are used to call its “color” and another called its “shape” – not its kind of color or shape, mind you, but these particular cases, this reddening, this occurrence or occasion of roundness, each as uniquely itself as a man, an earthquake, or a yell. With only a little more hardihood than christened the lollipops and sticks, we can christen our finer components “Harlac” and “Bantic” for the respective color components, let us say, and “Hamis” and “Boras” for the respective shape components.

In these four new names, the first and last letters are initials of “Heraplem” and “Boanerp,” and of “color” and “shape,” respectively, but this is a mnemonic device for us, irrelevant to their force as names. “Harlac,” for example, is

not to be taken as an abbreviation for the description, “the color component of Heraplem.” In a real situation like the one we are imagining, “Harlac” is defined ostensively, as one baptizes a child or introduces a man, present in the flesh; the descriptive phrase is only a scaffolding, a temporary device to bring attention to bear on the particular entity being denoted, as a mother of twins might admonish the vicar, “Boadicea is the cross-looking one.”

Heraplem and Boanerp are partially similar, then, not merely because the respective gross parts Paraplete and Merrinel (their sticks) are wholly similar but also because the respective fine parts, Hamis and Borcas (their “shapes”), are wholly similar – all this without prejudice to the fact that Hamis is numerically as distinct from Borcas, to which it is wholly similar, and from Harlac, with which it is conjoined in Heraplem, as Harlac is from Bantic, to which it is neither similar nor conjoined, and as the stick Paraplete is from the stick Merrinel, and as the whole lollipop, Heraplem, is from the whole Boanerp. The sense in which Heraplem and Boanerp “have the same shape” and in which “the shape of one is identical with the shape of the other” is the sense in which two soldiers “wear the same uniform” or in which a son “has his father’s nose” or our candy man might say “I use the same identical stick, Ledbetter’s Triple-X, in all my lollipops.” They do not have the same shape in the sense in which two children have the same father, or two streets have the same manhole in the middle of their intersection, or two college boys wear the same tuxedo (and so cannot go to dances together). But while similar in the indicated respects, Heraplem and Boanerp are partially dissimilar inasmuch as their knobs or heads are partially dissimilar, and these are partially dissimilar because some of their finer parts, for example, their colors, are dissimilar.

In like manner, to proceed, we note that Harlac, the color component of No. 1 (Heraplem), though numerically distinct from, is wholly similar to, the color component of No. 3. But No. 1 has not only a color component which is perfectly similar to the color component of No. 3; it has also a flavor component perfectly similar to the flavor component of No. 3. (It does not matter whether we think of the flavor as a phenomenal quality or as a molecular structure in the stuff of the candy.) The flavor-plus-color of No. 1 (and likewise of No. 3) is a complex whose own constituents are the flavor and the color, and so on for innumerable selections and combinations of parts, both gross and fine, which are embedded in any one such object or any collection of them.

Crucial here, of course, is the admission of a fine or subtle part, a diffuse or permeating one, such as a resident color or occurrent shape, to at least as good standing among the actual and individual items of the world’s furniture as a gross part, such as a stick. The fact that one part is thus finer and more diffuse than another and that it is more susceptible of similarity no more militates against its individual actuality than the fact that mice are smaller and more numerous than elephants makes them any the less real. To borrow now an old but pretty appropriate term, a gross part, like the stick, is “concrete,” as the whole lollipop is, while a fine or diffuse part, like the color component or shape component, is “abstract.” The color-plus-shape is less abstract or more nearly concrete than

the color alone, but it is more abstract or less concrete than color-plus-shape-plus-flavor, and so on up till we get to the total complex, which is wholly concrete.

I propose now that entities like our fine parts or abstract components are the primary constituents of this or any possible world, the very alphabet of being. They not only are actual but are the only actualities, in just this sense, that whereas entities of all other categories are literally composed of them, they are not in general composed of any other sort of entity. That such a crucial category has no regular name is quite characteristic of first principles and is one reason why the latter are worth pursuing. A description of it in good old phraseology has a paradoxical ring: our thin parts are “abstract particulars.” We shall have occasion to use “parts” for concrete entities and “components” for abstract ones (and “constituent” for both), as some British philosophers have used “component” for property and “constituent” for concrete part. Recalling, however, that Santayana used “trope” to stand for the essence of an occurrence,¹ I shall divert the word, which is almost useless in either his or its dictionary sense, to stand for the abstract particular which is, so to speak, the occurrence of an essence.

A trope then is a particular entity either abstract or consisting of one or more concrete entities in combination with an abstraction. Thus Napoleon and Napoleon’s forelock are not tropes, but Napoleon’s posture is a trope, and so is the whole whose constituents are his forelock and his posture, and so is his residing on Elba.

Turning now briefly from the alphabet of being to a glimpse of its syllabary, we observe two fundamental ways in which tropes may be connected with one another: the way of location and the way of similarity. These are categorially different and indeed systematic counterparts of one another – mirror images, as it were. Location is external in the sense that two tropes per se do not entail or necessitate or determine their location with respect to one another, while similarity is internal in the sense that, given any two tropes, there are entailed or necessitated or determined whether and how they are similar. (What further this *prima facie* difference amounts to we cannot pursue here.) Location is easiest thought of as position in physical space-time, but I intend the notion to include also all the analogous spreads and arrangements which we find in different conscious fields and indeed in any realm of existence which we can conceive – the whole interior stretch and structure of a Leibnizian monad, for example. Both modes of connection are describable in terms of distance and direction.

We are very familiar in a general way with the numberless distances and directions which compose locations in space and time, but not so used to thinking of the limiting value of such location (though very familiar with the phenomenon itself) – namely, being in the same place at the same time, the unique collocation and interpenetration which we call “belonging to or inhering in, or characterizing, the same thing.” Russell calls this “compresence”; I shall follow Whitehead, Keynes, and Mill in calling it “concurrence.” Plainly, what I called “color-plus-shape,” in the second paragraph back, is not just the sum of a color and a shape but their sum in concurrence; we might have said “color-cum-

shape.” We can now explain furthermore that Harlac and Bantic, our lollipop colors, are really complex, each consisting of a knob-color and a stick-color in a certain relative location, and similarly for the shapes. Since there are no short words (like “red” and “square”) which describe such complex colors and shapes, I shall ignore the sticks (supposed to be all alike) and use our trope names just for the qualities in the respective knobs.

It will not matter if the reader regards the use of “distance” and “direction” for resemblance relations as metaphorical so long as he gets the idea. Here we have no trouble with the notion of the limiting value, zero distance or precise similarity, but may need to think a little more about the lesser similarity or greater difference which holds, e.g., between a red and a purple, and still more, unless we are psychologists or phenomenologists, about such elaborate similarity distances and directions as are mapped on the color cone.

Any possible world, and hence, of course, this one, is completely constituted by its tropes and their connections of location and similarity, and any others there may be. (I think there are few others or none, but that is not necessary to the theory of tropes.) Location and similarity (or whatever else there is) provide all the relations, as the tropes provide the terms, but the total of the relations is not something over and above the total of the terms, for a relation R between tropes a and b is a constitutive trope of the complex $r'(a, b)$ (e.g., the concurrence-sum of Harlac and Hamis), while conversely the terms a and b will be in general composed of constituents in relation – though perhaps no more than the spread of a “smooth” quality, a “quale,” such as a color.

Any trope belongs to as many sets or sums of tropes as there are ways of combining it with other tropes in the world. Of special interest however are (1) the set or sum of tropes which have to it the relation of concurrence (the limiting value of location), and (2) the set or sum of those which have to it the relation of precise similarity (the limiting value of similarity, sometimes mischievously called “identity”). Speaking roughly now, the set or sum of tropes concurrent with a trope, such as our color component Harlac, is the concrete particular or thing which it may be said to characterize, in our example the lollipop Heraplem, or, to simplify the affair, the knob of the lollipop at a moment. By parallel, speaking roughly again, the set or sum of tropes precisely similar to a given trope, say Harlac again, may be supposed to be, or at least to

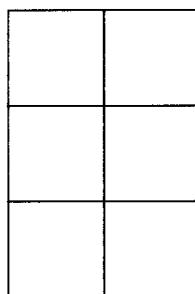


Figure 1

correspond formally to, the abstract universal or “essence” which it may be said to exemplify, in our illustration a definite shade of redness. (The tropes approximately similar to the given one provide a less definite universal.)

The phrase “set or sum” above is a deliberate hedge. A set is a *class* of which the terms are members; a sum is a whole of which the terms are parts, in the very primitive sense of “part” dealt with by recent calculi of individuals. In the accompanying figure (figure 1), for instance, the class of six squares, the class of three rows, and the class of two columns are different from each other and from the one figure; but the sum of squares, the sum of rows, and the sum of columns are identical with one another and with the whole.

What a difference of logical “type” amounts to, particularly in the philosophy of tropes, is far from clear, but everybody agrees that a sum is of the same type with its terms, as a whole is of the same type with its parts, a man of the same type with his arms and legs. The concept of a class or set, on the other hand, is notably more complex and questionable. A class has not been shown to be in any clear sense an abstract entity, but there is some excuse for considering it of a different type from its members. Convinced that a concrete thing is composed of tropes in a manner logically no different from that in which it is composed of any other exhaustive batch of parts, we have every incentive to say that a concrete thing is not a set but a sum of tropes; and let us so describe it. Whether the counterpart concept of the universal can be defined as the sum of similars – all merely grammatical difficulties aside – is not so clear; there is little doubt that the set or class will do the job.

All the paradoxes which attend the fashionable effort to equate the universal humanity, for example, with the class of concrete men (including such absurdities as that being a featherless biped is then the same as having a sense of humor) disappear when we equate it rather with our new set, the class of abstract particular humanities – the class whose members are not Socrates, Napoleon, and so forth, but the human trope in Socrates, the one in Napoleon, and so forth. Still wilder paradoxes resulted from the more radical nominalistic device of substituting the sum of concrete men for their class, and though most of these also are obviated by taking our sum of similar tropes instead, I am sure that some remain. Because concurrence and similarity are such symmetrical counterparts, I shall not be surprised if it turns out that while the concurrence complex must be a sum, the similarity complex must be a set.

In suggesting how both concrete particulars and abstract universals are composed of or “constructed from” tropes, I aver that those two categories do not divide the world between them. It does not consist of concrete particulars in addition to abstract universals, as the old scheme had it, nor need we admit that it must be constructible either from concrete particulars or from abstract universals, as recent innovators argue. The notions of the abstract and the universal (and hence of the concrete and the particular) are so far independent that their combinations box the logical compass. Socrates is a concrete particular. The component of him which is his wisdom is an abstract particular or “trope.” The one general wisdom of which all such wisdoms are members or examples is an abstract universal. The total Socratism of which all creatures exactly like him

are parts or members is a “concrete universal,” not in the idealistic but in a strictly accurate sense.

Having thus sorted out the rubrics, we can almost automatically do much to dispel the ancient mystery of predication, so influential in the idea of logical types. The prevalent theory has been that if y can be predicated of x , or inheres in or characterizes x , or if x is an instance of y , then x and y must be sundered by a unique logical and ontological abyss. Most of the horror of this, however, which has recently impelled some logicians to graceless contortions of language, is due to taking predication as one indissoluble and inscrutable operation and vanishes when our principles reveal predication to be composed of two distinct but intelligible phases. “Socrates is wise,” or generically “ a is ϕ ,” means that the concurrence sum (Socrates) includes a trope which is a member of the similarity set (wisdom-in-general). When we contrast a thing with a property or characteristic of it, a substantive with an adjective, we may intend either or both of these connections. . . .

A philosophy of tropes calls for completion in a dozen directions at once. Some of these I must ignore for the present because the questions would take us too far, some because I do not know the answers. . . . What in fact I shall do here is to defend the fundamental notion that there are entities at once abstract, particular, and actual, and this in two ways: the affirmative way of showing how experience and nature evince them over and over, and the negative way of settling accounts with old dialectical objections to them.

I deliberately did not use the word “abstract” to describe our tropes till we had done our best to identify them in other ways, lest the generally derogatory connotation of the word blind us to the reality of objects as plain as the sunlight (for indeed the sunlight is an abstract existent). The many meanings of “abstract” which make it repulsive to the empirical temper of our age suggest that anything abstract must be the product of some magical feat of mind, or the denizen of some remote immaterial eternity. . . .

At its broadest the “true” meaning of “abstract” is *partial, incomplete, or fragmentary*, the trait of what is less than its including whole. Since there must be, for everything but the World All, at least something, and indeed many things, of which it is a proper part, everything but the World All is “abstract” in this broad sense. It is thus that the idealist can denounce a cat as “abstract.” The more usual practice of philosophers, however, has been to require for “abstractness” the more special sort of incompleteness which pertains to what we have called the “thin” or “fine” or “diffuse” sort of constituent, like the color or shape of our lollipop, in contrast with the “thick,” “gross,” or chunky sort of constituent, like the stick in it.

If now one looks at things without traditional prepossessions, the existence of abstracta seems as plain as any fact could be. (To call them “abstracta” is bad in so far as the word is artificial, but it helps to avoid the prepossessions, including the implication of transcendent metaphysics, which hang about “abstract entities” and the psychological implication, as if such objects were our own private inventions, of “abstractions.”) . . .

I have no doubt that whole things like lollipops, trees, and the moon, do

exist in full-blooded concreteness, but it is not they which are present to the senses, and it is not awareness of abstracta which is “difficult, . . . not to be attained without pains and study.”² To claim primacy for our knowledge of concrete things is “mysticism” in the strict sense, that is, a claim to such acquaintance with a plethoric being as no conceivable stroke of psychophysics could account for. What we primarily see of the moon, for example, is its shape and color and not at all its whole concrete bulk. Generations lived and died without suspecting it had a concrete bulk. If now we impute to it a solidity and an aridity, we do it item by item quite as we impute wheels to a clock or a stomach to a worm. Evaluation is similarly focussed on the abstract. What most men have valued the moon for is its brightness; what a child wants of a lollipop is a certain flavor and endurance. He would much rather have these abstract qualities without the rest of the bulk than the bulk without the qualities. . . .

Though the uses of the trope to account for substances and universals are of special technical interest, the impact of the idea is perhaps greater in those many regions not so staled and obscured by long wont and old opinion and not so well supplied with alternative devices. While substances and universals can be “constructed” out of tropes, or apostrophized in toto for sundry purposes, the trope cannot well be “constructed” out of them and provides the one rubric which is hospitable to a hundred sorts of entity which neither philosophy, science, nor common sense can forgo. This is most obvious in any attempt to treat of the mind, just because the mind’s forte is the tuning, focussing, or spotlighting – in brief, the abstraction – which brings abstracta into relief against a void or nondescript background. A pain is a trope par excellence, a mysterious bright pain in the night, for example, without conscious context or classification, yet as absolutely and implacably its particular self as the Great Pyramid. But all other distinguishable contents are of essentially the same order: a love, or a sorrow, or a pleasure.

The notion, however, gets its best use in the theory of knowledge. The “sensible species” of the Scholastics, the “ideas” of Locke and Berkeley, the ideas and impressions of Hume, the sense data of later epistemology – once they are understood as tropes, and as neither things nor essences, a hundred riddles about them dissolve, and philistine attacks on theory of knowledge itself lose most of their point. We need not propose that a red sensum, for example, is perfectly abstract (whatever that might be). But even though it have such distinguishable components as a shape and a size as well as a color, and though the color itself involve the “attributes” of hue, brightness, and saturation, still it is abstract in comparison with a whole colored solid. According to reputable psychologists, furthermore, there can be data much more abstract, professed empiricists to the contrary notwithstanding: data which have color and no other character, or even hue and no other attribute.

. . . A whole soul or mind, if it is not a unique immaterial substance on its own, is a trope. In one manner or another everybody grants that there is a very considerable correlation between the components of conscious experience and the processes of the body. The physical correlates of conscious tropes are then in general physical tropes – the patterns of arrangement and motion which

behaviorally or physiologically are beliefs, discriminations, perceptions, desires, and the rest. In our happy-go-lucky way, however, the human functions we generally speak and think about are "mixed tropes," like the "mixed modes" of the Cartesians. A belief, a sensation, an emotion, a purpose – each is partly the conscious item and partly the behavioral one.

On the model of such mixed tropes we must understand a love affair, an act of contrition, or a piece of impudence. A word or a sentence in a particular occurrence is a trope, mental, physical, or mixed. The "same" word in many occurrences is the corresponding universal. This distinction differs from Peirce's between "token" and "type" inasmuch as it avoids the usual identification of the token with the concrete ink splotch, for example, in which our trope inheres – an ill-timed obsession with substance which is out of accord both with ordinary ideas and with the fact that most verbal tropes cannot plausibly be imputed to any special concrete objects anyhow. A word is a product of art, and art is full of analogous tropes. A statue is not a trope, but the connoisseur who gloats over its form, its texture, or its color is gloating over a trope. A musical performance, a song or a symphony, is a trope, and so is a musical theme – not the *kind* of theme, the universal which recurs throughout the same work in all its repetitions, but any single case of it. . . .

If a bit of perceptual behavior is a trope, so is any response to a stimulus, and so is the stimulus, and so therefore, more generally, is every effect and its cause. When we say that the sunlight caused the blackening of the film we assert a connection between two tropes; when we say that sunlight in general causes blackening in general, we assert a corresponding relation between the corresponding universals. Causation is often said to relate events, and generally speaking any event is a trope: a smile, a sneeze, a scream, an election, a cold snap, a storm, a lightning flash, a conspiracy, perhaps a wave, and so on up to such big and important events that they have proper names, like Lulu, the H-bomb explosion. We have called a trope a "case" of its universal, while the universal is the "kind" of the trope, and so it is no surprise that a medical "case" is a trope – in the sense, at any rate, in which a person is said to *have* a case of typhoid fever rather than to *be* a case of it (for the latter "case" stands for the whole concrete individual). A high-school boy, uncoached, has assured me, "Of course there's such a thing as redness – this pencil *has a case of it*."

When a scientist reports a temperature or a velocity or a viscosity, he is reporting a trope – not a universal, because it is a once-for-all occurrence, but not a concrete thing either, though doubtless a component of one. He is likely to call it an "aspect" of the thing or preferably a "state," and generally speaking, a "state" of a thing or a nation is a trope (though "state" too may mean a *kind* of state, the universal). Recent developments in subatomic physics, a none too reliable oracle, suggest that an electron, e.g., just is an existent state and that the common-sense philosophy of the concrete here abdicates altogether in favor of the trope.

Since events and processes are tropes, and also *cases* and *states*, one wonders about "facts," "states of affairs," and "what is the *case*." Mary's beauty is a trope; Mary-being-beautiful, the "fact" which makes "Mary is beautiful" true, seems a similar but queerer business. C. I. Lewis is surely right that a state of

affairs is “abstract and adjectival” rather than a “chunk,”³ and it would be delightful to say that a state of affairs stands to its proposition as an ordinary trope does to its property (universal). But I shrink from endowing the theory of tropes with either the assets or the deficits of a theory of facts, of states of affairs, or of propositions.

A variety of trope which has been much entertained by the philosopher unaware is that of geometrical figures, circles, triangles, and so forth. These have been alternately treated as if they were Platonic universals and as if they were concrete particulars, whereas in fact they are neither. Triangularity, to be sure, is an abstract universal, and a triangular object is a concrete particular, but a triangle is an abstract particular. Since a triangle, a circle, or any of the rest, while being particular, is an abstractum exhausted in one thin but salient character, the propensity of many generations of writers for taking them as typical “things” was perhaps largely responsible for that catastrophic doctrine of real essences by which truly concrete things like men or trees are supposed to be similarly dominated by a single essential character in each.

If a geometrical figure is a trope, so is a woman’s figure, and so is her complexion or her digestion – in that sense in which she is more concerned to take care of “her figure,” “her complexion,” “her digestion,” than those of anybody else, however similar. Thus too when someone tells her, “I love the sweetness of your voice and the serenity of your brow,” he does not mean, if he is wise or faithful, that *kind* of voice and brow, wherever they occur in the world manifold, but these particular cases. But while the complexion of a face, a smile on it, the whole expression of it, and every component of the expression, and the shape of the whole face or of any part of it, all are tropes, the face itself is a *surface*, and some logical philosophers who shy at “abstract entities” think that a surface escapes that epithet. Well, a surface does seem to occupy a sort of borderline status, but this is no more than our doctrine entails, for we have expressly denied that “between the abstract and the concrete there can be no intervening stages.”⁴

As the shape is to the surface, perhaps, so the surface is to the solid. The bigger difference is that a surface is “concrete” in two dimensions as a triangle on it is not concrete at all. This sort of quasiconcreteness, we note, belongs also to an instantaneous three-dimensional solid in comparison with one which is appreciably extended in time. Only an old familiarity with the terms of geometry, I think, makes anyone suppose that a surface or an instantaneous solid has in any fundamental way a more robust being than a four-dimensional shape or temperature. Similar questions and answers may be expected to attend such entities as the equator or a hole.

. . . The arguments which I wish especially to weigh now . . . are those which assert that the status of abstractness itself is incompatible with actual existence. Because our tropes are advocated not as entities additional to concrete things but as constituents of them, any effective denial or defense of them must be an argument, not for or against a transcendent realm of being, but concerning what sorts of constituents are real and which, if any, are not, and hence the rights and significance of analysis.

When the issue is thus narrowed, the principal dialectical objection may be summed up in the old maxim that a true existent must be such as can exist by and in itself, *per se* and *in se*. We have called to witness the idealists that if this is taken without reservation, then ordinary concrete things – men's limbs as well as their temperaments, the men as well as their limbs – cannot be real either, for only the world as an eternal whole exists *per se* and *in se*. To preserve the advantage of the concrete over the abstract, we must reinterpret "*per se*" to accommodate the former but not the latter. If we do not altogether beg the question by defining "*per se*" to mean concrete, the most we can say is that the concrete is comparatively independent of its context and that it can, within wide limits, be moved around without losing its identity. Whereas we can pull the stick or an atom out of a lollipop and even put it back on demand, we cannot strip off its color and shape or extract the pure flavor of it, and still more obviously we cannot assemble a lollipop from such components.

Even this difference, however, fades out under examination. It is merely an accident of physical fact, after all, that sticks are not dissipated when removed from lollipops, or wheels from watches, as a volume of chlorine is when let out of a flask. Many concrete parts are physically incapable of removal, as the Mississippi River is from the Mississippi Valley, and most of them which are removable, as a whelk is from its shell, are so damaged by the operation that they are, as we say, not the same thing at all. But whether removable or not in the ordinary sense of "removable," they are always irremovable in the one queer respect which is cardinal to our kind of question. For the actual events which compose the existence of the watch wheel now before me on the table are numerically as distinct from those which compose the wheel inside the watch ten minutes ago, or back inside the watch again two hours from now, as any of these is from my fingers or from Jupiter. Their community consists logically of only a continuity of similar events or states strung between.

To bring this out best let us use the word "constituents" (what "stand together") for parts or components as they exist within a complex object as we describe it, and the word "ingredients" (what "go into" the object) for those entities with which we operate when we start generating it or when we are through disintegrating it. The wheels of a watch or the stick of a lollipop qua "ingredients" happen to be conspicuously affiliated with the wheels and stick qua "constituents." The milk, sugar, eggs, and flour which went to making a cake, however, or the flaccid and ruined organs dissected out of an animal body, are much less fairly described as "the same as" the constituents of the object while it lasted.

The atomic theory was the great triumph of the feeling that things ought to have concrete parts which are at once constituents and ingredients, but with discontinuous and identityless electrons taking the place of atoms, this reassurance, limited to begin with, has become worth next to nothing. If now we turn back to our abstractions, the situation seems much the same with them. They often cannot be "moved" in even the crude sense in which some concrete things can, but in whatever sense "the same" wheel survives when taken from a watch, in that sense, if we can believe our eyes, the color of a blouse, for example, may

be transferred to the wash water, or the glare of an electric light survives for a moment in the positive afterimage.

There remains one severe question, whether abstract qualities do not logically or metaphysically *require* their contexts as concrete entities do not. On the idealist logic of internal relations, everything requires or entails its whole cosmic matrix, but it seems at first sight that even those who deny this extravagance of idealism would have to grant it inconceivable that an abstractum should exist by itself, like the grin left behind by the Cheshire cat, and not as a component of something concrete.

This raises first the question how we define "concrete." If it means merely what does exist unjoined with further components, then it is a verbalism that whatever exists must be concrete or a component of something concrete. The real question then is whether an entity which is "abstract" in the sense that it is conjoined in something concrete with other abstracta, as the shape of a watch is, for example, may be duplicated elsewhere by an entity precisely similar internally but not thus conjoined with anything. Our instincts say "No," that there is a sort of cosmic standard of concreteness, a certain degree of richness or thickness, which perhaps is a general maximum that nothing can exceed, but which at any rate is a general minimum that an entity must attain in order, as the Scholastics say, "to be apt for existence," or that, in Aristotle's phrase, it "can exist apart."⁵

Plausible though it be, however, that a color or a shape cannot exist by itself, I think we have to reject the notion of a standard concreteness. For it means that from the awareness of even the thinnest abstractum – indeed the thinner the better – we could *deduce* the presence of the rest of a concrete thing, if not its specific character then at least that there is something concrete there, as Descartes deduced from a conscious state the existence of a spiritual substance in which it inhered. It seems to me an analytic principle that all deduction must be analytic so that while any proper component is deducible from the composite which contains it, no composite is deducible from any of its proper components, and hence that abstracta must in principle be as independent of their contexts as concrete things are.

Though it has been interesting to observe, for its own sake, that the abstract and the concrete entities are much alike with respect to independence and manipulability, this was nearly superfluous for our main purpose because the actuality of entities does not in any event depend on whether they are independent and manipulable but only on whether they are *there*. This was the import of our differentiation between constituents and ingredients. The constituents of the universe are not the ingredients of which God made it, if He made it of any, nor the fragments which will supervene when it decomposes, but are the stars and atoms and men, the shapes and tastes and numbers, which are present in it now; and in the same way the constituents of a lollipop, for example, are not the stuffs which went into the kettle, nor the shards which would result from running it through a grinder, but the sectors, the facets, the atoms, the structures, and the qualities which are its current parts and components *in situ*. That things consist of tropes does not imply either that they were

made by putting tropes together or that they can be dismantled by taking tropes apart. . . .

The most that can be done by a thesis in first philosophy like ours is to prepare the way for more concrete and synoptic inquiry. We are only beginning to philosophize till we turn from the bloodless proposition that things in any possible world must consist of tropes to specific studies of the sorts of tropes of which the things in this world actually consist. It is a virtue of our thesis that it does not strangle or eviscerate the great problems in the philosophical cradle but keeps them alive to face the test of experience and logic. It will be a further virtue if it assists, as I think it will, in their formulation and appraisal. Are there only physical objects and energies, or only minds or spirits, or are there both? How, specifically, is a physical object constituted, and how a mind, and how are they related? These topics of gigantic hypothesis are the last of philosophy for which the first is made.

Notes

- 1 Santayana, George, *The Realm of Matter*, ch. 6, in *Works* (New York: Scribner's, 1937), vol. 14, pp. 288–304.
- 2 Berkeley, George, *Principles of Human Knowledge* (Dublin, 1710), Introd., Sec. 10.
- 3 C. I. Lewis, *An Analysis of Knowledge and Valuation* (La Salle, Ill.: Open Court, 1946, p. 55).
- 4 Ibid., p. 475. Lewis, who elsewhere suggests there are degrees of abstractness, is here equating abstractness with universality, and concreteness and *universality*, we know, are just incommensurate.
- 5 Aristotle, *Metaphysics*, 1070b 36.

3 The Principle of Individuation: From *Human Knowledge, its Scope and Limits**

Bertrand Russell

I shall discuss in this chapter the modern form of a very old problem, much discussed by the scholastics, but still, in our day, far from being definitively solved. The problem, in its broadest and simplest terms, is this: “How shall we define the diversity which makes us count objects as two in a census?” We may put the same problem in words that look different; e.g., “What is meant by a ‘particular’?” or “What sort of objects can have proper names?”

* From Bertrand Russell, *Human Knowledge: its Scope and Limits* (New York: Simon and Schuster, 1948). Reprinted by permission of Routledge and the Bertrand Russell Peace Foundation.

Three views have been influentially advocated.

First: a particular is constituted by qualities; when all its qualities have been enumerated, it is fully defined. This is the view of Leibniz.

Second: a particular is defined by its spatio-temporal position. This is the view of Thomas Aquinas as regards material substances.

Third: numerical diversity is ultimate and indefinable. This, I think, would be the view of most modern empiricists if they took the trouble to have a definite view.

The second of the above three theories is reducible to either the first or the third according to the way in which it is interpreted. If we take the Newtonian view, according to which there actually are points, then two different points are exactly alike in all their qualities, and their diversity must be that bare numerical diversity contemplated in the third theory. If, on the other hand, we take – as everyone now does – a relational view of space, the second theory will have to say, “If A and B differ in spatio-temporal position, then A and B are two.” But here there are difficulties. Suppose A is a shade of color: it may occur in a number of places and yet be only one. Therefore our A and B must not be qualities, or, if they are, they must be qualities that never recur. If they are not qualities or bundles of qualities, they must be particulars of the sort contemplated in our third theory; if they are qualities or bundles of qualities, it is the first of our three theories that we are adopting. Our second theory, therefore, may be ignored.

The construction of points and instants in our three preceding chapters used “events” as its raw material. Various reasons, of which the theory of relativity has been the most influential, have made this procedure preferable to one which, like Newton’s, allows “points,” “instants,” and “particles” as raw material. It has been assumed, in our constructions, that a single event may occupy a finite amount of space-time, that two events may overlap both in space and in time, and that no event can recur. That is to say, if A wholly precedes B, A and B are not identical. We assumed also that if A wholly precedes B, and B wholly precedes C, then A wholly precedes C. “Events” were provisionally taken as “particulars” in the sense of our third theory. It was shown that if a raw material of this sort is admitted, space-time points and space-time order can be constructed.

But we are now concerned with the problem of constructing space-time points and space-time order when our first theory is adopted. Our raw material will now contain nothing that *cannot* recur, for a quality can occur in any number of separate places. We have therefore to *construct* something that *does not* recur, and until we have done so we cannot explain space-time order.

We have to ask ourselves what is meant by an “instance.” Take some definite shade of color, which we will call “C.” Let us assume that it is a shade of one of the colors of the rainbow, so that it occurs wherever there is a rainbow or a solar spectrum. On each occasion of its occurrence, we say that there is an “instance” of C. Is each instance an unanalyzable particular, of which C is a quality? Or is each instance a complex of qualities of which C is one? The former is the third of the above theories; the latter is the first.

There are difficulties in either view. Taking first the view that an instance of C

is an unanalyzable particular, we find that we encounter all the familiar difficulties connected with the traditional notion of "substance." The particular cannot be defined or recognized 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 recognized to be dangerous.

It is difficult to see how something so unknowable as such a particular would have to be can be required for the interpretation of empirical knowledge. The notion of a substance as a peg on which to hang predicates is repugnant, but the theory that we have been considering cannot avoid its objectionable features. I conclude, therefore, that we must, if possible, find some other way of defining space-time order.

But when we abandon particulars in the sense which we have just decided to reject, we are faced, as observed above, with the difficulty of finding something that will not be repeated. A simple quality, such as the shade of color C, cannot be expected to occur only once. We shall seek to escape this difficulty by considering a "complex" of qualities. What I mean will be most easily understood if stated in psychological terms. If I see something and at the same time hear something else, my visual and auditory experiences have a relation which I call "compresence." If at the same moment I am remembering something that happened yesterday and anticipating with dread a forthcoming visit to the dentist, my remembering and anticipating are also "compresent" with my seeing and hearing. We can go on to form the whole group of my present experiences and of everything compresent with all of them. That is to say, given any group of experiences which are all compresent, if I can find anything else which is compresent with all of them I add it to the group, and I go on until there is nothing further which is compresent with each and all of the members of the group. I thus arrive at a group having the two properties: (a) that all the members of the group are compresent, (b) that nothing outside the group is compresent with every member of the group. Such a group I shall call a "complete complex of compresence."

Such a complex I suppose to consist of constituents most of which, in the natural course of events, may be expected to be members of many other complexes. The shade of color C, we supposed, recurs every time anybody sees a rainbow distinctly. My recollection may be qualitatively indistinguishable from a recollection that I had yesterday. My apprehension of dental pain may be just what I felt before my last visit to the dentist. All these items of the complex of compresence may occur frequently, and are not essentially dated. That is to say, if A is one of them, and A precedes (or follows) B, we have no reason to suppose that A and B are not identical.

Have we any reason, either logical or empirical, to believe that a complete complex of compresence, as a whole, cannot be repeated? Let us, in the first place, confine ourselves to one person's experience. My visual field is very complex, though probably not infinitely complex. Every time I move my eyes, the visual qualities connected with a given object which remains visible undergo changes: what I see out of the corner of my eyes looks different from what is in

the center of my field of vision. If it is true, as some maintain, that my memory is colored by my whole past experience, then it follows logically that my total recollections cannot be exactly similar on two different occasions; even if we reject this doctrine, such exact similarity seems very improbable.

From such considerations I think we ought to conclude that the exact repetition of my total momentary experience, which is what, in this connection, I call a "complete complex of compresence," is not logically impossible, but is empirically so exceedingly improbable that we may assume its non-occurrence. In that case, a complete complex of compresence will, so far as one person's experience is concerned, have the formal properties required of "events"; i.e., if A, B, C are complete complexes of compresence, then if A wholly precedes B, A and B are not identical; and if B also wholly precedes C, then A wholly precedes C. We thus have the requisites for defining the time-order in one person's experience.

This, however, is only part, and not the most difficult part, of what we have to accomplish. We have to extend space-time order beyond one person's experience to the experiences of different people and to the physical world. In regard to the physical world, especially, this is difficult.

So long as we confine ourselves to one person's experience, we need only concern ourselves with time. But now we have also to take account of space. That is to say, we have to find a definition of "events" which shall insure that each event has not merely a unique temporal position but a unique spatio-temporal position.

So long as we confine ourselves to experiences, there is no fresh difficulty of a serious kind. It may be taken as virtually certain, on empirical grounds, that my visual field, whenever my eyes are open, is not exactly similar to that of anyone else. If A and B are looking simultaneously at the same scene, there are differences of perspective; if they change places, A will not see exactly what B was seeing, because of differences of eyesight, changes of lighting meanwhile, and so on. In short, the reasons for supposing that no total momentary experience of A is ever exactly like some total momentary experience of B are of the same kind as the reasons for supposing that no two total momentary experiences of A are ever exactly alike.

This being granted, we can establish a spatial order among percipients by means of the laws of perspective, provided there is any physical object that all the percipients concerned are perceiving. If there is not, a process by means of intermediate links can reach the same result. There are of course complications and difficulties, but they are not such as concern our subject at all closely, and we may safely ignore them.

What can be said about the purely physical world is hypothetical, since physics gives no information except as to structure. But there are reasons for supposing that, at every place in physical space-time, there is at every moment a multiplicity of occurrences, just as there is in a mind. "Compresence," which I take to have a merely ostensive definition, appears in psychology as "simultaneity in one experience" but in physics as "overlapping in space-time." If, as I maintain, my thoughts are in my head, it is obvious that these are different

aspects of one relation. However, this identification is inessential to my present argument.

When I look at the stars on a clear night, each star that I see has an effect on me, and has an effect on the eye before it has an effect on the mind. It follows that, at the surface of the eye, something causally connected with each visible star is happening. The same considerations apply to ordinary objects seen in daylight. At this moment I can see white pages covered with writing, some books, an oval table, innumerable chimneys, green trees, clouds, and blue sky. I can see these things because there is a chain of physical causation from them to my eyes and thence to the brain. It follows that what is going on at the surface of my eye is as complex as my visual field, in fact as complex as the whole of what I can see. This complexity must be physical, not merely physiological or psychological; the optic nerve could not make the complex responses that it does make except under the influence of equally complex stimuli. We must hold that wherever the light of a certain star penetrates, something connected with that star is happening. Therefore in a place where a telescope photographs many millions of stars, many millions of things must be happening, each connected with its own star. These things are only "experienced" in places where there is a recording nervous system, but that they happen in other places also can be shown by cameras and dictaphones. There is therefore no difficulty of principle in constructing "complexes of compresence," where there are no percipients, on the same principles as we employed in dealing with momentary experiences.

Abandoning speculations about the physical world, about which our knowledge is very limited, let us return to the world of experience. The view which I am suggesting, as preferable to the assumption of such wholly colorless particulars as points of space or particles of matter, may be expressed as follows:

There is a relation, which I call "compresence," which holds between two or more qualities when one person experiences them simultaneously – for example, between high C and vermillion when you hear one and see the other. We can form groups of qualities having the following two properties: (a) all members of the group are compresent; (b) given anything not a member of the group, there is at least one member of the group with which it is not compresent. Any one such complete group of compresent qualities constitutes a single complex whole, defined when its constituents are given, but itself a unit, not a class. That is to say, it is something which exists not merely because its constituents exist but because, in virtue of being compresent, they constitute a single structure. One such structure, when composed of mental constituents, may be called a "total momentary experience."

Total momentary experiences, as opposed to qualities, have time relations possessing the desired characteristics. I can see blue yesterday, red today, and blue again tomorrow. Therefore, so far as qualities are concerned, blue is before red and red is before blue, while blue, since it occurs yesterday and tomorrow, is before itself. We cannot therefore construct, out of qualities alone, such a relation as will generate a series. But out of total momentary experiences we can do this, provided no total momentary experience ever exactly recurs. That this

does not happen is an empirical proposition, but, so far as our experience goes, a well-grounded one. I regard it as a merit in the above theory that it gets rid of what would otherwise be synthetic a priori knowledge. That if A precedes B, B does not precede A, and that if A precedes B and B precedes C, then A precedes C are synthetic propositions; moreover, as we have just seen, they are not true if A and B and C are qualities. By making such statements (in so far as they are true) empirical generalizations, we overcome what would otherwise be a grave difficulty in the theory of knowledge.

I come back now to the conception of "instance." An "instance" of a quality, as I wish to use the word, is a complex of compresent qualities of which the quality in question is one. In some cases this view seems natural. An instance of "man" has other qualities besides humanity: he is white or black, French or English, wise or foolish, and so on. His passport enumerates enough of his characteristics to distinguish him from the rest of the human race. Each of these characteristics, presumably, exists in many other instances. There are baby giraffes who have the height mentioned in his passport, and parrots who have the same birthday as he has. It is only the assemblage of qualities that makes the instance unique. Every man, in fact, is defined by such an assemblage of qualities, of which humanity is only one.

But when we come to points of space, instants of time, particles of matter, and such stock in trade of abstract science, we feel as if a particular could be a "mere" instance, differentiated from other instances by relations, not by qualities. To some degree, we think this of less abstract objects: we say "as like as two peas," suggesting that between two peas there are no qualitative differences. We think also that two patches of color may be *merely* two, and may differ only numerically. This way of thinking, I maintain, is a mistake. I should say that when the same shade of color exists in two places at once, it is one, not two; there are, however, two complexes, in which the shade of color is combined with the qualities that give position in the visual field. People have become so obsessed with the relativity of spatial position in physics that they have become oblivious of the absoluteness of spatial position in the visual field. At every moment, what is in the center of my field of vision has a quality that may be called "centrality"; what is to the right is "dexter," what to the left "sinister," what above "superior," what below "inferior." These are *qualities* of the visual datum, not relations. It is the complex consisting of one such quality combined with a shade of color that is distinct from the complex consisting of the same shade elsewhere. In short, the multiplicity of instances of a given shade of color is formed exactly as the multiplicity of instances of humanity is formed, namely, by the addition of other qualities.

As for points, instants, and particles, in so far as they are not logical fictions similar considerations apply. Take first instants. It will be found that what I call a "momentary total experience" has all the formal properties required of an "instant" in my biography. And it will be found that where there is only matter, the "complete complex of compresence" may serve to define an instant of Einsteinian local time, or to define a "point-instant" in cosmic space-time. Points in perceptual space are defined without any trouble, since the qualities of up-

and-down, right-and-left, in their various degrees, have already all the properties that we require of “points.” It is indeed this fact, together with perception of depth, that has led us to place such emphasis on the spatial characteristics of the world.

I do not think “particles” can be dealt with quite in the above manner. In any case, they are no longer part of the fundamental apparatus of physics. They are, I should say, strings of events interconnected by the law of inertia. They are no longer indestructible, and have become merely convenient approximations. . . .

4 Distinct Indiscernibles and the Bundle Theory*

Dean W. Zimmerman

A: Like Locke, you think an object must be something more than its properties. So you posit a mysterious “substratum,” an unreachable “kernel” that bears properties but is not itself a property. This is metaphysics at its most gratuitous and pernicious. All we observe or detect are the properties of things, and a particular substance is nothing more than a bundle of properties.

B: So you say. But remember the puzzle I put to you many years ago:

Isn’t it logically possible that the universe should have contained nothing but two exactly similar spheres? We might suppose that each was made of chemically pure iron, had a diameter of one mile, that they had the same temperature, color, and so on, and that nothing else existed. Then every quality and relational characteristic of the one would also be a property of the other. Now if what I am describing is logically possible, it is not impossible for two things to have all their properties in common.¹

You have yet to satisfy me that your bundle theory of substance is compatible with the possibility of such a world.

A: I must admit, your two-sphere world had me worried for a time. But it now seems clear to me that the possibility you described poses no real threat to the bundle theory.

B: How so? The spheres have to be bundles of the very same universals; and they can’t be distinguished by their relations to one another, either. Throwing

* Portions of this paper originally appeared under the same title in *Mind*, 106 (1997), pp. 305–9. Reprinted by permission of the author and publisher.

in relations to different *places* won't help, since the places in question are indiscernible, too. If you posit distinct but indiscernible places, doesn't this amount to the recognition of things that are something more than mere bundles of universals? The only way out is to deny that the two-sphere world is really possible. But I know you too well to think that you'll take that route; you're not one of these "modally-challenged"² philosophers, unable to recognize a possibility when they see one.

A: Ah, but there *is* another way out.³ In order to see it, you must first recall that the universals I'm bundling into substances are not, of course, Platonic entities existing outside of space and time somewhere. They're "immanent universals," located right where and when their instances are.

B: Oh, you're bundling *tropes*, particular *instances* of universals, which can differ *solo numero*. That will solve the problem – but it's cheating, from the point of view of the traditional bundle theory. You've brought brute particularity back into your metaphysical picture; you're not bundling real *universals* any longer.

A: No, immanent universals aren't tropes; they're real universals, wholly present in each instance. They differ from Platonic universals only in being spatio-temporally located.

B: So the blueness on the surface of one sphere, say, is numerically identical with the blueness on the surface of another sphere of exactly the same hue?

A: Right. And you're probably beginning to see how I'll answer the two-sphere challenge. It's quite simple really: the situation you describe is surely possible; but it is a world in which a *single* bundle of universals – the universals of solidity, mass, shape, color, etc. colocated in one of the spheres – is *at some distance from itself*.

B: "At some distance from itself"? Surely that's a contradiction.

A: If so, then the very idea of an immanent universal is contradictory. An immanent universal will routinely be "at some distance from itself," in the sense that it is wholly present in more than one place. If you grant me immanent universals, then you must allow that my redescription of the sphere-world is consistent.

B: But it is a *redescription*, is it not? The world I described contains *two* spheres. But your bi-located bundle is just *one* thing that shows up in two places.

A: Granted, you *say* the world has two distinct spheres in it; but to insist on including this as part of the description of the world is to beg the question against the bundle theorist. I submit that the possibility your story illustrates is simply this: *a symmetrical universe*, a world in which the pattern of properties

exemplified on one side of a certain plane is precisely mirrored on the opposite side. You want to insist that, in addition, the objects on the one side of the plane of symmetry are *numerically distinct* from those on the other. But it is not at all clear to me that *that* is possible.

B: Let me see if I understand how your immanent universals fit together to constitute an object. Take one of the homogeneous spheres. Suppose it's uniformly blue all over. You want to say that the blueness on the one side is identical with the blueness on the other, the blueness on the top half is identical with the blueness on the bottom half, and so on?

A: Exactly; although the example presents certain difficulties – no one seems to agree about what blueness is, or about whether it's an intrinsic property on the surfaces of objects. But we can just let blueness stand for some real intrinsic property present all over the surface of the homogeneous sphere.

B: Fine. Now you would agree, I imagine, that the sphere possesses certain causal capabilities in virtue of the universals involved in it.

A: I suppose so; although I have no settled views about the metaphysics of causation.

B: Nor have I. But one thing that does seem clear to me is that, for example, a sphere translates motion to another ball when it strikes it in virtue of its mass and speed, but not in virtue of its color; and it causes a blue image to appear on a Polaroid in virtue of its color but not its mass; and so on. And generally, when the sphere causes something to happen, it will usually be only some of its properties that are causally significant. Furthermore, it is often only the properties possessed by a *part* of an object that are relevant to its producing a given effect. For instance, a polaroid of one of our blue spheres has a blue image because of the blueness on only one side. Its backside could have been red or green for all it matters to this particular causal transaction.

A: I think I'm beginning to see where you're headed. You want to say that the blueness on the side of the sphere facing the camera has to be distinct from the blueness on the back of the sphere, since only the former figures among the causes of the blue image.

B: Yes, that's more or less what I'm working up to.

A: I answer that causal relations among universals, like so many other relations among them, must be relativized in various ways. Take distance, for example. If the spheres are five feet in diameter, and one diameter apart, then blueness will be five feet from itself, and ten feet from itself, and fifteen feet from itself, and so on. There is no contradiction here, since distance relations are relativized: blueness is five feet from itself relative to the surfaces that are closest; ten feet from

itself relative to the inner surface of one and the far side of the other. Similarly, blueness causes the blue image on the film relative to one side of the sphere, but not relative to the other side.

B: You want to say, then, that blueness “causes-from-*here*” the image, where “here” is the location of the surface facing the camera; but it doesn’t “cause-from-*there*” the image, if “there” is the location of the surface on the other side?

A: Yes, something like that. However, talk about relativizing to *locations* may not be the best way to put it.

B: Right. For I’ll ask what these locations are like. If they’re literal parts of some kind of substantival space, then I’ll ask if they are or are not identical with bundles of universals. If they’re not, then you’ve brought in brute particularity again. If they are bundles of universals, I’ll want to know what distinguishes the one from the other.

A: But in the case you’ve described, there *is* a difference between the region of space occupied by the surface facing the camera and the region occupied by the surface facing away from it. The former is *closer to a camera* than the latter. So it is relative-to-the-space-that-is-closer-to-a-camera that blueness causes the image; and this space can be identified with a bundle of space-properties (pointhood or regionhood, say) plus the relational property of *being such-and-such a distance from a camera*.

And now that I put it this way, I see that I needn’t even go along with your initial suggestion that causing must be relativized to locations in space. I can just as easily say that it is blueness relative to the surface of one half of the sphere that causes the image; and the causally relevant half can be identified with a bundle of universals that includes the relational property *being such-and-such a distance from a camera*. This property isn’t part of the otherwise indiscernible bundle making up the sphere’s other half. So I can distinguish between the causing that blueness does from *here*, and the causing it does (or doesn’t do) from *there*, by reference not to different parts of space but to the different objects that are here and there – objects that are different because they are identical with bundles that include different properties.

B: Suppose there is an exactly similar camera on the opposite side as well. Then each half of the sphere will be made of precisely the same bundle of universals, even if you include these funny relational properties like “being a certain distance from a camera.” Doesn’t this force you to admit that the causal contribution made by blueness must be relativized to a *place*, and not to a part of the object?

A: No; for now there is no need to distinguish between the causal role played by blueness relative to the front half of the sphere and the causal role it plays relative to the back of the sphere. If this is a perfectly symmetrical world, the

cameras, like the spheres in the original two-sphere world, are really one and the same bundle at some distance from itself; and the causal relation holding between blueness and camera on one side is identical with the causal relation holding between blueness and camera on the other.

B: This is making me dizzy. Crossing the axis of symmetry in the camera-sphere-camera world is like passing through the looking glass: the camera you leave behind is identical with the one in front of you!

But now suppose there is something indeterministic about these cameras; whether or not the blueness causes a blue image depends upon some quantum-mechanical fluctuation, say. Up to a given time t , the causal relations between blueness and camera are the same on either side; and at t an image forms on the film of both cameras. The image is caused by the way the sphere looked right up to t , but not because of the way it looked at t – in other words, it's not a case of simultaneous causation. Now, given that the process is indeterministic, and that the cameras are independent, there was some chance that only one camera would work, that only one blue image would be formed. But how can you make sense of this possibility? One wants to say: if the shutter of the camera on *this* side had failed to open, as it could well have done, then the blueness on this side would have failed to cause a blue image while the blueness on the other side would have succeeded. You can't say that, though; for the blueness on this side is identical with the blueness on the other side, and the camera on this side is identical with the camera on that side – in fact this side of the sphere is identical with that side! So it is impossible that the one do something that the other does not do.

A: There are a couple of ways I can try to allow for the possibilities you describe, depending on the theory of time I hold. Suppose that I'm a "presentist" – in other words, that I think the only things that exist are those that exist *now*. Someone who holds this view cannot, I expect, hold that two bundles of universals differ now in virtue only of their relations to future things. But the presentist bundler can admit that there is a possible world just like the one you describe up to t , but diverging thereafter in virtue of the fact that one camera works and the other doesn't. This is a world in which a single camera-ish bundle of universals is located at some distance from itself for a while, and then is located at some distance from a *distinct* camera-ish bundle of universals – a non-functioning camera-ish bundle of universals, as it happens.

B: But, since the before- t cameras and blueness were really one camera and one blueness, the presentist won't be able to say that there is a world in which *this* camera works and *that* one doesn't, and another world in which *that* camera works and *this* one doesn't.

A: True enough; although the possibility that *can* be allowed for is still fairly close to the one you're after.

B: Not by my lights. But in any case, I know you're not a presentist; you're a "four-dimensionalist." You think everything "co-exists" in a big space-time block. So what would you say?

A: What I would say about your two cameras is this: a world in which one camera fails to function is a world in which the camera-ish bundles were distinguished *from the start* in virtue of their differing spatiotemporal relations to later camera-ish bundles. But you are right that I cannot allow that it's possible, in the world where "both" cameras work, to pick out one side of the sphere and one camera and assert that they could have failed to produce a picture while the other side and the other camera succeeded. For, on my view, "they" are not two but one.

B: It seems to me that the possibility you are able to recognize is not even close to the one I pointed out. I said that blueness could have failed to bring about its result relative to *this* side and *this* camera, while succeeding relative to the *other* side and the *other* camera. All you can admit is that blueness could have succeeded relative to this side and this camera while also being a part of a hemisphere distinct from any camera or hemisphere in the original world. The whole sphere in this other possible world has a *different* side relative to which blueness fails to cause an image in a *different* camera.

A: I grant you that. But doesn't your description of the case just presuppose the falsity of my bundle-theoretic approach, which says that a "pair" of complete indiscernibles is really one and the same bundle at some distance from itself? You ask: Couldn't the pair of this side and this camera have been differently causally related than the pair of that side and that camera? But, in the situation you've described, I say the "two" pairs are really one pair in close proximity to itself. So of course I can't allow that "one" of the pairs can do something which the "other" pair doesn't do.

B: I don't think *I'm* the one begging the question here. Let's simplify the example a bit. Suppose that nothing exists save two electrons – or, if you like, that the same bundle of electron-ish properties appears on opposite sides of a symmetrical universe. Suppose further that electrons obey indeterministic laws. In that case, even though the electron on the one side is now indiscernible from the one on the other, it remains possible that differences will emerge later on – in other words, it is possible that *this* one have a future differing from *that* one. And even in the case of an eternally symmetrical, two-electron universe in which differences never emerge, such differences were nonetheless possible – both logically or metaphysically possible, and physically or causally possible, too. But you cannot recognize this possibility: on your view the "electrons" must *really* be a single bundle, and so nothing could be true of the one but false of the other.

A: I fear we may have reached an impasse; for this objection of yours depends

upon the resolution of another long-standing quarrel between us: namely, how best to analyze statements involving necessity and possibility.

B: You're still peddling your "counterpart theory," I imagine?

A: Naturally. And your argument tacitly assumes the falsity of a counterpart-theoretic analysis of what it means to say that something is necessarily so-and-so, or possibly such-and-such.⁴ On my view, the possibilities open to a given object are not determined by what *it* does in other possible worlds, but rather by what its *counterparts* do in other possible worlds – and a counterpart is similar to, but not identical with, the original object. Take the fact that I could have been a doctor. Being a doctor represents a real possibility for me just because there is a doctor in some merely possible world who is more like me than anyone else there – he is my "counterpart" in that world. A world with twins who tie in resembling me more than any others in their world is one in which I have two counterparts; there are two possibilities open for me in such a world – each of the twins is a "possible me."⁵ I can say something similar about the statement "The (so-called) 'two' electrons could have diverged": this statement is true just in case the single bi-located electron-bundle in the symmetrical world has *two* counterparts inhabiting some other world, and the two counterparts differ there in the required ways.

B: Well, you know what I think of your counterpart semantics for modal ascriptions. But even if I grant you that, I don't see how it helps. Suppose I let you say that the one bi-located electron could have been either of its two counterpart electrons in the divergent world. Still, something more is possible in the symmetrical world: the electron on the one side could have developed differently while the one on the other side did not. But if "they" are identical, "they" must have the same counterparts in every possible situation – and so there's no possible world in which the one *but not the other* has a counterpart with a particular future.

A: Perhaps counterpart theory by itself doesn't dissolve the problem entirely. But the more I think about this alleged possibility, the less troubled I am by it. Who says it has to be possible for the "one" electron to change its state without "the other" doing so as well? I say they're the same bundle; so when I think about it, I have a hard time even imagining what you're talking about.

B: Just another case of philosophical theory corrupting modal judgment. Let me try one last time to help you grasp the eminently possible situation I'm thinking about: there are just two electrons, they are and remain exactly alike, and the behavior of each one evolves independently of the other. Think of them as being far apart, and moving away from each other at a constant rate. Now, given that each behaves in accordance with slightly indeterministic laws, there are ways in which each one *could* come to differ from the other in the future, although in fact they remain in synch. But there is no way you can allow for the *possibility* of the one doing something that the other does not.

A: I'll admit that there is something a little bit odd about ruling out the alleged possibilities you describe. But why can't I just hold a bundle theory for the objects in the *actual* world? After all, it is only in these bizarrely symmetrical universes that problems arise.

B: So you want to restrict your thesis to just those worlds that lack distinct indiscernibles?

A: Something like that, yes.

B: Doesn't the *ad hoc* nature of the restriction bother you? Pick a couple of almost indiscernible particles in the actual world. Couldn't they (or at least a pair just like them) exist in a world by themselves, where they remain distinct because of some small change which the one undergoes but the other does not?

A: As long as they remain distinct bundles of universals, I can't see that your earlier objections have any foothold.

B: Not in that world. But we are only a small step from *another* world where the one fails to undergo this little change. Suddenly, you have to give up your metaphysics of pure bundles, and posit underlying substrata or some such things. Surely it is implausible to suppose that a tiny change in the global distribution of intrinsic properties would require a radical change in ontology!⁶

A: That does sound a bit unsatisfactory. But your objections have begun to seem less and less pressing to me. Perhaps there is really no need to retreat to a contingent version of the bundle theory. Call me "modally challenged" if you like, but I'm no longer at all sure that the two independently evolving but indiscernible electrons you describe really are possible; it's not obvious to me that, given that the "two" really are indiscernible, one of them *could* behave in a certain way while the other does not.

B: Well, it would be obvious to you, if you weren't such a devotee of the bundle theory.

Notes

- 1 Max Black, "The Identity of Indiscernibles," in his *Problems of Analysis* (Ithaca, N. Y.: Cornell University Press, 1954), p. 83.
- 2 Compare Richard Gale's epithet, "modally other-abled" in Gale, "Some Difficulties in Theistic Treatments of Evil," in *The Evidential Argument from Evil*, ed. Daniel Howard-Snyder (Bloomington and Indianapolis: Indiana University Press, 1996), p. 213.
- 3 This response to Black's spheres is given by John O'Leary-Hawthorne, "The Bundle Theory of Substance and the Identity of Indiscernibles," *Analysis*, 55 (1995), pp. 191–6.

- 4 David Lewis defends counterpart theory in “Counterpart Theory and Quantified Modal Logic,” reprinted in his *Philosophical Papers*, vol. 1 (Oxford: Oxford University Press, 1983). I hasten to add that *A* is not intended to resemble Lewis in any other respect.
- 5 Compare David Lewis’s strategy for introducing haecceitistic modal differences for joint possibilities in his *On the Plurality of Worlds* (Oxford: Blackwell, 1986), pp. 230–3.
- 6 This argument comes from Robert M. Adams, “Primitive Thisness and Primitive Identity,” *Journal of Philosophy*, 76 (1979), pp. 5–26; cf. also D. M. Armstrong, *Universals: an Opinionated Introduction* (Boulder, Col.: Westview Press, 1989), pp. 64–70.

What is Time? What is Space?

5 Time: an Excerpt from *The Nature of Existence** ---

J. McT. E. McTaggart

303. It will be convenient to begin our enquiry by asking whether anything existent can possess the characteristic of being in time. I shall endeavour to prove that it cannot.

It seems highly paradoxical to assert that time is unreal, and that all statements which involve its reality are erroneous. Such an assertion involves a departure from the natural position of mankind which is far greater than that involved in the assertion of the unreality of space or the unreality of matter. For in each man's experience there is a part – his own states as known to him by introspection – which does not even appear to be spatial or material. But we have no experience which does not appear to be temporal. Even our judgments that time is unreal appear to be themselves in time.

304. Yet in all ages and in all parts of the world the belief in the unreality of time has shown itself to be singularly persistent . . .

305. 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. To constitute such a series there is required a transitive asymmetrical relation, and a collection of terms such that, of any two of them, either the first is in this relation to the second, or the second is in this relation to the first. We may take here either the relation of 'earlier than' or the relation of 'later than', both of which, of course, are transitive and asymmetrical. If we take the first, then the terms have to be such that, of any two of them, either the first is earlier than the second, or the second is earlier than the first.

In the second place, each position is either Past, Present, or Future. The distinctions of the former class are permanent, while those of the latter 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.

306. Since distinctions of the first class are permanent, it might be thought that they were more objective, and more essential to the nature of time, than

* From J. McT. E. McTaggart, *The Nature of Existence*, vol. II (Cambridge: Cambridge University Press, 1927). Reprinted with the permission of the Cambridge University Press.

those of the second class. I believe, however, that this would be a mistake, and that the distinction of past, present, and future is as *essential* to time as the distinction of earlier and later, while in a certain sense it may . . . be regarded as more *fundamental* than the distinction of earlier and later. And it is because the distinctions of past, present, and future seem to me to be essential for time, that I regard time as unreal.

For the sake of brevity I shall give the name of the *A* series to that series of positions which runs from the far past through the near past to the present, and then from the present through the near future to the far future, or conversely. The series of positions which runs from earlier to later, or conversely, I shall call the *B* series. The contents of any position in time form an event. The varied simultaneous contents of a single position are, of course, a plurality of events. But, like any other substance, they form a group, and this group is a compound substance. And a compound substance consisting of simultaneous events may properly be spoken of as itself an event.¹

307. The first question which we must consider is whether it is essential to the reality of time that its events should form an *A* series as well as a *B* series. It is clear, to begin with, that, in present experience, we never *observe* events in time except as forming both these series. We perceive events in time as being present, and those are the only events which we actually perceive. And all other events which, by memory or by inference, we believe to be real, we regard as present, past, or future. Thus the events of time as observed by us form an *A* series.

308. It might be said, however, that this is merely subjective. It might be the case that the distinction of positions in time into past, present, and future, is only a constant illusion of our minds, and that the real nature of time contains only the distinctions of the *B* series – the distinctions of earlier and later. In that case we should not perceive time as it really is, though we might be able to *think* of it as it really is.

This is not a very common view, but it requires careful consideration. I believe it to be untenable, because, as I said above, it seems to me that the *A* series is essential to the nature of time, and that any difficulty in the way of regarding the *A* series as real is equally a difficulty in the way of regarding time as real.

309. It would, I suppose, be universally admitted that time involves change. In ordinary language, indeed, we say that something can remain unchanged through time. But there could be no time if nothing changed. And if anything changes, then all other things change with it. For its change must change some of their relations to it, and so their relational qualities. The fall of a sand-castle on the English coast changes the nature of the Great Pyramid.

If, then, a *B* series without an *A* series can constitute time, change must be possible without an *A* series. Let us suppose that the distinctions of past, present, and future do not apply to reality. In that case, can change apply to reality?

310. What, on this supposition, could it be that changes? Can we say that, in a time which formed a *B* series but not an *A* series, the change consisted in the fact that the event ceased to be an event, while another event began to be an event? If this were the case, we should certainly have got a change.

But this is impossible. If *N* is ever earlier than *O* and later than *M*, it will always be, and has always been, earlier than *O* and later than *M*, since the relations of earlier and later are permanent. *N* will thus always be in a *B* series. And as, by our present hypothesis, a *B* series by itself constitutes time, *N* will always have a position in a time-series, and always has had one. That is, it always has been an event, and always will be one, and cannot begin or cease to be an event.

Or shall we say that one event *M* merges itself into another event *N*, while still preserving a certain identity by means of an unchanged element, so that it can be said, not merely that *M* has ceased and *N* begun, but that it is *M* which has become *N*? Still the same difficulty recurs. *M* and *N* may have a common element, but they are not the same event, or there would be no change. If, therefore, *M* changed into *N* at a certain moment, then at that moment, *M* would have ceased to be *M*, and *N* would have begun to be *N*. This involves that, at that moment, *M* would have ceased to be an event, and *N* would have begun to be an event. And we saw, in the last paragraph, that, on our present hypothesis, this is impossible.

Nor can such change be looked for in the different moments of absolute time, even if such moments should exist. For the same argument will apply here. Each such moment will have its own place in the *B* series, since each would be earlier or later than each of the others. And, as the *B* series depends on permanent relations, no moment could ever cease to be, nor could it become another moment.

311. Change, then, cannot arise from an event ceasing to be an event, nor from one event changing into another. In what other way can it arise? If the characteristics of an event change, then there is certainly change. But what characteristics of an event can change? It seems to me that there is only one class of such characteristics. And that class consists of the determinations of the event in question by the terms of the *A* series.

Take any event – the death of Queen Anne, for example – and consider what changes can take place in its characteristics. That it is a death, that it is the death of Anne Stuart, that it has such causes, that it has such effects – every characteristic of this sort never changes. ‘Before the stars saw one another plain,’ the event in question was the death of a Queen. At the last moment of time – if time has a last moment – it will still be the death of a Queen. And in every respect but one, it is equally devoid of change. But in one respect it does change. It 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.²

Such characteristics as these are the only characteristics which can change. And, therefore, if there is any change, it must be looked for in the *A* series, and

in the *A* series alone. If there is no real *A* series, there is no real change. The *B* series, therefore, is not by itself sufficient to constitute time, since time involves change.

312. The *B* series, however, cannot exist except as temporal, since earlier and later, which are the relations which connect its terms, are clearly time-relations. So it follows that there can be no *B* series when there is no *A* series, since without an *A* series there is no time.

313. We must now consider three objections which have been made to this position. The first is involved in the view of time which has been taken by Mr Russell, according to which 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. Thus it is only past, present, or future, in relation to some assertion. If there were no consciousness, there would be events which were earlier and later than others, but nothing would be in any sense past, present, or future. And if there were events earlier than any consciousness, those events would never be future or present, though they could be past.

If *N* were ever present, past, or future in relation to some assertion *V*, it would always be so, since whatever is ever simultaneous to, earlier than, or later than, *V*, will always be so. What, then, is change? We find Mr Russell's views on this subject in his *Principles of Mathematics*, Section 442. '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.' That is to say, there is change, on Mr Russell's view, if the proposition 'at the time *T* my poker is hot' is true, and the proposition 'at the time *T'* my poker is hot' is false.

314. I am unable to agree with Mr Russell. I should, indeed, admit that, when two such propositions were respectively true and false, there would be change. But then I maintain that there can be no time without an *A* series. If, with Mr Russell, we reject the *A* series, it seems to me that change is essential, goes with it, and that therefore time, for which change is essential, goes too. In other words, if the *A* series is rejected, no proposition of the type 'at the time *T* my poker is hot' can ever be true, because there would be no time.

315. It will be noticed that Mr Russell looks for change, not in the events in the time-series, but in the entity to which those events happen, or of which they are states. If my poker, for example, is hot on a particular Monday, and never before or since, the event of the poker being hot does not change. But the poker changes, because there is a time when this event is happening to it, and a time when it is not happening to it.

But this makes no change in the qualities of the poker. It is always a quality of

that poker that it is one which is hot on that particular Monday. And it is always a quality of that poker that it is one which is not hot at any other time. Both these qualities are true of it at any time – the time when it is hot and the time when it is cold. And therefore it seems to be erroneous to say that there is any change in the poker. The fact that it is hot at one point in a series and cold at other points cannot give change, if neither of these facts change – and neither of them does. Nor does any other fact about the poker change, unless its presentness, pastness, or futurity change.

316. Let us consider the case of another sort of series. The meridian of Greenwich passes through a series of degrees of latitude. And we can find two points in this series, *S* and *S'*, such that the proposition ‘at *S* the meridian of Greenwich is within the United Kingdom’ is true, while the proposition ‘at *S'* the meridian of Greenwich is within the United Kingdom’ is false. But no one would say that this gave us change. Why should we say so in the case of the other series?

Of course there is a satisfactory answer to this question if we are correct in speaking of the other series as a time-series. For where there is time, there is change. But then the whole question is whether it is a time-series. My contention is that if we remove the *A* series from the *prima facie* nature of time, we are left with a series which is not temporal, and which allows change no more than the series of latitudes does.

317. If, as I have maintained, there can be no change unless facts change, then there can be no change without an *A* series. For, as we saw with the death of Queen Anne, and also in the case of the poker, no fact about anything can change, unless it is a fact about its place in the *A* series. Whatever other qualities it has, it has always. But that which is future will not always be future, and that which was past was not always past.

It follows from what we have said that there can be no change unless some propositions are sometimes true and sometimes false. This is the case of propositions which deal with the place of anything in the *A* series – ‘the battle of Waterloo is in the past’, ‘it is now raining’. But it is not the case with any other propositions.

318. Mr Russell holds that such propositions are ambiguous, and that to make them definite we must substitute propositions which are always true or always false – ‘the battle of Waterloo is earlier than this judgment’, ‘the fall of rain is simultaneous with this judgment’. If he is right, all judgments are either always true, or always false. Then, I maintain, no facts change. And then, I maintain, there is no change at all.

I hold, as Mr Russell does, that there is no *A* series. (My reasons for this will be given below. And I regard the reality lying behind the appearance of the *A* series in a manner not completely unlike that which Mr Russell has adopted. The difference between us is that he thinks that, when the *A* series is rejected, change, time, and the *B* series can still be kept, while I maintain that its

rejection involves the rejection of change, and, consequently, of time, and of the *B* series . . .

324. We conclude, then, that the distinctions of past, present, and future are essential to time, and that, if the distinctions are never true of reality, then no reality is in time. . . .

325. I now pass to the second part of my task. Having, as it seems to me, succeeded in proving that there can be no time without an *A* series, it remains to prove that an *A* series cannot exist, and that therefore time cannot exist. This would involve that time is not real at all, since it is admitted that the only way in which time can be real is by existing. . . .

329. Past, present, and future are incompatible determinations. Every event must be one or the other, but no event can be more than one. If I say that any event is past, that implies that it is neither present nor future, and so with the others. And this exclusiveness is essential to change, and therefore to time. For the only change we can get is from future to present, and from present to past.

The characteristics, therefore, are incompatible. But every event has them all.³ 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 the three characteristics belong to each event. How is this consistent with their being incompatible?

330. It may seem that this can easily be explained. Indeed, it has been impossible to state the difficulty without almost giving the explanation, since our language has verb-forms for the past, present, and future, but no form that is common to all three. 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.

331. But what is meant by 'has been' and 'will be'? And what is meant by 'is', when, as here, it is used with a temporal meaning, and not simply for predication? When we say that *X* has been *Y*, we are asserting *X* to be *Y* at a moment of past time. When we say that *X* will be *Y*, we are asserting *X* to be *Y* at a moment of future time. When we say that *X* is *Y* (in the temporal sense of 'is'), we are asserting *X* to be *Y* at a moment of present time.

Thus our first statement about *M* – that it is present, will be past, and has been future – means that *M* is present at a moment of present time, past at some moment of future time, and future at some moment of past time. But every moment, like every event, is both past, present, and future. And so a similar difficulty arises. If *M* is present, there is no moment of past time at which it is past. But the moments of future time, in which it is past, are equally moments of past time, in which it cannot be past. Again, that *M* is future and will be

present and past means that *M* is future at a moment of present time, and present and past at different moments of future time. In that case it cannot be present or past at any moments of past time. But all the moments of future time, in which *M* will be present or past, are equally moments of past time.

332. And thus again we get a contradiction, since the moments at which *M* has any one of the three determinations of the *A* series are also moments at which it cannot have that determination. If we try to avoid this by saying of these moments what had been previously said of *M* itself – that some moment, for example, is future, and will be present and past – then ‘is’ and ‘will be’ have the same meaning as before. Our statement, then, means that the moment in question is future at a present moment, and will be present and past at different moments of future time. This, of course, is the same difficulty over again. And so on infinitely.

Such an infinity is vicious. The attribution of the characteristics past, present, and future to the terms of any series leads to a contradiction, unless it is specified that they have them successively. This means, as we have seen, that they have them in relation to terms specified as past, present, and future. These again, to avoid a like contradiction, must in turn be specified as past, present, and future. And, since this continues infinitely, the first set of terms never escapes from contradiction at all.⁴

The contradiction, it will be seen, would arise in the same way supposing that pastness, presentness, and futurity were original qualities, and not, as we have decided that they are, relations. For it would still be the case that they were characteristics which were incompatible with one another, and that whichever had one of them would also have the other. And it is from this that the contradiction arises.

333. The reality of the *A* series, then, leads to a contradiction, and must be rejected. And, since we have seen that change and time require the *A* series, the reality of change and time must be rejected. And so must the reality of the *B* series, since that requires time. Nothing is really present, past, or future. Nothing is really earlier or later than anything else or temporally simultaneous with it. Nothing really changes. And nothing is really in time. Whenever we perceive anything in time – which is the only way in which, in our present experience, we do perceive things – we are perceiving it more or less as it really is not. . . .

Notes

- 1 It is very usual to contemplate time by the help of a metaphor of spatial movement. But spatial movement in which direction? The movement of time consists in the fact that later and later terms pass into the present, or – which is the same fact expressed in another way – that presentness passes to later and later terms. If we take it the first way, we are taking the *B* series as sliding along a fixed *A* series. If we take it the second way, we are taking the *A* series as sliding along a fixed *B* series. In the first case time presents itself as a movement from future to past. In the second case it

presents itself as a movement from earlier to later. And this explains why we say that events come out of the future, while we say that we ourselves move towards the future. For each man identifies himself especially with his present state, as against his future or his past, since it is the only one which he is directly perceiving. And this leads him to say that he is moving with the present towards later events. And as those events are now future, he says that he is moving towards the future.

Thus the question as to the movement of time is ambiguous. But if we ask what is the movement of either series, the question is not ambiguous. The movement of the *A* series along the *B* series is from earlier to later. The movement of the *B* series along the *A* series is from future to past.

- 2 The past, therefore, is always changing, if the *A* series is real at all, since at each moment a past event is further in the past than it was before. This result follows from the reality of the *A* series, and is independent of the truth of our view that all change depends exclusively on the *A* series. It is worth while to notice this, since most people combine the view that the *A* series is real with the view that the past cannot change – a combination which is inconsistent.
- 3 If the time-series has a first term, that term will never be future, and if it has a last term, that term will never be past. But the first term, in that case, will be present and past, and the last term will be future and present. And the possession of two incompatible characteristics raises the same difficulty as the possession of three.
- 4 It may be worth while to point out that the vicious infinite has not arisen from the impossibility of *defining* past, present, and future, without using the terms in their own definitions. On the contrary, we have admitted these terms to be indefinable. It arises from the fact that the nature of the terms involves a contradiction, and that the attempt to remove the contradiction involves the employment of the terms, and the generation of a similar contradiction.

6 McTaggart's Arguments against the Reality of Time: an Excerpt from *Examination of McTaggart's Philosophy**¹

C. D. Broad

We come at last to McTaggart's destructive arguments. . . .

3.1. The Main Argument. We take as an established premise that any series which could count as a temporal series would have to consist of terms which have *A*-characteristics and which individually change in respect of their *A*-characteristics. McTaggart tries to prove that there is a contradiction involved

* From C.D. Broad, *Examination of McTaggart's Philosophy*, vol. II, part I (Cambridge: Cambridge University Press, 1938). Reprinted with the permission of the Cambridge University Press.

in this condition, and therefore that nothing could be a temporal series. If he is right, then, the characteristic of being a *B*-series, i.e., a series in which the terms are events and the relation is that of 'earlier than', is a delusive characteristic.

The essence of the argument is as follows: (i) The various determinate *A*-characteristics are incompatible with each other, in the usual way in which different determinates under the same determinable are so. McTaggart confines his statement to past, present, and future. But, of course, if it is true at all, it is equally true of any two degrees of pastness or of futurity.

(ii) Every event has all the *A*-characteristics; for every event has all degrees of futurity, has presentness, and has all degrees of pastness. The only possible exceptions would be that last event, if there were one, and the first event, if there were one. But, even so, the last event would have presentness and all degrees of futurity, though it would not have pastness. And the first event would have presentness and all degrees of pastness, though it would not have futurity. Thus every event has a plurality of determinate *A*-characteristics, whilst no two *A*-characteristics are compatible with each other.

(iii) McTaggart admits that, at first sight, this seems to lead to no difficulty. After all, no event has two different *A*-characteristics at any *one* moment; though each event has a different *A*-characteristic at each different moment.

(iv) McTaggart claims to show, however, that this attempted answer is useless, because it leads either to a contradiction or to a vicious infinite regress. His argument is as follows.

Suppose we try to avoid the contradiction of a term *M* being past, present, and future by saying that *M* is now present, will be past, and has been future; or by saying that *M* is now future, will be present, and will be past; or by saying that *M* is now past, has been present, and has been future. We must then raise the question of what we mean by these temporal copulas. According to McTaggart, there is only one possible analysis. To say that *S* has been *P* means 'There is a moment *t*, such that *S* has *P* at *t* and *t* is past.' To say that *S* is now *P* means 'There is a moment *t*, such that *S* has *P* at *t* and *t* is present.' To say that *S* will be *P* means 'There is a moment *t*, such that *S* has *P* at *t* and *t* is future.'

Now substitute *M* for *S*, and substitute the *A*-characteristics for *P*. We get the following results. '*M* is now present' means 'There is a moment *t*, such that *M* has presentness at *t* and *t* is present.' Again, '*M* will be past' means 'There is a moment *t*, such that *M* has pastness at *t* and *t* is future.' Lastly, '*M* has been future' means 'There is a moment *t*, such that *M* has futurity at *t* and *t* is past.'

The next stage of the argument will be found in the last paragraph of §331. It is very difficult to follow, as stated by McTaggart; but I have no doubt as to what is the essential point of it. I shall first quote McTaggart's argument in his own words, and shall then restate in my own way what is substantially the same argument. McTaggart's statement runs as follows: '... every moment, like every event, is both past, present, and future. . . . If *M* is present, there is no moment of past time at which it is past. But the moments of future time, in which it is past, are equally moments of past time, in which it cannot be past. Again, that *M* is future and will be present and past means that *M* is future at a moment of present time, and present and past at different moments of future time. In

that case it cannot be present or past at any moments of past time. But all the moments of future time, in which M will be present or past, are equally moments of past time.'

I will now try to put the essential points of this very obscure argument clearly. The question is whether the three propositions ' M is now present, M has been future, and M will be past' are mutually compatible. McTaggart wants to show that they are not. (a) Consider the proposition ' M will be past.' According to McTaggart, this means 'There is a moment t , such that M has pastness at t and t is future.' But, according to him, any moment that is future is also *present*. Therefore it follows that there is a moment t , such that M has pastness at t and t is *present*. But this is equivalent to the proposition ' M is now past.' This is incompatible with the proposition ' M is now present.' Thus ' M will be past' entails ' M is now past', and the latter is inconsistent with ' M is now present.' Therefore ' M will be past' is inconsistent with ' M is now present.'

(b) Now consider the proposition ' M has been future.' According to McTaggart, this means 'There is a moment t , such that M has futurity at t and t is past.' But, according to him, any moment that is past is also *present*. Therefore it follows that there is a moment t , such that M has futurity at t and t is *present*. But this is equivalent to the proposition ' M is now future.' This is incompatible with the proposition ' M is now present.' Thus ' M has been future' entails ' M is now future', and the latter is inconsistent with ' M is now present.' Therefore ' M has been future' is inconsistent with ' M is now present.'

(c) If the argument in paragraphs (a) and (b) were valid, it would have proved that both the propositions ' M will be past' and ' M has been future' are inconsistent with the proposition ' M is now present.' It remains to show that these two propositions are inconsistent with *each other*. This is easily done. From the argument in paragraph (a) we conclude that ' M will be past' entails ' M is now past.' From the argument in paragraph (b) we conclude that ' M has been future' entails ' M is now future.' But the two propositions ' M is now past' and ' M is now future' are incompatible with each other. Therefore the two proposition ' M will be past' and ' M has been future' are incompatible with each other. Thus, if the argument is valid, it would prove that *each* of the three propositions ' M is now present', ' M has been future', and ' M will be past' is incompatible with the other two. I believe this to be a fair and clear statement of the line of argument which McTaggart had in mind in the last paragraph of §331.

If we had started, instead, with the three propositions ' M is now past, M has been present, and M has been future', or ' M is now future, M will be present, and M will later on be past', a similar argument would have led to a similar result. So McTaggart claims to have shown that the original contradiction of M being past, present, and future breaks out again in the amended statement that M is now present, has been future, and will be past; and in the amended statement that M is now past and has been present, and future; and in the amended statement that M is now future and will be present and past.

(v) Of course there is *prima facie* a perfectly simple answer to this alleged contradiction, which McTaggart mentions in §332. Instead of admitting in paragraph (a) above that the future moment at which M has pastness is also

present, we ought only to have admitted that it *will be* present. And, instead of admitting in paragraph (b) above that the past moment at which *M* has futurity is also present, we ought only to have admitted that it *has been* present. The argument would then have broken down at the first move.

McTaggart rejects this answer on the following grounds. According to him, we shall have to analyse the statement that a certain *moment* is now present, has been future, and will be past, in a similar way to that in which we analysed the corresponding statements about the *event M*. To say that *t* *will be* present, e.g., must mean that there is a moment *t'* such that *t* has presentness at *t'* and *t'* is future. To say that *t* *has been* present must mean that there is a moment *t'*, such that *t* has presentness at *t'* and *t'* is past. Thus the same contradiction will arise at the second stage about *moments* as arose at the first stage about *events*. Any attempt to remove it in the same way will merely lead to a third stage at which the same contradiction will break out. We start on an infinite regress; which is vicious, because each step is needed in order to remove a contradiction in the previous stage, and at each stage the same contradiction breaks out again.

This is the main argument by which McTaggart persuaded himself that nothing can have *A*-characteristics. If nothing can have them, nothing can change in respect of them. If nothing can change in respect of *A*-characteristics, there can be no processes of qualitative change. And, if there can be no processes of qualitative change, no series can be a *B*-series. And so neither *A*-characteristics, nor *B*-relations, nor qualitative change or persistence, can apply to anything. All these ostensible characteristics are delusive.

3.11. Criticism of the Main Argument. We must now consider whether this argument of McTaggart's is valid. I should suppose that every reader must have felt about it as any healthy-minded person feels about the Ontological Argument for the existence of God, viz., that it is obviously wrong somewhere, but that it may not be easy to say precisely what is wrong with it.

(i) I cannot myself see that there is any contradiction to be avoided. When it is said that pastness, presentness, and futurity are incompatible predicates, this is true only in the sense that no one term could have two of them *simultaneously* or *timelessly*. Now no term ever appears to have any of them timelessly, and no term ever appears to have any two of them simultaneously. What appears to be the case is that certain terms have them *successively*. Thus there is nothing in the temporal appearances to suggest that there is a contradiction to be avoided.

(ii) What are we to say, then, about McTaggart's alleged vicious infinite regress? In the first place we must say that, since there is no contradiction to be avoided, there is no need to start on any regress in order to avoid a contradiction. Secondly, we may well ask why McTaggart should assume that, e.g., '*M* is now present' *must* be analysed into 'There is a moment *t*, such that *M* has presentness at *t* and *t* is present.' Similarly, we may ask why he should assume that, e.g., 'The moment *t* has been future' *must* be analysed into 'There is a moment *t'*, such that *t* has futurity at *t'* and *t'* is past.'

(a) In the first place, we note that McTaggart has suddenly introduced the notion of *moments*, in addition to that of *events*. No justification whatever has

been given for this. It would seem to imply that the temporal copulas ‘is now’, ‘has been’, and ‘will be’ presuppose some form of the Absolute Theory of Time. This is surely not obvious.

(b) The real motive of this analysis, and the real cause of the subsequent infinite regress, seems to me to be a certain assumption which McTaggart tacitly makes. He assumes that what is meant by a sentence with a *temporal copula* must be completely (and more accurately) expressible by a sentence or combination of sentences in which there is no temporal copula, but only *temporal predicates* and non-temporal copulas. And the regress arises because there remains at every stage a copula which, if taken as non-temporal, involves the *non-temporal* possession by a term of certain temporal predicates which could belong to it only *successively*.

Take, e.g., the general analysis of ‘*S* is now *P*’ into ‘There is a moment *t*, such that *S* has *P* at *t* and *t* is present.’ The only motive for making this analysis is that it seems at first sight to have got rid of the temporal copula ‘is now’. The predicate ‘having *P* at *t*’ may be said to belong to *S* timelessly or sempiternally if it belongs to *S* at all. And we are tempted to think that the ‘is’ in ‘*t* is present’ is a timeless copula too. Now the source of McTaggart’s regress is that, if you take the ‘is’ in ‘*t* is present’ to be timeless, you will have to admit that *t* is also past and future in the same timeless sense of ‘is’. Now this is impossible, for it is obvious that *t* can have these predicates only in succession. If, to avoid this, you say that the ‘is’ in ‘*t* is present’ means ‘is now’, you have not got rid of temporal copulas. Therefore, if you are committed at all costs to getting rid of them, you will not be able to rest at this stage. At every stage of the analysis you will have a copula which, if taken to be *non-temporal*, leads to a contradiction, and, if taken to be *temporal*, needs to be analysed further in terms of temporal predicates and non-temporal copulas.

Now it seems to me that the proper interpretation of the regress is that it disproves the assumption that temporal copulas can be replaced by temporal predicates and non-temporal copulas. Since there is nothing necessary or self-evident about this assumption, the regress raises no objection to the *prima facie* appearance that events become and pass away and that they stand to each other in relations of temporal sequence and simultaneity.

(iii) It may be worth while to go into a little more detail about the question of temporal copulas and temporal predicates before leaving this topic. Let us take the sentences ‘It will rain’, ‘It is now raining’, and ‘It has rained.’ The utmost that can be done with the first is to analyse it into ‘There is (in some non-temporal sense of “is”) an event characterised non-temporally by raininess, and it is now future.’ The corresponding analyses of the second and third would be got by substituting ‘it is now present’ and ‘it is now past’, respectively, for ‘it is now future’ in the analysis of the first. Even if this kind of analysis be accepted as correct, we have not got rid of the temporal copula ‘is now’.

Another type of analysis would be to make ‘It will rain’ equivalent to ‘There is (in some non-temporal sense of “is”) an event characterised non-temporally by raininess, and it will be present.’ The corresponding analyses of the second and third would be got by substituting ‘it is now present’ and ‘it has been

'present', respectively, for 'it will be present' in the analysis of the first. Here we get rid of two out of the three *A*-characteristics, but have to keep all three temporal copulas. In the previous kind of analysis we got rid of two out of the three temporal copulas, but had to keep all three *A*-characteristics. So, on neither kind of analysis, can we get rid of *all* temporal copulas; and, on both kinds of analysis, we have to introduce at least one temporal predicate in addition to temporal copulas. Now the original sentences 'It will rain', 'It is now raining', and 'It has rained' express the facts in the most natural and simple way without introducing temporal predicates in addition to temporal copulas. So both kinds of analysis seem to be worthless. They complicate instead of simplifying; they make nothing intelligible which was not intelligible before; and they suggest false analogies with non-temporal propositions.

Quite apart from the fact that such 'analyses' serve no useful purpose, it seems to me that they fail to express what we have in mind when we use such sentences as 'It has rained' or 'It will rain.' When I utter the sentence 'It has rained', I do *not* mean that, in some mysterious non-temporal sense of 'is', there *is* a rainy event, which momentarily possessed the quality of presentness and has now lost it and acquired instead some determinate form of the quality of pastness. What I mean is that raininess has been, and no longer is being, manifested in my neighbourhood. When I utter the sentence 'It will rain', I do *not* mean that, in some mysterious non-temporal sense of 'is', there *is* a rainy event, which now possesses some determinate form of the quality of futurity and will in course of time lose futurity and acquire instead the quality of presentness. What I mean is that raininess will be, but is not now being, manifested in my neighbourhood.

The fact is that what are called 'statements about past events' are statements to the effect that certain characteristics, which constitute descriptions of possible events, have been and no longer are being manifested. What are called 'statements about future events' are statements to the effect that certain characteristics, which constitute descriptions of possible events, will be but are not yet being manifested.

To sum up. I believe that McTaggart's main argument against the reality of Time is a philosophical 'howler' of the same kind as the Ontological Argument for the existence of God. The fallacy of the Ontological Argument consists in treating being or existence as if it were a predicate like goodness, and in treating instantial propositions as if they were characterising propositions. The fallacy in McTaggart's argument consists in treating absolute becoming as if it were a species of qualitative change, and in trying to replace temporal copulas by non-temporal copulas and temporal adjectives. Both these 'howlers', like the Fall of Adam, have been over-ruled to good ends. In each case one can see that there is something radically wrong with the argument; and one's desire to put one's finger on the precise point of weakness stimulates one to clear up linguistic confusions which would otherwise have remained unnoticed and unresolved. I suspect that plenty of other philosophers have made the same mistake as St Anselm and the same mistake as McTaggart. But, since they did not draw such startling consequences from their confusions as these eminent men did, these errors have been allowed to rest in decent obscurity.

7 The Notion of the Present*

A. N. Prior

Before directly discussing the notion of the present, I want to discuss the notion of the real. These two concepts are closely connected; indeed on my view they are one and the same concept, and the present simply is the real considered in relation to two particular species of unreality, namely the past and the future. So let's begin with the real in general.

Philosophers often speak as if the real world were just one of a number of different big boxes in which various things go on, the other boxes having such labels as 'the mind' or 'the world of Greek mythology'. For example, centaurs exist in the world of Greek mythology but not in the real world, aeroplanes exist in the real world but not in the world of Greek mythology, and horses and men exist both in the real world and in the world of Greek mythology. Again, Anselm addresses himself to people who held that God does not exist in the real world but only in the mind, and claimed to have a proof that if God exists in the mind he must exist in the real world too. Leibniz contrasted the real or actual world with an infinity of merely possible worlds in which various things happen which do not happen in the actual world. All these ways of talking suggest that the real world or the actual world is just a *region* of some larger universe which contains other regions as well – possible worlds, imaginary worlds, and so on.

I want to suggest – I don't of course claim that there's anything original in this suggestion – that this way of conceiving the relation between the real and the unreal is profoundly mistaken and misleading. The most important way in which it is misleading is that it minimises, or makes a purely arbitrary matter, the vast and stark *difference* that there is between the real and every form of unreality. For talking of the real as one 'region' among others immediately suggests the question, 'In that case, what is so special about the real world in contrast with all other regions? – is it not a kind of narrow-mindedness and parochialism to think that it has anything special about it that none of the others have?' One philosopher, Meinong, has indeed said precisely that it *is* just narrow-mindedness and parochialism to single out the real world as a region of special interest; the 'prejudice in favour of the actual', he called it. Well, I want to argue that this is *not* just narrow-mindedness and parochialism, and that it becomes obvious enough what is so special about the real world as soon as we drop this metaphor of boxes or regions and become a little more literal.

To say that there are centaurs in the world of Greek mythology is surely *not* to say that there are centaurs in some remote and peculiar region, but just to say

* From A. N. Prior, 'The Notion of the Present', *Studium Generale*, 23 (1970), pp. 245–8. Reprinted by permission of Springer-Verlag New York, Inc.

that Greek myth-makers have said that there are centaurs. Similarly, to say that there are centaurs in some person's mind is to say that that person thinks or imagines that there are centaurs. And to say that there are possible worlds in which there are centaurs is just to say that it could be that there are centaurs. In general, to say that X is the case in some non-real world is just to say ' X is the case' with some modifying prefix like 'Greek myth-makers have said that', 'Jones imagines that', or 'It could be that'. But to say that X is the case in the real or the actual world, or that it is really or actually or in fact the case, is just to say that it is the case – flat, and without any prefix whatever. To say that there are centaurs in the real world, for example, is not to say that there are centaurs in some region of the universe in which we happen to have more interest than in others; it is simply to say that there are centaurs. Talk of the real world, in other words, is not a metaphorical fudging-up of talk in which our sentences have a special kind of prefix, but a fudging-up of talk in which the relevant sentences have no prefixes at all. 'Really', 'actually', 'in fact', 'in the real world' are strictly redundant expressions – that, and not any prejudice or provincialism, is their specialness.

So to say that although there are no centaurs in the real world there are some in the world of Greek mythology, is just to say that although there are no centaur's Greek myth-makers have said that there are; to say that although God does not exist in reality he exists in the mind, is just to say that although God does not exist people may imagine that he does; to say that although Sextus raped Lucretia in the real world there is a possible world in which he didn't, is just to say that although Sextus raped Lucretia he need not have done so. There is, if you like, no other place than the real world for God or centaurs to exist in or for Sextus to rape Lucretia in; for God or centaurs to exist in the real world, or for Sextus to rape Lucretia in the real world, is just for God or centaurs to exist, or for Sextus to rape Lucretia. Again, 'Greek myth-makers have said that there are centaurs in the real world' is all one with 'Greek myth-makers have said that there are centaurs', and so is 'Greek myth-makers in the real world have said that there are centaurs.'

And now the present. It is tempting to think of the present as a region of the universe in which certain things happen, such as the war in Vietnam, and the past and the future as other regions in which other things happen, such as the battle of Hastings and men going to Mars. But to this picture there is the same objection as to the picture of the 'real world' as a box or region among other boxes or regions. It doesn't bring out what is so special about the present; and to be more specific, it doesn't bring out the way in which the present is real and the past and future are not. And I want to suggest that the reality of the present consists in what the reality of anything else consists in, namely the absence of a qualifying prefix. To say that Whitrow's lecture is past is to say that it has been the case that Whitrow is lecturing. To say that Scott's lecture is future is to say that it will be the case that Scott is lecturing. But to say that my lecture is present is just to say that I am lecturing – flat, no prefixes. The pastness of an event, that is to say its having taken place, is not the same thing as the event itself; nor is its futurity; but the presentness of an event is just the event. The presentness of my

lecturing, for instance, is just my lecturing. Moreover, just as a real thought of a centaur, and a thought of a real centaur, are both of them just a thought of a centaur, so the present pastness of Whitrow's lecture, and its past presentness, are both just its pastness. And conversely, its pastness is its present pastness, so that although Whitrow's lecture isn't now present and so isn't real, isn't a fact, nevertheless its pastness, its *having* taken place, *is* a present fact, *is* a reality, and will be one as long as time shall last.

Notoriously, much of what is present isn't present permanently; the present is a shifting, changing thing. That is only to say that much of what is the case, of what is real and true, is constantly changing. Not everything, of course; some things that are the case also have always been the case and will always be the case. I imagine scientists have a special interest in such things. And among the things that not only are the case but always have been and always will be, are the laws of change themselves, I mean such laws as that if anything *has* occurred then for ever after it *will have* occurred (like Whitrow's lecture). These are the laws of what is now called *tense logic*, and the conception of the present that I have just been suggesting is deeply embedded in the syntax of that discipline. So that conception underlies, or anyhow seems to underlie, what is now a pretty flourishing systematic enterprise. . . .

8 The General Problem of Time and Change: an Excerpt from *Scientific Thought**

C. D. Broad

Alice sighed wearily, 'I think you might do something better with the time,' she said, 'than waste it asking riddles with no answers.'

'If you knew Time as well as I do,' said the Hatter, 'you wouldn't talk about wasting it.'

Lewis Carroll, *Alice in Wonderland*

. . . At first sight the problems of Time look very much like those of Space, except that the single dimension of Time, as compared with the three of Space, seems to promise greater simplicity. We shall point out these analogies at the beginning; but we shall find that they are somewhat superficial, and that Time and Change are extremely difficult subjects, in which spatial analogies help us but little.

The physicist conceives Time in much the same way as he conceives Space.

* From C. D. Broad, *Scientific Thought* (London: Routledge and Kegan Paul, 1923). Reprinted by permission of Routledge.

Just as he distinguishes Space from the matter in it, so he distinguishes Time from events. Again, mere difference of position in Time is supposed to have no physical consequences. It is true that, if I go out without my overcoat at 2 a.m., I shall probably catch cold; whilst, if I do so at 2 p.m., I shall probably take no harm. But this difference is never ascribed to the mere difference in date, but to the fact that different conditions of temperature and dampness will be contemporary with my two expeditions. Again, Time, like Space, is supposed to be continuous, and physicists suppose (or did so until quite lately) that there is a single time-series in which all the events of nature take place. This series is of one dimension, so that, as far as appears at present, Time is like a very simple Space consisting of a single straight line.

Just as we treat our geometry in terms of unextended points and their relations, so we treat our chronometry in terms of moments without duration and *their* relations. Duration in Time corresponds to extension in Space. Now, just as we never perceive points or even unextended particles, so we are never aware of moments or of momentary events. What we are aware of is finite events of various durations. By an event I am going to mean anything that endures at all, no matter how long it lasts or whether it be qualitatively alike or qualitatively different at adjacent stages in its history. This is contrary to common usage, but common usage has nothing to recommend it in this matter. We usually call a flash of lightning or a motor accident an event, and refuse to apply this name to the history of the cliffs at Dover. Now the only relevant difference between the flash and the cliffs is that the former lasts for a short time and the latter for a long time. And the only relevant difference between the accident and the cliffs is that, if successive slices, each of one second long, be cut in the histories of both, the contents of a pair of adjacent slices may be very different in the first case and will be very similar in the second case. Such merely quantitative differences as these give no good ground for calling one bit of history an event and refusing to call another bit of history by the same name. . . .

So far, the analogy between Time and Space has seemed to work very well. Duration has corresponded to length, before and after to right and left, and simultaneity to complete mutual overlapping. But, if we reflect a little more carefully, we shall see that the analogy between before and after and right and left is not so illuminating as it seems at first sight. The peculiarity of a series of events in Time is that it has not only an intrinsic *order* but also an intrinsic *sense*. Three points on a straight line have an intrinsic order, i.e. B is between A and C, or C is between B and A, or A is between C and B. This order is independent of any tacit reference to something traversing the line in a certain direction. By difference of sense I mean the sort of difference which there is between, say, ABC and CBA. Now the points on a straight line do not have an intrinsic sense. A sense is only assigned to them by correlation with the left and right hands of an imaginary observer, or by thinking of a moving body traversing the line in such a way that its presence at A is earlier than its presence at B, and the latter is earlier than its presence at C. In fact, if we want a spatial analogy to Time, it is not enough to use a straight line; we need a straight line with a fixed sense, i.e. the sort of thing which we usually represent by a line with an arrow-head on it.

Now the points on straight lines do not have any intrinsic sense, and so the meaning of the arrow-head is only supplied by reference to something which is at one point *before* it gets to another. Thus to attempt to understand before and after by analogy with a directed line is in the end circular, since the line only gets its sense through a tacit correlation with a series of events in Time.

Now the intrinsic sense of a series of events in Time is essentially bound up with the distinction between past, present, and future. A precedes B because A is past when B is present. . . .

We are naturally tempted to regard the history of the world as existing eternally in a certain order of events. Along this, and in a fixed direction, we imagine the characteristic of presentness as moving, somewhat like the spot of light from a policeman's bull's-eye traversing the fronts of the houses in a street. What is illuminated is the present, what has been illuminated is the past, and what has not yet been illuminated is the future. The fact that the spot is of finite area expresses the fact that the Specious Present is not a mere point but is of finite, though short, duration. Such analogies may be useful for some purposes, but it is clear that they explain nothing. On this view the series of events has an intrinsic order, but no intrinsic sense. It gains a sense, and we become able to talk of one event as earlier than another, and not merely of one event as between two others, because the attribute of presentness moves along the series in a fixed direction. But, in the first place, the lighting of the characteristic of presentness now on one event and now on another is itself an event, and ought therefore to be itself a part of the series of events, and not simply something that happens to the latter from outside. Again, if events have no intrinsic sense but only an intrinsic order, what meaning can we give to the assertion that the characteristic of presentness traverses the series of events *in a fixed direction*? All that we can mean is that this characteristic is *present* at B when it is *past* at A. Thus all the problems which the policeman's bull's-eye analogy was invented to solve are simply taken out of other events to be heaped on that particular series of events which is the movement of the bull's-eye. . . .

The difficulty about past, present, and future in general can be summed up in two closely connected paradoxes. (i) Every event has all these characteristics, and yet they are inconsistent with each other. And (ii) *events* change in course of time with respect to these characteristics. Now we believe ourselves to understand stand change in *things*, but to talk of *events* changing seems almost unintelligible. The connexion between the two paradoxes is, of course, that we get into the second directly we take the obvious step to avoid the first.

We have plenty of experience of things which appear to have incompatible characteristics, such as redness and greenness, or greatness and smallness. As a rule we remove this apparent inconsistency by pointing out that the facts have been stated elliptically, and that really a relation is involved. In the first example we say that what has been omitted is a relation to two different times. The full statement is that the thing is red at one time and green at another, and there is no inconsistency in this. In the second example we have no need even to bring in a relation to two different times. It is enough to point out that the predicates great and small themselves tacitly assume relations; so that the full statement is

that the thing is at once great as compared with one object and small as compared with another. In one of these two ways we always proceed when we have to deal with the apparent co-inherence of incompatible predicates in a single subject. We therefore naturally try one of these expedients to deal with the fact that every event is past, present, and future, and that these predicates are incompatible.

It seems natural and childishly simple to treat the problem in the way in which we treated the thing that was both red and green. We say: 'Of course the event E has futurity for a certain stretch of time, then it has presentness for a short subsequent stretch, and it has pastness at all other moments.' Now the question at once arises: 'Can we treat the change of an *event* in respect to its *temporal* qualities as just like the change of a *thing* with respect to qualities like red and green?'

To answer this question we must try to see what we mean when we say that a certain thing T changes from red to green. So far as I can see, our meaning is somewhat as follows: There is a certain long-lasting event in the history of the world. This stands out in a noticeable way from other events which overlap it wholly or partly. If successive short sections in time be taken of this long event, adjacent sections have spatial continuity with each other, and predominant qualitative resemblance to each other. On these grounds the whole long event is treated as the history of a single thing T. But, although adjacent short sections are *predominantly* alike in their qualities, there may be adjacent sections which differ very markedly in *some* quality, such as colour. If you can cut the history of the thing in a certain moment, such that a slice of its history before that is red and a slice after that is green, we say that the thing T has changed from red to green at that moment. To say that a thing changes, thus simply means that its history can be cut up into a series of adjacent short slices, and that two adjacent slices may have qualitative differences.

Can we treat the change of an event from futurity, through presentness, to pastness in the way in which we have treated the change of a thing (say a signal lamp) from red to green? I think it is certain that we cannot; for two closely connected reasons. In the first place, the attempt would be circular, because the change of things will be found on further analysis to involve the change of events in respect to their temporal characteristics. We have assumed that the history of our signal lamp can be analysed into a series of shorter adjacent events, and that it was true of a certain pair of these that the earlier was red and the later green. But to say that this series of events passes from earlier to later (which is necessary if we are to distinguish between a change from red to green and a change from green to red) simply means that the red sections are past when the green ones are present and that the red ones are present when the green ones are future. Thus the notion of the history of the lamp as divisible into a series of sections, following each other in a certain direction, depends on the fact that each of these sections itself changes from future, through present, to past. It would therefore be circular to attempt to analyse changes in events in the way in which we have analysed changes in things, since the latter imply the former.

Apart from this objection, we can see directly that the change of events can-

not be treated like the changes of things. Let us take a short section of the history of the lamp, small enough to fall into a Specious Present, and such that the light from the lamp is red throughout the whole of this section. This short event was future, became present, and then became past. If we try to analyse this change, in the way in which we analysed the change of the lamp from red to green we shall have to proceed as follows: We shall have to divide this red event into shorter successive sections, and say that the latest of these have futurity, the middle ones presentness, and the earliest ones pastness. Now this analysis obviously does not fit the facts. For the fact is that *the whole* event was future, became present, and is now past. Clearly no analysis which splits up the event into successive sections with different characteristics is going to account for the change in the temporal attributes of the event as a whole.

We see then that the attempt to reconcile the incompatible temporal qualities of the same event by appealing to change, in the ordinary sense of the word, is both circular and ineffective. The circularity becomes specially glaring when put in the following way: The changes of things are changes *in Time*; but the change of events or of moments from future, through present, to past, is a change *of Time*. We can hardly expect to reduce changes of Time to changes in Time, since Time would then need another Time to change in, and so on to infinity.

We seem, therefore, to be forced back to the other type of solution, viz., that the predicates, *past*, *present*, and *future*, are of their very nature relational, like *large* and *small*. Unfortunately we have already had occasion to look at some solutions of this type – the policeman's bull's-eye and the different cognitive relations – and the omens are not very favourable.

If we reflect, we shall notice that there are two quite different senses in which an entity can be said to change its relational properties. An example of the first is where Tom Smith, the son of John Smith, becomes taller than his father. An example of the second is where Tom Smith ceases to be the youngest son of John Smith, and becomes the last son but one. What is the difference between these two cases? In the first we have two partially overlapping life-histories, T and J. If we cut up both into successive short sections we find that the earlier sections of T have the relation of 'shorter than' to the contemporary sections of J, whilst the later sections of T have the relation of 'taller than' to the contemporary sections of J. In the second we have quite a different state of affairs. When we say that T is the youngest son of J we mean that there is no entity in the universe of which it is true to say both that it is a son of J and that it is younger than T. When we say that T has ceased to be the youngest son of J we mean that the universe does contain an entity of which it is true to say both that it is a son of J and that it is younger than T. In the first case then, we simply have a difference of relation between different corresponding sections of two existing long events. In the latter, the difference is that a certain entity has changed its relational properties because a second entity, which did not formerly exist (and therefore could stand in *no* relation whatever to T), has begun to exist, and consequently to stand in certain relations to T, who is a member of the same universe as it.

Now it is obvious that the change that happens to an event when it ceases to

be present and becomes past is like the change of Tom Smith when he ceases to be the youngest son of John Smith; and the continuous retreat of an event into the more and more remote past is like the successive departure of Tom from being the 'baby' of the family, as John Smith (moved by the earnest exhortations of the Bishop of London) produces more and more children. A Specious Present of mine is just the last thin slice that has joined up to my life-history. When it ceases to be present and becomes past this does not mean that it has changed its relations to anything to which it was related when it was present. It will simply mean that other slices have been tacked on to my life-history, and, with their existence, relations have begun to hold, which could not hold before these slices existed to be terms to these relations. To put the matter in another way: When an event, which was present, becomes past, it does not change or lose any of the relations which it had before; it simply acquires in addition new relations which it *could* not have before, because the terms to which it now has these relations were then simply non-entities.

It will be observed that such a theory as this accepts the reality of the present and the past, but holds that the future is simply nothing at all. Nothing has happened to the present by becoming past except that fresh slices of existence have been added to the total history of the world. The past is thus as real as the present. On the other hand, the essence of a present event is, not that it precedes future events, but that there is quite literally *nothing* to which it has the relation of precedence. The sum total of existence is always increasing, and it is this which gives the time-series a sense as well as an order. A moment t is later than a moment t' if the sum total of existence at t includes the sum total of existence at t' together with something more.

We are too liable to treat change from future to present as if it were analogous to change from present to past or from the less to the more remote past. This is, I believe, a profound mistake. I think that we must recognise that the word 'change' is used in three distinct senses, of which the third is the most fundamental. These are (i) Change in the attributes of things, as where the signal lamp changes from red to green; (ii) Change in events with respect to pastness, as where a certain event ceases to be present and moves into the more and more remote past; and (iii) Change from future to present. I have already given an analysis of the first two kinds of change. It is clear that they both depend on the third kind. We analysed the change in colour of the signal lamp to mean that a red section of its history was followed by a green section of its history. This is sufficient analysis for a past change of quality, dealt with reflectively in retrospect. But, when we say that the red section precedes the green section, we mean that there was a moment when the sum total of existence included the red event and did not include the green one, and that there was another moment at which the sum total of existence included all that was included at the first moment and also the green event. Thus a complete analysis of the qualitative changes of things is found to involve the coming into existence of events.

Similarly we have seen that the second kind of change involves the third. For the change of an event from present to past turned out to depend on the fact

the sum total of existence increases beyond the limits which it had when our given event came into existence.

Let us call the third kind of change *Becoming*. It is now quite evident that becoming cannot be analysed into either of the two other kinds of change, since they both involve it. Moreover, we can see by direct inspection that becoming is of so peculiar a character that it is misleading to call it change. When we say that a thing changes in quality, or that an event changes in pastness, we are talking of entities that exist both before and after the moment at which the change takes place. But, when an event becomes, it *comes into existence*; and it was not anything at all until it had become. You cannot say that a future event is one that succeeds the present; for a present event is defined as one that is succeeded by nothing. We can put the matter, at choice, in one of two ways. We can either say that, since future events are non-entities, they cannot stand in any relations to anything, and therefore cannot stand in the relation of succession to present events. Or, conversely, we can say that, if future events succeeded present events, they would have the contradictory property of succeeding something that has no successor, and therefore they cannot be real.

It has long been recognised that there are two unique and irreducible, though intimately connected types of judgment. The first asserts that S is or exists; and is called an *existential* judgment. The second asserts that S is so and so, or has such and such a characteristic. This may be called a *characterising* judgment. The connexion between the two is that a thing cannot be so and so without *being*, and that it cannot be without being *so and so*.¹ Meinong, with the resources of the German tongue at his disposal, coins the convenient words *Sein* and *Sosein*. Now it seems to me that we have got to recognise a third equally fundamental and irreducible type of judgment, viz., one of the form: S becomes or comes into existence. Let us call these *genetic* judgments. I think that much of the trouble about Time and Change comes from our obstinate attempts to reduce such judgments to the characterising form. Any judgment can be *verbally* reduced to this form. We can reduce 'S is' to 'S is existent.' But the reduction is purely verbal, and those who take it seriously land in the sloughs of the Ontological Argument. Similarly 'S is future' is verbally a judgment that ascribes a characteristic to an event S. But, if we are right, this must be a mistake; since to have a characteristic implies to exist (at any rate in the case of particulars, like events), and the future does not exist so long as it is future.

Before passing on there is one more verbal ambiguity to be noted. The same word *is* is used absolutely in the existential judgment 'S is', and as a connective tie in the characterising judgment 'S is P.' Much the same is true of the word *becomes*. We say 'S becomes', and we say 'S becomes P.' The latter type of judgment expresses qualitative change, the former expresses coming into existence.

The relation between existence and becoming (and consequently between characterisation and becoming) is very intimate. Whatever is has become, and the sum total of the existent is continually augmented by becoming. There is no such thing as *ceasing* to exist; what has become exists henceforth for ever. When we say that something has ceased to *exist* we only mean that it has ceased to be

present; and this only means that the sum total of existence has increased since any part of the history of the thing became, and that the later additions contain no events sufficiently alike to and sufficiently continuous with the history of the thing in question to count as a continuation of it. For complete accuracy a slight modification ought to be made in the statement that ‘whatever is has become’. Long events do not become bodily, only events short enough to fall in Specious Presents become, as wholes. Thus the becoming of a long event is just the successive becoming of its shorter sections. We shall have to go more fully into the question of Specious Presents at a later stage.

We are left with two problems which we may hope that the previous discussions will help us to solve. (i) If the future, so long as it is future, be literally nothing at all, what are we to say of judgments which profess to be about the future? And (ii) What, in the end, is our answer to the original difficulty that every event is past, present, and future, and that these characteristics are mutually incompatible?

(i) Undoubtedly we do constantly make judgments which profess to be about the future. Weather forecasts, nautical almanacs, and railway time-tables, are full of such judgments. Admittedly no judgment about the future is absolutely certain (with the possible exception of the judgment that there will always be events of some kind or other); but this is irrelevant for our present purpose. No historical judgment about the past is absolutely certain either; and, in any case, our question is not whether we can have *certain* knowledge about the future, but is the prior question: What are we really *talking about* when we profess to make judgments about the future, and what do we *mean* by the truth or falsity of such judgments?

We cannot attempt to answer these questions till we have cleared up certain points about the nature of judgments in general. First, we must notice that the question: ‘What is a certain judgment about?’ is ambiguous. It may mean: ‘What is the subject or subjects of the judgment?’ or: ‘To what fact does the judgment refer?’ The fact to which a judgment refers is the fact that renders it true or false. It is true, if it has the peculiar relation of concordance to the fact to which it refers; and false, if it has the relation of discordance to this fact. Discordance, I think, is a positive relation which is incompatible with concordance; it is not the mere absence of concordance. I see no reason to suppose that the reference of a judgment to a fact is a third independent relation over and above the relations of concordance and discordance. I take it to be just the disjunction ‘concordance-or-discordance’; and I suppose that to say that J refers to F simply means that F is the fact which either makes J true by concording with it or false by discording with it.

Now people make many judgments, which have nothing to do with the future, but are nevertheless apparently about objects which do not, in fact, exist. Many English peasants, in the Middle Ages, must have made the judgments ‘Puck exists’ or ‘Puck has turned the milk.’ And the latter of these, of course, implies the former. I will assume (in spite of Sir Conan Doyle) that Puck does not in fact exist. What were these men referring to, in our sense of the word? To answer this we have simply to ask: What fact made their judgments false? The

answer is that it is the negative fact that no part of the universe was characterised by the set of characteristics by which they described Puck to themselves. Their judgment boils down to the assertion that some part of the existent is characterised by this set of characteristics, and it is false because it discords with the negative fact that the set in question characterises no part of the universe. Naturally they did not know that this was what their judgment referred to, or they would not have made it. But, in our sense of reference, there is no reason why a person who makes a judgment should know what it refers to.

Now it would obviously be absurd to say that what these men were *talking about* was the negative fact that no part of the universe has the characteristics which they ascribe to Puck. Hence we see the need of distinguishing between what a judgment refers to and what the person who makes the judgment is talking about. What they were talking about was a certain set of characteristics, viz., those by which they described Puck to themselves. This may be called the logical subject of their judgment. It is something real and independent of the judging mind; having the kind of reality and independence which is characteristic of universals, and not, of course, that which is characteristic of particular existents. Thus, although there is no such being as Puck, people who profess to be judging about him are not judging about nothing (for they are judging about a set of characteristics which is itself real, though it does not happen to characterise any particular existent). Nor are they referring to nothing (for they are referring – though they do not know it – to an important negative fact about the existent).

Since the non-existence of Puck is compatible with the fact that the judgment ‘Puck exists’ is an intelligible statement about something real, we may hope that the non-existence of the future may prove to be compatible with the existence and intelligibility of judgments which profess to be about the future. Up to a point the two kinds of judgment can be treated in much the same way. The judgment which is *grammatically* about ‘Puck’ proves to be *logically* about the set of characteristics by which the assertor describes Puck to himself. Similarly the judgment ‘To-morrow will be wet’, which is grammatically about ‘to-morrow’, is logically about the characteristic of wetness. The non-existence of to-morrow is therefore consistent with the fact that the judgment is about something.

Still there is one very important difference between the two kinds of judgment. Judgments like ‘Puck exists’ are not only *about* something; they also *refer to* some fact which makes them true or false. This fact may be negative, but it is a real fact about the existent world. If we ask what fact judgments ostensibly about the future refer to, we must answer that there is no such fact. If I judge, to-day that to-morrow will be wet, the only fact which this judgment can refer to, in our sense of the word, is the fact which renders it true or false. Now it is obvious that this fact is the wetness or fineness of to-morrow when to-morrow comes. To-day, when I make the judgment, there is no such fact as the wetness of to-morrow and there is no such fact as the fineness of to-morrow. For these facts can neither of them begin to be till to-morrow begins to be, which does not happen till to-morrow becomes to-day. Thus judgments which profess to

be about the future do not refer to any fact, whether positive or negative, at the time when they are made. They are therefore at that time neither true nor false. They will become true or false when there is a fact for them to refer to; and after this they will remain true or false, as the case may be, for ever and ever. If you choose to define the word *judgment* in such a way that nothing is to be called a judgment unless it be either true or false, you must not, of course, count 'judgments' that profess to be about the future as judgments. If you accept the latter, you must say that the Law of Excluded Middle does not apply to all judgments. If you reject them, you may say that the Law of Excluded Middle applies to all genuine judgments; but you must add that 'judgments' which profess to be about the future are not genuine judgments when they are made, but merely enjoy a courtesy title by anticipation, like the eldest sons of the higher nobility during the lifetime of their fathers. For convenience, I shall continue to speak of them as judgments.

So far then, we have determined two facts about judgments which profess to be concerned with the future. (a) They are about something, viz., some characteristic or set of characteristics; and (b) they do not refer to any fact at the time when they are made. This is clearly not a complete analysis. Two further points need to be cleared up. (a) If such judgments when made do not refer to anything, how is it that, if certain events become, the judgment is verified, and, if other events become, it is refuted? (b) If such judgments are about characteristics, what precisely is it that they assert about these characteristics?

(a) Suppose I judge to-day that to-morrow will be wet. Nothing that may happen to-morrow will be relevant to this judgment except the state of the weather, and nothing will then make it true except the wetness of the weather. This is true enough, but it does not prove that the judgment refers to any fact, in our sense of reference. With *any* judgment we can tell what *kind* of fact will verify or refute it, as soon as we know what the judgment is about and what kind of assertion it makes. But no amount of inspection of a judgment itself will show us *the particular fact* which makes it true if it is true and false if it is false. There is therefore no inconsistency between the statement that we can know at once what *kind of fact* would verify a judgment about the future, and the statement that such judgments do not refer to any *fact* when made.

(b) As regards any judgment we have to consider not only what it is about, but also what it asserts about its subject or subjects. These two questions are not altogether free from ambiguity, and this ambiguity must be cleared up before we consider the special question as to what judgments that profess to be about the future assert. (1) There is the confusion between what a judgment is about and what it refers to. This we have already dealt with. (2) There is the distinction between what a judgment is ostensibly about and what it is really about. If you had asked a peasant, who said that Puck had turned the milk, what he was talking about, he would have said that he was talking about a certain individual fairy. This is what the judgment professes to be about. What it is really about is a certain set of characteristics. Roughly speaking, we may say that what a judgment professes to be about can be determined by a grammatical analysis of the sentence in which the judgment is expressed. Although there is

always a connexion between the grammatical structure of a sentence and the logical structure of a judgment, it is highly dangerous to suppose that what the sentence is grammatically about is the name of what the judgment is logically about. (3) When these two confusions have been set aside and we are quite definitely dealing with the *judgment*, and neither with the *fact* to which it refers nor the *sentence* which expresses it, there is still a difficulty as to how much is to be included under the head of what the judgment is about and how much is to be included under the head of what the judgment asserts. Take first a very simple characterising judgment, like '3 is a prime'. What is this about, and what does it assert? We should all agree that it is at any rate about the number 3. But is it about the characteristic of primeness too? If you say Yes, what is there left for it to assert? If you say No, how can you face the obviously equivalent judgment 'Primeness is a characteristic of 3'? Exactly the same kind of difficulty arises over a relational proposition, like '3 is greater than 2'. We should all at this time of day agree that it is at least about the numbers 2 and 3. But is it or is it not about the relation of greater? I think that we must say that the former judgment is about primeness as much as it is about the number 3, and that the latter is about the relation of greater as much as it is about the numbers 2 and 3. Really it is as misleading to say that the first asserts primeness as to say that it asserts 3. The minimum that it asserts is the primeness of 3. Similar remarks apply to the second. If we like to use the useful word *tie*, which Mr W.E. Johnson² has lately introduced into logic, we might say that the first judgment is about the number 3 and the characteristic of primeness, and asserts that they are connected by the characterising tie. The second is about the numbers 3 and 2 and the relation greater, and asserts that they are connected by the relational tie in the order 3 to 2. But we might equally well distinguish different kinds of assertion, and say that the first is about the number 3 and the characteristic of primeness, and makes a characterising assertion about them. In the case of the second we should talk of a relating assertion.

So far we have purposely chosen examples which are about timeless objects, like numbers. Let us now take the series of judgments: 'It has rained', 'It is raining', and 'It will rain', which are about events, and contain an essential reference to time. The first may be analysed as follows: 'There is an event which is characterised by raininess, and the sum total of existence when the judgment is made includes all and more than all which it includes when this event becomes'. The second may be analysed as follows: 'There is an event which is characterised by raininess, and the sum total of existence is the same when this event becomes and when the judgment is made.' Thus judgments about the past and the present can be analysed into judgments which involve the four familiar types of assertion – the existential, the characterising, the genetic, and the relational. But the judgment that it will rain cannot be analysed in a similar way. It cannot mean anything that begins with the statement: 'There is an event', for the only events that there are are the events that have become up to the time when the assertion is made; the sum total of existence does not contain future events. We can only restate the judgment in the form: 'The sum total of existence will increase beyond what it is when the judgment is made, and some part

of what will become will be characterised by raininess'. We cannot then analyse *will* away, as we can *has been* and *is now*. Every judgment that professes to be about the future would seem then to involve two peculiar and not further analysable kinds of assertion. One of these is about becoming; it asserts that further events will become. The other is about some characteristic; it asserts that this will characterise some of the events which will become. If then we ask: What are judgments which profess to be about future events really about? the answer would seem to be that they are about some characteristic and about becoming. And if it be asked: What do such judgments assert? the only answer that I can give is that they assert that the sum total of existence will increase through becoming, and that the characteristic in question will characterise some part of what will become. These answers are compatible with the non-existence of the future. The only 'constituents' of the judgment, when it is made, are the characteristic – which has the kind of reality which universals possess – and the concept of becoming. About these the judgment makes certain assertions of a quite peculiar and not further analysable kind. Something called *to-morrow* is not a constituent of judgments which are grammatically about 'to-morrow', any more than an individual called *Puck* is a constituent of judgments which profess to be about 'Puck'.

I have thus tried to show that there is an extreme difference between judgments which profess to be about future events and those which are about past or present events. The former, when made, do not refer to anything, and therefore are not literally true or false, though it is possible for anyone who understands their meaning to see what kind of fact *will* eventually make them true or false as the case may be. Again, *is now* and *has been* need not be taken as new and ultimate types of assertion, but *will be* apparently must be so taken. Nevertheless, although the future is nothing and although judgments which profess to be about future events refer to nothing, they are not about nothing. They are about some characteristic and about becoming; and, so far as I can see, they make an unique and not further analysable kind of assertion about these terms. . . .

Notes

- 1 *Über die Stellung der Gegenstandstheorie* (Leipzig: R. Voitlander, 1907), and elsewhere [e.g., Alexius Meinong, 'The Theory of Objects', in R. M. Chisholm (ed.), *Realism and the Background of Phenomenology* (Glencoe, Ill.: The Free Press, 1960)].
- 2 W. E. Johnson *Logic* (Cambridge: Cambridge University Press) vol. i.

9 The Space–Time World: an Excerpt from *Philosophy and Scientific Realism**

J. J. C. Smart

Anthropocentricity of some Temporal Concepts

There is one feature of common ways of thinking which projects another sort of anthropocentric idea on to the universe at large. One can easily get the idea that the notions of past, present, and future apply objectively to the universe. In contrast, I shall argue that the concepts of past, present, and future have significance relative only to human thought and utterance and do not apply to the universe as such. They contain a hidden anthropocentricity. So also do tenses. On the other hand, the concepts of ‘earlier’, ‘simultaneous’, and ‘later’ are impeccably non-anthropocentric. I shall argue for a view of the world as a four-dimensional continuum of space–time entities, such that out of relation to particular human beings or other language users there is no distinction of ‘past’, ‘present’, and ‘future’. Moreover, the notion of the flow of time is the result of similar confusions. Our notion of time as flowing, the transitory aspect of time as Broad has called it, is an illusion which prevents us seeing the world as it really is.

The Space–Time World

A man or stone or star is commonly regarded as a three-dimensional object which nevertheless *endures* through time. This enduring through time clearly brings a fourth dimension into the matter, but this fact is obscured by our ordinary language. In our ordinary way of talking we stress the three-dimensionality of bodies, and by our notion of the permanent in change we conceal the fact that bodies extend through time. For philosophical reasons, therefore, it is of interest to discuss a way of talking which does not make use of the notion of the permanent in change. This explicitly four-dimensional way of talking has had important applications in physics. It needs, however, a bit of philosophical tidying up.

In what follows I shall want to make use of tenseless verbs. I shall indicate tenselessness by putting these verbs in italics. Tenseless verbs are familiar in logic and mathematics. When we say that two plus two *equals* four we do not mean that two plus two equals four at the present moment. Nor do we mean that two plus two always equalled four in the past, equals four now, and will

* From J. J. C. Smart, *Philosophy and Scientific Realism* (London: Routledge, 1963). Reprinted by permission of the author.

always equal four in the future. This would imply that two plus two will equal four at midnight tonight, which has no clear sense. It could perhaps be taken to mean that if someone says ‘two plus two *equals* four’ at midnight tonight, then he will speak truly, but then ‘at midnight tonight’ does not occur in the proposition that is mentioned.

It is perfectly possible to think of things and processes as four-dimensional space-time entities. The instantaneous state of such a four-dimensional space-time solid will be a three-dimensional ‘time slice’ of the four-dimensional solid. Then instead of talking of things or processes changing or not changing we can now talk of one time slice of a four-dimensional entity *being* different or not different from some other time slice. (Note the tenseless participle of the verb ‘to be’ in the last sentence.)

When we think four-dimensionally, therefore, we replace the notions of change and staying the same by the notions of the similarity or dissimilarity of time slices of four-dimensional solids. It may be objected that there is one sort of change which cannot be thus accommodated. For of any event, or of any time slice, it may be said on a certain occasion that it is in the future, and that later on it becomes present, and that later still it becomes past. It seems essential to say such things as that, for example, event E was future, is present, and will become past. The notion of change seems to be reintroduced into our four-dimensional scheme of things.

The objector is going too fast. If we are going to eliminate the notion of change we had better, to preserve consistency, eliminate also words such as ‘past’, ‘present’, ‘future’, and ‘now’. Let us replace the words ‘is past’ by the words ‘*is* earlier than this utterance’. (Note the transition to the tenseless ‘*is*’.) Similarly, let us replace ‘is present’ and ‘now’ by ‘*is* simultaneous with this utterance’, and ‘is future’ by ‘*is* later than this utterance’. By ‘utterance’ here, I mean, in the case of spoken utterances the actual sounds that are uttered. In the case of written sentences (which extend through time) I mean the earliest time slices of such sentences (ink marks on paper). Notice that I am here talking of self-referential *utterances*, not self-referential *sentences*. (The same sentence can be uttered on many occasions.) We can, following Reichenbach, call the utterance itself a ‘token’, and this sort of reflexivity ‘token-reflexivity’. Tenses can also be eliminated, since such a sentence as ‘he will run’ can be replaced by ‘he *runs* at some future time’ (with tenseless ‘*runs*’) and hence by ‘he *runs* later than this utterance’. Similarly, ‘he runs’ means ‘he *runs* (tenseless) simultaneous with this utterance’, and ‘he ran’ means ‘he *runs* (tenseless) earlier than this utterance’.² All the jobs which can be done by tenses can be done by means of the tenseless way of talking and the self-referential utterance ‘this utterance’. Of course, every time you use the words ‘this utterance’ you refer to a different utterance. So though I have just said that ‘all the jobs’ we can do with tenses and with words such as ‘past’, ‘present’, ‘future’, and ‘now’ can be done in our tenseless language together with the self-referential utterance ‘this utterance’, there is nevertheless one sort of thing that we cannot say in our tenseless language. We cannot translate a sentence of the form ‘This event was future, is present and will be past.’

So far from this last fact being a criticism of the tenseless way of talking, it is, I think, pure gain. The inability to translate talk of events changing in respect of pastness, presentness, and futurity into our tenseless language can be taken simply as a proof of the concealed token reflexivity of tenses and of words such as ‘past’, ‘present’, ‘future’, and ‘now’. If ‘past’ means ‘earlier than this utterance’ it is going to have a different reference every time it is used. If uttered in 1950 it refers to events earlier than 1950 and if uttered in 1965 it refers to events earlier than 1965. The notion of events ‘changing from future to past’ is simply a confused acknowledgment of this quite simple sort of fact. Once we see this we banish from the universe much unnecessary mystery.

If past, present, and future were real properties of events, then it would require explanation that an event which becomes present in 1965 becomes present at that date and not at some other (and this would have to be an explanation over and above the explanation of why an event of this sort *occurred* in 1965). Indeed, every event is ‘now’ at some time or another, and so the notion of ‘now’ cannot be that of an objective property in nature which singles out some events from others. When we talk in our four-dimensional language of space-time we must clearly talk neither of events nor of things changing, since we have replaced the notion of a thing as the permanent in change by that of a four-dimensional entity, some of whose time slices *are* or *are not* different from others. But even in our language of the permanent in change we must still not think of *events* changing. Things (and processes) come into existence, change, or stay the same, whereas to say that an event (such as the beginning of a football match) ‘came into existence’ or ‘changed’ would be absurd. The only exception to this rule is that we *can* say that events ‘become present’, or ‘become past’, or even ‘become probable’ or ‘become unlikely’. (On the other hand, it is somewhat strained to say that a *thing* becomes past or probable.) These phenomena of language can be neatly explained once we recognise the fact that utterances of words such as ‘past’, ‘present’, and ‘future’ refer to themselves. So also with ‘probable’ and ‘unlikely’, since here ‘probable’ and ‘unlikely’ mean ‘probable, or unlikely, in terms of *present* evidence’.

Some philosophers have talked as though events ‘become’ or ‘come into existence’. ‘Become’ is a transitive verb, and so to say that an event ‘becomes’ must presumably mean that it ‘becomes present’, and this, we have seen, misleads by concealing the token-reflexivity of ‘present’ and suggesting that the becoming present of an event is a real change like, for example, the becoming brown of a grassy hillside in summer. Similarly, an event cannot come into existence – a new building can come into existence, but the building of it cannot meaningfully be said to come into existence. (In the four-dimensional way of talking, of course, we must not say even that *things* come into existence – we replace talk of a building coming into existence at *t* by talk of the earliest time slice of the building *being* at *t*.) Some philosophers have erected these misconceptions about the grammar of the verbs ‘to become’ and ‘to come into existence’ into a metaphysics, as when, for example, Whitehead said that ‘actual occasions become’.

We can also see how misleading it is to talk of the flow of time, or of our

advance through time. To say that by next year a year of time will have gone by is simply to say that our conscious experiences of a year later than this utterance *are* (tenseless) a year later than this utterance. Our consciousness does not literally advance into the future, because if it did we could intelligibly ask ‘How fast does it advance?’ We should need to postulate a hyper-time with reference to which our advance in time could be measured (seconds per hyper-seconds), but there seems to be no reason to postulate such an entity as a hyper-time. (There is still something odd about movement in time even if it is said, as it might be, that the hyper-time has an *order* but no metric. This would rule out talk of ‘seconds per hyper-seconds’, but it would not affect the fact that change in time would still be a change with respect to hyper-time. Moreover, anyone who thought that time-flow was necessary for time would presumably want to say that hyper-time-flow was necessary for hyper-time. He would therefore be driven to postulate a hyper-hyper-time, and so on without end.)

It is true that sometimes in relativity theory it is said that time ‘runs more slowly’ in a moving system than it does in a system at rest relative to us. This, however, is not to imply any movement or ‘running’ of time. What is meant, by this misleading locution, is that according to the conventions of simultaneity of our system of axes the space-time interval between events on our clock is greater than that between simultaneous events on a clock in the moving system. Equally, since we are moving relative to the other system, clocks in our system, ‘run slow’ relative to the moving system. Indeed, so far from relativity leading to difficulties for us, the reverse is the case. The four-dimensional way of talking which we have advocated could still have been possible in pre-relativity days, but it has derived additional theoretical advantages from Minkowski’s discovery that the Lorentz transformations of special relativity can be regarded simply as a rotation of axes in space-time. This is not the place to go into an exposition of relativity, but I wish to record the conviction that many of the puzzles and paradoxes of relativity (or rather those things which are sometimes wrongly thought to be puzzles and paradoxes) can most easily be resolved by drawing diagrams of Minkowski space-time, in which most of these at first sight counter-intuitive facts will at once look quite obvious. (We must, of course, bear in mind that the geometry of space-time is not Euclidean.)

If I am right in supposing that ‘now’ is equivalent to ‘simultaneous with this utterance’, then I am able, as we have seen, to reject the notion of an objective ‘now’, the notion that even in past ages when there were perhaps no sentient beings there was nevertheless a moment which was distinguishable as ‘the present’ or ‘now’.³ An utterance of the word ‘now’ refers to itself, since it refers to the set of events simultaneous with itself. Now the special theory of relativity shows that there is no unique set of events which is ‘now’ or ‘simultaneous with this utterance’. Which time slice of the four-dimensional manifold constitutes a ‘now’ depends on the frame of reference in which we are at rest. Our four-dimensional cake can be sliced at different angles. It is worth mentioning this consideration, since I have known one very eminent disciple of Whitehead (and therefore of an objective ‘becoming’) to have been genuinely worried by it. For our purposes we can easily modify the notions of ‘now’ or ‘present’ to mean ‘simulta-

neous, relative to the utterer's frame of reference, with this utterance'. Similar modifications must be made for 'past' and 'future'.

The notions of 'past', 'present', and 'future' are more complex than those of 'earlier' and 'later', since the former notions do, and the latter notions do not, involve reference to the utterer's position in space-time. 'Earlier' and 'later' fit into the tenseless locution that I have advocated, whereas 'past', 'present', and 'future' do not.

It may now be objected: 'So much the worse for the tenseless way of talking.' For it may be said that so far from the tensed language being definable in terms of the tenseless one (together with the notion of self-referential utterances), the tenseless '*is*' has to be defined in terms of the tensed one. As Wilfrid Sellars has objected,⁴ a tenseless sentence ' x is ϕ at t ' is equivalent to the tensed one 'Either x was ϕ at t or is ϕ at t or will be ϕ at t '. So ' x is ϕ at t ' is not like '7 is a prime number', which does *not* mean '7 was, is, or will be a prime number'.

Now there is, I agree, a difference between ' x is ϕ at t ' and '7 is a prime number'. But it does not appear to be happily expressed by saying that the former sentence is not really tenseless. It is better expressed by saying that '*is* a prime number at such and such a time' is not a meaningful predicate. The difference can be brought out within the *predicates* of ' x is ϕ at t ' and '7 is a prime number' and has nothing to do with the copula. It is true that in extending the tenseless way of talking from pure mathematics to discourse about the space-time world it is natural to introduce ' x is ϕ at t ' via the locution ' x was, is, or will be ϕ at t '. This is because it is tacitly agreed that x is a space-time entity and so earlier, simultaneous with or later than our present utterance, though in the present context which it is does not matter. But though it is natural to wean users of tensed language from their tenses in this way, it is by no means logically necessary that a tenseless language should be introduced in this manner.

A fable may be of use here. Consider a tribe whose religious and social life depended on the exact numerical age in years of the king, and that for this reason their very language made a difference between three sorts of numbers: those numbers which were less than the number of years which was the king's age, the number which was equal to this number, and the numbers which were greater than this number. Indeed, our tribe do not think of the three sorts of numbers as numbers, but believe that there are three sort of entities, alphas, betas, and gammas. They are, of course, slightly puzzled that every year (until the king dies) a gamma becomes a beta and a beta becomes an alpha. Someone might get the bright idea of introducing the notion of number as 'number = alpha or beta or gamma'. Would this show that the notion of 'number' had anything to do with the age of the king? It has indeed been introduced by reference to notions that have to do with the age of the king, but in such a way that this kingly reference 'cancels out'. Sellars argues that Tom, in 1955, Dick, in 1956, and Harry, in 1957, could agree that Eisenhower should be (tenselessly) President in 1956, but that their reasons would be different. Tom's reason would be 'Eisenhower will be President in 1956', Dick's reason would be 'Eisenhower is President in 1956', and Harry's reason would be 'Eisenhower was President in 1956'. These considerations, says Sellars, make it quite clear that

the tenseless present, introduced via ‘was, is, or will be’, is quite other than the tenseless present of mathematics. As against this, I would say this: the fact that, since they speak from different temporal perspectives, Tom, Dick, and Harry give different reasons for saying ‘Eisenhower *is* (tenseless) President in 1956’ does not show that they mean anything non-tenseless. For a reason ‘*q*’ offered for ‘*p*’ in the explanation ‘*p* because *q*’, may well contain extraneous and irrelevant elements. It does not therefore seem to me that Sellars has given any convincing reason for saying that there is any important difference between the tenseless ‘*is*’ of ‘Eisenhower *is* President in 1956’ and ‘7 + 5 *is* equal to 12’. Of course Eisenhower is a temporal entity, and so ‘in 1956’ has sense in relation to him, and numbers are non-temporal entities, and so there is no question of ‘in 1956’ in the case of the second proposition. This distinction can perfectly well be made explicit in the *predicates* of the two sentences and need not be done in the *copulae*. This also explains why it is natural (though there is no need to suppose that it is logically *necessary*) to introduce the tenseless *is* in the case of ‘Eisenhower *is* President in 1956’ *via* the idiom ‘was, is, or will be’, whereas it would, as Sellars notes, not be natural to do so in cases like ‘7 + 5 *is* equal to 12’.

A sentence of the form ‘*x* is ϕ at *t*’ is, of course, not timeless, any more than ‘*x* is ϕ at such and such a place at *t*’ is *spaceless*. Timelessness is not the same as tenselessness, ‘7 *is* a prime number’ is both tenseless and timeless. (There is no sense in saying ‘7 *is* a prime number at *t*’.) The tenseless way of talking does not therefore imply that physical things or events are eternal in the way in which the number 7 is.

As we have already noted, it is sometimes said that ‘this utterance’ is to be analysed as ‘the utterance which is *now*’. If so, of course, tenses or the notions of past, present, and future *are* fundamental. My reply to this move is to say that this is simply a dogmatic rejection of the analysis in terms of token-reflexiveness. On this analysis ‘now’ is elucidated in terms of ‘this utterance’, and not vice-versa. This seems to me to be a perfectly legitimate procedure. How does one settle the argument with someone who says that ‘this utterance’ has to be analysed in terms of ‘utterance now’? Any analysis is a way of looking at language, and there is no one way. I advocate my way, because it fits our ordinary way of talking much more closely to our scientific way of looking at the world and it avoids unnecessary mystification. If someone is adamant that his analysis is the correct analysis of ordinary language I am prepared to concede him this rather empty point. Ordinary language is, then, on his account, more at variance with science than is my version of ordinary language. Nevertheless, the two analyses are in practice pretty well equivalent: in ordinary life a linguist will detect no difference between ‘ordinary language’, as in accordance with my analysis, and ‘ordinary language’, as in accordance with my opponent’s analysis. Our ordinary language is just not quite so ‘ordinary’ as is our opponent’s, but it is just as good even for ordinary purposes. It is perhaps more ‘ordinary’ to say that sugar ‘melts’ than that it ‘dissolves’, but the greater scientific correctness of the latter locution does not in any way unfit it for even the most practical purposes. Similarly, the additional theoretical advantages of looking at temporal language in the present way suggest that we should prefer this analysis to the other. Perhaps the objector is saying that

the present analysis is impossible for any language, whether ‘ordinary’ or scientific. But it is not at all evident why the objector should think that an utterance like ‘this utterance’ cannot be *directly* self-referential. We hear a token of the form ‘this utterance’ and simply understand that this token utterance is the one referred to. We can at a later date *say* what the utterance referred to was: we can enumerate sufficient of its characteristics to identify it. It is always logically possible, of course, that some *other* utterance should possess this list of characteristics – we can misidentify an utterance just as we can misidentify a stone, a tree, or a person. But in fact we need not and do not. Moreover, if we *did* misidentify it, how would the proposal to elucidate ‘this’ in terms of ‘now’ have prevented us?

The self-reference of specific utterances of words such as ‘here’ and ‘now’ is sufficient to deal with the following puzzle: it is logically possible that in remote regions of space–time the universe might repeat itself exactly.⁵ We cannot therefore uniquely single out an entity (say this table) by referring to it by means of some set of properties – elsewhere in the universe there might be another table with exactly the same qualities and relations to other objects. A token-reflexive expression can, however, uniquely pick out this table – ‘this table is near the utterance of *this token*’. Of course there may well be other Smarts in other regions of space–time uttering precisely similar tokens, but they can all refer uniquely to their environments by token-reflexive means. There is, however, no need for words such as ‘now’ or tenses – ‘this utterance’ or ‘this token’ is always enough to do the trick. Sellars makes a similar point when he argues that token-reflexives are needed to distinguish the real world from fictional worlds. (The real world is a system of entities which includes *this*.) There are obvious difficulties here, which perhaps can be got round only if one accepts Sellars’ own interesting but debatable views on the concept of existence. I should wish to say too, however, that tenses and words such as ‘present’ or ‘now’ are unimportant here, and that a simple token-reflexive device (corresponding to ‘this utterance’) is enough to do the trick. For cosmological *theory*, moreover, token-reflexivity is *not* needed. Here one can simply assert, as part of the theory, either that the universe repeats itself in remote parts of space–time or that it does not. It is only in *applying* the theory to observations that unique references have actually to be made.

It should be hardly necessary, at this stage, I should hope, to emphasise that when in the tenseless way of talking we banish tenses, we really must banish them. Thus, when we say that future events exist we do *not* mean that they exist now (present tense). The view of the world as a four-dimensional manifold does not therefore imply that, as some people seem to have thought, the future is already ‘laid up’. To say that the future is already laid up is to say that future events exist *now*, whereas when I say of future events that they *exist* (tenselessly) I am doing so simply because, in this case, they *will* exist. The tensed and tenseless locutions are like oil and water – they do not mix, and if you try to mix them you get into needless trouble. We can now see also that the view of the world as a space–time manifold no more implies determinism than it does the fatalistic view that the future ‘is already laid up’. It is compatible both with determinism and with indeterminism, i.e. both with the view that earlier time slices of the

universe are determinately related by laws of nature to later time slices and with the view that they are not so related.

When we use tenses and token-reflexive words such as 'past', 'present', and 'future', we are using a language which causes us to see the universe very much from the perspective of our position in space-time. Our view of the world thus acquires a certain anthropocentricity, which can best be eliminated by passing to a tenseless language. By the use of such expressions as 'earlier than this utterance' and 'later than this utterance' we make quite explicit the reference to our particular position in space-time. Once we recognise this anthropocentric reference and bring it out into the open we are less likely to project it on to the universe. The tenseless and minimally token-reflexive language enables us to see the world, in Spinoza's phrase, *sub specie aeternitatis*.

Notes

- 1 This vivid expression is used by J. H. Woodger. See his 'Technique of Theory Construction', *International Encyclopedia of Unified Science*, vol. 2, no. 5 (University of Chicago Press, 1939).
- 2 H. Reichenbach has given an excellent discussion of tenses and similar notions in terms of 'token-reflexivity' in §§ 50–1 of his *Elements of Symbolic Logic* (New York, Macmillan, 1947).
- 3 See the passage from H. Bergmann, *Der Kampf um das Kausalgesetz in der jüngsten Physik* (Braunschweig, 1929), pp. 27–8, which is quoted in A. Grünbaum's paper 'Carnap's Views on the Foundations of Geometry', in P.A. Schilpp (ed.), *The Philosophy of Rudolf Carnap* (La Salle, Ill.: Open Court, 1963). Grünbaum's paper contains an excellent critique of the idea of an objective 'now'.
- 4 In his essay 'Time and the World Order', in H. Feigl and G. Maxwell (eds), *Minnesota Studies in the Philosophy of Science*, vol. III (University of Minnesota Press, 1962), pp. 527–616, see p. 533.
- 5 See A. W. Burks, 'A Theory of Proper Names', *Philosophical Studies*, vol. 2, (1951), pp. 36–45, and N. L. Wilson, 'The Identity of Indiscernibles and the Symmetrical Universe', *Mind*, vol. 62 (1953), pp. 506–11.

10 Topis, Soris, Noris: an Excerpt from *The Existence of Space and Time**

Ian Hinckfuss

... [I]t is often claimed that our sense of time-flow – of the feeling we have that events come out of the future, become present, and then recede into the

* From Ian Hinckfuss, *The Existence of Space and Time* (Oxford: Oxford University Press, 1975). Reprinted by permission of the author and Oxford University Press.

past – is an illusion. Such a belief is common among the mystics of the Orient; but it is shared also by many physicists and philosophers of science. Thus on p. 132 of his book *Philosophy and Scientific Realism*, J. L. C. Smart maintains: ‘Our notion of time as flowing . . . is an illusion which prevents us seeing the world as it really is’.

Smart claims that this illusion is largely maintained by the fact that we speak a tensed language. We use the so-called past tense in describing events which occur earlier than the time of making the description, the present tense in describing events which occur at a time simultaneous with our description of them, and the future tense in describing events which occur later than the description. This makes it seem that we are always describing events as falling into one of three categories: being in the past, being in the present; or being in the future. Thus we are able to say of some particular event that is now occurring, that it *was* in the future, it *is now* present, and it *will be* past.

In the remainder of this section, I shall argue that our ability to describe events as being in the past or present or future is not sufficient to give us the idea of a flow of time. The reason for this is that the situation as so far described could be deemed to be perfectly symmetrical between the past and the future. We use one sort of tense for the past, another for the present and yet another for the future. Even if we add the fact that what we mean by the past is anything *earlier than* the present, and what we mean by the future is anything *later than* the present, the situation could still be deemed to be symmetrical and static.

Consider a corresponding spatial analogy. Let us imagine that we use the words ‘the planar’ to refer to a plane parallel to the equatorial plane, and which passes through the speaker at the time of his uttering the words. Assume further, that we used the words ‘the north’ to refer to those events north of such a plane and ‘the south’ to refer to any events south of the plane. Further, let us assume that we modified our verbs depending on whether we were speaking of events which were to the north of us, or which were at the planar, or which were to the south of us, in the following way. If the event is to the north, we use the prefix ‘nor’ before the verb. If the event is at the planar, we use the prefix ‘top’. If the event is to the south, we use the prefix ‘sor’. Thus a speaker in Britain in August 1972 would say ‘There soris a war in Indo-China.’

The correct grammar for a speaker in Australia would be ‘There noris a war in Indo-China’, while a person in Indo-China, or for that matter, in Thailand or Burma, would say ‘There topis a war in Indo-China.’ We could also say of any event that topis occurring, that it *soris* to the north, it *topis* at the planar, and it *noris* to the south. That is, speaking normally, south of here, the event is to the north, here the event is on this plane, and north of here, the event is to the south. Further, corresponding to the facts that the past is *earlier than* the present and the future is *later than* the present, there would be the facts that the south is *south of* the planar and the north is *north of* the planar.

In his *Philosophy and Scientific Realism*, Smart has claimed that if we were to rid our language of tenses and of the phrases ‘in the future’, ‘at present’, and ‘in the past’, we could defuse our tendency to think of time as flowing. The idea is that we use only one tense, or rather that we use a tenseless form of the verb,

and that we use the descriptions 'earlier than this utterance', 'simultaneous with this utterance', and 'later than this utterance' to give the relative temporal position of the utterance with the event being described. In Smart's reduction 'this utterance' is meant to refer to the utterance at the time it is being produced. Thus 'There will be peace' is rendered 'There *is* peace later than this utterance', 'There is a war going on' is rendered 'There *is* a war going on simultaneous with this utterance', and 'There was a war' is rendered 'There *is* a war earlier than this utterance', where the italicized '*is*' is the tenseless 'is' of 'Two is an even number' or 'All ravens are black.' Confusion about time-flow, it is claimed, tends to arise due to the fact that our normal tensed manner of speaking fails to emphasize the *relational* nature of tenses, the relations being the temporal relations of 'earlier than', 'simultaneous with', and 'later than' between the events being described and the utterance which describes them. Likewise the phrases 'in the future', 'at present', and 'in the past' tend, we are told, to give us the impression that events can have a property of being in the future or being present or being in the past, whereas what is actually the case is that there are no such one-place properties of single events. What there are, are two-place relations between events, the relations of being earlier than, simultaneous with, or later than.

Now it might be true that the use of a tenseless language in which these relationships were made explicit would cure us of our time-flow illusion, if illusion it is. But how it would do this I do not know, for what is certain is that such an illusion could not be *explained* in terms of our tensed language and our ability to refer to the past, present, and future, particularly if our tensed language is regarded as being like its corresponding spatially tensed counterpart. For the corresponding spatially tensed language of the north, the planar, and the south illustrates the absolute symmetry of this mode of speech. Even if we were confused enough to think of 'being at planar' as an intrinsic property of events, rather than a relation between events and ourselves, there would be no reason to think of the planar moving northwards any more than there would be reason to think of it moving southwards. Yet we feel that the present encroaches on the future and that it does not encroach upon the past. We feel ourselves to be approaching death, not birth. Alternatively we sometimes think of the present as something static, past which events flow in a vast stream, as when we think '1972 will soon be past'. But there is no more reason to think of events to the north flowing past the planar into the south any more than there is to think of events to the south flowing past the planar to the north. . . .

11 Some Free Thinking about Time*

A. N. Prior

There's a dispute among philosophers – indeed there has always been this dispute among philosophers – as to whether time is real. Some say yes, and some say no, and some say it isn't a proper question; I happen to be one of the philosophers who say yes. All attempts to deny the reality of time founder, so far as I can see, on the problem of explaining the *appearance* of time's passage: for appearing is itself something that occurs in time. Eddington once said that events don't happen, we merely come across them; but what is *coming across* an event but a happening?

So far, then, as I have anything that you could call a philosophical creed, its first article is this: I believe in the reality of the distinction between past, present, and future. I believe that what we see as a progress of events *is* a progress of events, a *coming to pass* of one thing after another, and not just a timeless tapestry with everything stuck there for good and all.

To bring out the difference of viewpoint I have in mind, let me mention a small logical point. Logic deals, at bottom, with statements. It enquires into what statements follow from what – but logicians aren't entirely agreed as to what a statement *is*. Ancient and medieval logicians thought of a statement as something that can be true at one time and false at another. For example, the statement 'Socrates is sitting down' is true so long as he *is* sitting down, but becomes false when he gets up. Most modern logicians, however, say that if a statement is true at any time, it's true all the time – once true, always true. Confronted with the example 'Socrates is sitting down', they would say that this isn't really a statement, but only a piece of a statement. It needs to be completed by some unambiguous specification of the time at which he is sitting down, for example, at exactly 3 p.m. (Greenwich mean time) on June 15th, 326 BC. And when we say that he *is* sitting down at this time and date, we don't need to change this 'is' to 'was', because in this sort of statement 'is' hasn't any tense at all – the complete statement tells us a timeless property of a date or moment; that date or moment just *is*, eternally, a Socrates-sitting-downy date or moment.

Such a notion of what a statement is seems clearly to reflect what I have called the tapestry view of time, and I believe accordingly that this is a point at which logicians ought to retrace their steps. I think the logically primary sense of the word 'statement' is the old sense, the sense in which a statement which is true at one time may be false at another time, and in which the *tense* of statements must

* From B.J. Copeland, ed., *Logic and Reality: Essays on the Legacy of Arthur Prior* (Oxford: Oxford University Press, 1996). Reprinted by permission of Oxford University Press.

be taken seriously. I don't think these are just fragments of 'statements' in some more fundamental sense of the word; on the contrary, the allegedly tenseless statements of modern logic are just a special case of statements in the old sense – they are statements which happen to be either always false or always true, and the 'is' that occurs in them is not really a tenseless 'is' but is just short for 'is', always has been, and always will be'.

This belief, or prejudice, of mine is bound up with a belief in real freedom. One of the big differences between the past and the future is that once something has become past, it is, as it were, out of our reach – once a thing has happened, nothing we can do can make it not to have happened. But the future is to some extent, even though it is only to a very small extent, something we can make for ourselves. And this is a distinction which a tenseless logic is unable to express. In my own logic with tenses I would express it this way: We can lay it down as a law that whatever *now is* the case *will always have been* the case; but we can't interchange past and future here and lay it down that whatever *now is* the case *has always been going to be* the case – I don't think that's a logical law at all; for if something is the work of a free agent, then it wasn't going to be the case until that agent decided that it was. But if happenings are just properties timelessly attached to dates, I don't see how you can make this distinction.

This general position that I want to uphold has come under fire from different quarters at different times. In the Middle Ages it was menaced by the theologians, many of whom, like Thomas Aquinas, taught that God doesn't experience time as passing, but has it present all at once. In other words, God sees time as a tapestry. Other medieval theologians such as Duns Scotus argued, I think very sensibly, that since time *isn't* a tapestry, either God *doesn't* see it that way or He has an illusion about it, and since He hasn't any illusions He *doesn't* see it that way but sees it as it is, as passing. I would go further than Duns Scotus and say that there are things about the future that God *doesn't* yet know because they're not yet there to be known, and to talk about knowing them is like saying that we can know falsehoods. God cannot know that 2 and 2 are 5, because 2 and 2 *aren't* 5, and if He's left some matter to someone's free choice, He cannot know the answer to the question 'How will that person choose?' because there isn't any answer to it until he has chosen.

Nowadays it's not so much the theologians we have to contend with as the scientists, and the philosophical interpreters of the scientists. Many philosophical upholders of what I've called the tapestry view of time claim that they have on their side a very august scientific theory, the theory of relativity, and of course it wouldn't do for mere philosophers to question august scientific theories. Well, I've tried to find out recently exactly what is the strength of this argument, and I'll discuss it with you now as simply as I can, though I'll have to warn you that it's not *very* simple. The physical facts seem to be more or less like this: *My* experience has a quite definite time-order, of which I am immediately aware; and *your* experience has a definite time-order, of which *you* are immediately aware; and similarly for any observer, no matter where he is, or how he is moving. Moreover, if you were to calculate the time-order of my experiences, I would agree with your result, and similarly, if I were to calculate yours. The

trouble arises when we come to *compare* one another's experiences – when, for example, I want to know whether I saw a certain flash of light before you did, or you saw it before I did. Even about points like this there is often agreement all round, but we can't depend on it. It could happen that if I assumed myself to be stationary and you moving, I'd get one result – say that I saw the flash first – and if you assumed that you were stationary and I moving, you'd get a different result. I could explain your result by saying that the speed of your movement had made your measuring instruments go haywire; but you could explain my results in the same way. And it appears to be established that in such a case there would be no physical way of deciding which of us is right; that is, there is no way of determining whether the light-signal first crossed my path or yours. And the conclusion drawn in the theory of relativity is that this question – the question as to which of us is right, which of us really saw it first – is a meaningless question; outside our private paths, the time-direction and space-direction just aren't as distinct as that.

Now I don't want to be disrespectful to people whose researches lie in other fields than my own, but I feel compelled to say that this just won't do. I think we have excellent grounds for insisting that the question in question is *not* a meaningless one, and I'll try and explain what its meaning is. People who are doing relativity physics are concerned with the relations of before and after and simultaneity, but these aren't the first things as far as the real passage of time is concerned – the first thing is the sequence of past, present, and future, and this is not just a private or local matter, different for each one of us; on the contrary, pastness, presentness, and futurity are properties of events that are independent of the observer; and under favourable conditions they are *perceived* properties of events. We all know what it is to wait for something – an examination, for example; or coming home from the war; or Christmas. What we're waiting for begins by being future; it *hasn't yet* come to pass. Then a time comes when it does come to pass – when it's *present*, and we're aware of its presentness, and there's no mistaking it. And then it's past, and we say, perhaps, 'Thank goodness all that's over'; and we all know quite well what this 'being over' is, and couldn't mistake it for anything else. I have a very good friend and colleague in Australia, Professor Smart of Adelaide, with whom I often have arguments about this. He's an advocate of the tapestry view of time, and says that when we say 'X is now past' we just mean 'The latest part of X is earlier than this utterance.' But, when at the end of some ordeal I say 'Thank goodness that's over', do I mean 'Thank goodness the latest part of that is earlier than this utterance'? I certainly do not; I'm not thinking about the utterance at all, it's the *overness*, the *now-endedness*, the *pastness* of the thing that I'm thankful for, and nothing else. Past and future are in fact not to be defined in terms of earlier or later, but the other way round – 'X is earlier than Y means 'At some time X was past and Y was present', and 'X is later than Y means the opposite of this.

Coming back to this allegedly meaningless question as to whether you or I saw the light-flash first, surely what it means is just this: When I was seeing the flash, *had* you already seen it, or *had* you not? In other words, when my seeing it was a *present* fact, had your seeing it become a *past* fact, or *had* it not? And I

just cannot be persuaded that such a question is meaningless – its meaning seems to me perfectly obvious. When an event X is happening, another event Y either *has* happened or *has not* happened – ‘having happened’ is not the kind of property that can attach to an event from one point of view but not from another. On the contrary, it’s something like *existing*; in fact to ask what has happened *is* a way of asking what exists, and you can’t have a thing existing from one point of view but not existing from another, although of course its existence may be *known* to one person or in one region, without being known to or in another.

So it seems to me that there’s a strong case for just digging our heels in here and saying that, relativity or no relativity, if I say I saw a certain flash before you, and you say you saw it first, one of us is just wrong – or misled it may be, by the effect of speed on his instruments – even if there is just no physical means whatever of deciding which of us it is. To put the same point another way, we may say that the theory of relativity isn’t about *real* space and time, in which the earlier-later relation is defined in terms of pastness, presentness, and futurity; the ‘time’ which enters into the so-called space-time of relativity theory isn’t this, but is just part of an artificial framework which the scientists have constructed to link together observed facts in the simplest way possible, and from which those things which are systematically concealed from us are quite reasonably left out.

This sort of thing has happened before, you know. When that formidable mathematical engine the differential calculus was first invented, its practitioners used to talk a mixture of excellent mathematics and philosophical nonsense, and at the time the nonsense was exposed for what it was by the philosopher Berkeley, in a pamphlet entitled ‘A Defence of Free Thinking in Mathematics’. And the mathematicians saw in the end that Berkeley was right, though it took them about a century and a half to come round to it. They came round to it when they became occupied with problems which they could solve only by being accurate on the points where Berkeley had shown them to be loose; then they stopped thinking of the things he had to say as just a reactionary bishop’s niggling, and began to say them themselves. Well, it may be that some day the mathematical physicists will want a sound logic of time and tenses; and meanwhile the logician had best go ahead and construct it, and abide his time.

12 The Fourth Dimension: an Excerpt from *The Ambidextrous Universe**

Martin Gardner

Immanuel Kant, the great German philosopher of the eighteenth century, was the first eminent thinker to find a deep philosophical significance in mirror imagery. That an asymmetric object could exist in either of two mirror-image forms seemed to Kant both puzzling and mysterious. Before discussing some of the implications Kant drew from left-right asymmetry, let us first see if we can recapture something of the mood in which he approached this topic.

Imagine that you have before you, on a table, solid models of the enantiomeric polyhedrons shown in figure 2. The two models are *exactly alike* in all geometrical properties. Every edge of one figure has a corresponding edge of the same length on the other figure. Every angle of one figure is matched by a duplicate angle on the other. No amount of measurement or inspection of either figure will disclose a single geometrical feature not possessed by the other. They are, in a sense, identical, congruent figures. Yet clearly they are *not* identical!

This is how Kant expressed it, in Section 13 of his famous *Prolegomena to All Future Metaphysics*: “What can more resemble my hand or my ear, and be in all points more like, than its image in the looking-glass? And yet I cannot put such a hand as I see in the glass in the place of its original. . . .”

That two objects can be exactly alike in all properties, yet unmistakably different, is certainly one reason why the looking-glass world has such an eerie quality for children and for primitive people when they encounter it for the first time. Of course the major source of spookiness is simply the appearance behind

* From Martin Gardner, *The Ambidextrous Universe* (New York: Basic Books, 1964). Reprinted by permission of the author.

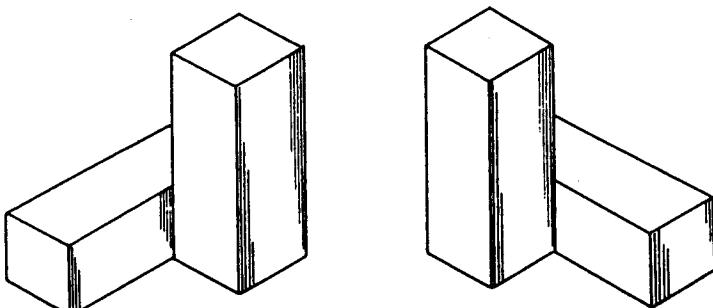


Figure 2 Enantiomeric polyhedrons

the glass of a world that looks as real as the world in front, yet is completely illusory. If you want to puzzle and fascinate a small child, stand him in front of a large wall mirror at night, in a dark room, and hand him a flashlight. When he shines the flashlight into the mirror the beam goes straight into the room behind the glass and illuminates any object toward which he aims it! This strong illusion of a duplicate room is spooky enough, but it grows even spookier when one becomes aware of the fact that everything in the duplicate room “goes the other way.” It is the *same* room, yet it isn’t.

Exactly what Kant made of all this is a tangled, technical, controversial story. . . .

Kant’s first published paper, *Thoughts on the True Estimation of Living Forces* (1747), contains a remarkable anticipation of n -dimensional geometry. Why, he asks, is our space three-dimensional? He concludes that somehow this is bound up with the fact that forces such as gravity move through space, from a point of origin, like expanding spheres. Their strength varies inversely with the square of the distance. Had God chosen to create a world in which forces varied inversely with the *cube* of the distance, a space of four dimensions would have been required. (Similarly, though Kant did not mention it, forces in 2-space, moving out from a point source in expanding circles, would vary only inversely with the distance.) Kant here adopted a view of space which had been put forth a century earlier by Gottfried Wilhelm von Leibnitz, the great German philosopher and mathematician. Space has no reality apart from material things; it is nothing more than an abstract, mathematical description of relations that hold between objects. Although the notion of a fourth dimension had occurred to mathematicians, it had been quickly dropped as a fanciful speculation of no possible value. No one had hit on the fact that an asymmetric solid object could (in theory) be reversed by rotating it through a higher space; it was not until 1827, eighty years after Kant’s paper, that this was first pointed out by August Ferdinand Moebius, the German astronomer for whom the Moebius strip is named. It is surprising, therefore, to find Kant writing as early as 1747: “A science of all these possible kinds of space [spaces of more than three dimensions] would undoubtedly be the highest enterprise which a finite understanding could undertake in the field of geometry.” . . . “If it is possible,” he adds, “that there are extensions with other dimensions, it is also very probable that God has somewhere brought them into being; for His works have all the magnitude and manifoldness of which they are capable.” Such higher spaces would, however, “not belong to our world, but must form separate worlds.”

In 1768, in a paper *On the First Ground of the Distinction of Regions in Space*, Kant abandoned the Leibnitzian view of space for the Newtonian view. Space is a fixed, absolute thing – the “ether” of the nineteenth century – with a reality of its own, independent of material objects. To establish the existence of such a space, Kant turned his attention toward what he called “incongruent counterparts” – asymmetric solid figures of identical size and shape but opposite handedness, such as snail shells, twining plants, hair whorls, right and left hands. The existence of such twin objects, he argued, implies a Newtonian space. To prove it, he made use of a striking thought experiment which can be stated as follows.

Imagine that the cosmos is completely empty save for one single human hand. Is it a left or right hand? Since there are no intrinsic, measurable differences between enantiomorph objects, we have no basis for calling the hand left or right. Of course if you imagine yourself looking at the hand, naturally you will see it as either left or right, but that is equivalent to putting yourself (with your sense of handedness) into 3-space. You must imagine the hand in space to be completely removed from all relationships with other geometrical structures. Clearly it would be as meaningless to say that the hand is left or right as it would be to say it is large or small, or oriented with its fingers pointing up or down.

Suppose now that a human body materializes in space near the hand. The body is complete except for both hands; they have been severed at the wrist and are missing. It is evident that the hand will not fit both wrists. It will fit only one – say the left wrist. Therefore it is a left hand. Do you see the paradox confronting us? If it proves to be a left hand, by virtue of fitting the left wrist, it must have been a left hand *before* the body appeared. There must be some basis, some ground, for calling it “left” even when it is the sole object in the universe! Kant could see no way of providing such a ground except by assuming that space itself possessed some sort of absolute, objective structure – a kind of three-dimensional lattice that could furnish a means of defining the handedness of a solitary, asymmetric object.

A modern reader, familiar with n -dimensional geometry, should have little trouble seeing through the verbal confusion of Kant’s thought experiment. In fact, while I was writing this chapter, Kant’s error was effectively exposed by an episode in Johnny Hart’s syndicated comic strip called *B.C.*, in newspapers of July 26, 1963. One of Hart’s cavemen has just invented the drum. He strikes a log with a stick held in one hand and says, “That’s a left flam.” Then he hits the log with a stick in his other hand and says, “That’s a right flam.”

“How do you know which is which?” asks a spectator.

The drummer points to the back of one hand and replies, “I have a mole on my left hand.”

Let us see how this relates to Kant’s error. Imagine that Flatland contains nothing but a single, flat hand. It is true that it is asymmetric, but it is meaningless to speak of it as left or right if there is no other asymmetric structure on the plane. This is evident from the fact that we in 3-space can view the hand from either side of the plane and see it in either of its two mirror-image forms. The situation changes if we introduce a handless Flatlander and *define* “left” as, say, the side on which his heart is located. This by no means entails that the hand was “left” or “right” before introducing the Flatlander, because *we can introduce him in either of two enantiomorphic ways*. Place him in the plane one way, the hand becomes a left hand. Turn him over, place him the other way, and the hand becomes a right hand – “right” because it will fit the wrist on the side opposite the heart.

Does this mean that the hand alters its handedness, or that the Flatlander’s heart magically hops from one side of his body to the other? Not at all. Neither the hand nor the Flatlander changes in any respect. It is simply that their relations to each other in 2-space are changed. It is all a matter of words. “Left”

and “right” are words which mean, as Humpty Dumpty said, whatever we want them to mean. The solitary hand can be labeled with either term. So can the sides of a solitary Flatlander. It is only when the two asymmetric objects are present in the same space, and a choice of labels has been made with respect to one that labels applied to the other cease to be arbitrary.

It is the same in 3-space. Not until we introduce the handless body, with the understanding that “left” is the side the heart is on, do we have a basis for deciding what to call the hand. If the body is “turned over” by rotating it through 4-space, the hand’s label automatically changes. Suppose we first label the solitary hand, calling it, say, a “right” hand. When the body appears, its “right” wrist will be, by simple definition, the wrist on which the hand fits. The important point is that the initial choice of terms is wholly arbitrary. Hart’s caveman who chose to call one hand “left” because it had a mole on it was making a completely rational first step in defining handedness. The humor of the strip lies in the way the caveman phrased his reply. Instead of saying that he knew the difference between left and right flams because he had a mole on his left hand, he should have said: “Because I have decided to call ‘left’ the hand that has a mole on it.” There is nothing paradoxical about such a situation, therefore no need to introduce Newton’s absolute space.

Actually, even a fixed, Newtonian ether is no help in providing a label for the solitary hand unless the structure of space itself is somehow asymmetrical. If the hand floats inside a spherical, cylindrical, or conical cosmos, or in an infinite space criss-crossed with the lines of a cubical lattice, we are no better off than before. If the cosmos has the shape of one enormous human hand, the situation changes. We could call the cosmic hand “right” (or “plus” or “Yin”); then, if the solitary human hand is of opposite handedness, we are forced to call it “left” (or “minus” or “Yang”). We could also define the hand’s handedness on the basis of an asymmetric “grain” in space, a submicroscopic lattice of geodesics (straightest possible paths) like the asymmetric lattice of quartz or cinnabar. . . .

13 Incongruent Counterparts and Higher Dimensions

James Van Cleve

Incongruent Counterparts and Absolute Space

Incongruent counterparts are asymmetrical objects that come in mirror-image forms, such as left and right human hands. In a paper published in 1768, Kant argued that the existence of such objects is relevant to the debate between Newton and Leibniz on the ontological status of space.¹ Newton regarded space

as a thing in its own right – a vast aetherial container without walls, in which everything else that exists lives and moves and has its being. Leibniz believed that the existence of such an entity would generate various absurd pseudopossibilities, such as the possibility that the entire material cosmos might have been shifted three miles to the east or four miles to the west. How could God have had any reason to actualize one rather than another of these possibilities? Accordingly, Leibniz proposed that space is not a genuine entity, but simply a *façon de parler*; all talk of space may be replaced by talk of the spatial relations among material things. Kant agreed with Leibniz early in his career, but reflection on incongruent counterparts led him to believe that Newton was right after all – space is an absolute being, not just a system of relations.

The argument of Kant's 1768 paper may be assembled and set forth as follows:

- (1) A hand is left or right (as the case may be) either (a) solely in virtue of the *internal* relations among the parts of the hand or (b) at least partly in virtue of the *external* relations of the hand to something outside it – if not other material objects, then space itself.
- (2) But a hand is not left or right solely in virtue of its internal relations, since these are the *same* for right and left. ("The right hand is similar and equal to the left hand. And if one looks at one of them on its own, examining the proportion and the position of its parts to each other, and scrutinising the magnitude of the whole, then a complete description of the one must apply in all respects to the other, as well.")²
- (3) Nor is a hand right or left even partly in virtue of its relations to other material objects, since a hand that was all alone in the universe would still be right or left. ("Imagine that the first created thing was a human hand. That hand would have to be either a right hand or a left hand.")³
- (4) Therefore, a hand is left or right (as the case may be) at least partly in virtue of its relation to absolute space. ("Our considerations, therefore, make it clear that differences, and true differences at that, can be found in the constitution of bodies . . . [which] relate exclusively to *absolute* and *original* space . . .")⁴

Since the argument is plainly valid, there are four possible responses to it. (i) One may reject the first premise, maintaining that left and right do not consist in relations of any sort, but are irreducible intrinsic properties. This view is not very plausible, and I do not know of anyone who has advocated it. (ii) One may reject the second premise, maintaining that right and left, though intrinsic properties of hands as wholes, consist in relations among a hand's own parts. Call this view *internalism*. (iii) One may reject the third premise, maintaining that right and left consist in relations to material objects outside the hand. Call this view *externalism*. (iv) One may accept the conclusion, and with it the existence of absolute space. Call this view *absolutism*. In recent critical commentary on

Kant's argument, internalism has been advocated by Earman, externalism by Gardner, and absolutism by Nerlich.⁵

Kant rejected internalism because he thought the relevant internal relations were limited to distances between points and angles between lines; these relations are indeed the same in a left hand as in a right. Defenders of internalism may protest that Kant has overlooked a key internal relation, namely, the *direction* in which some points lie from others. May we not say that as you look at the palms of your hands, the direction from thumb to fingertips to wrist is clockwise in the left hand and counterclockwise in the right? But advocates of the other two positions we have distinguished would question whether direction is really an internal relation. Externalists would say that direction can be defined only by reference to an outside material thing (e.g., a clock), and Kant himself maintained that it can be defined only by reference to space as an entity. ("The direction, however, in which this order of parts is orientated, refers to the space outside the thing.")⁶

One phenomenon relevant to the evaluation of Kant's argument is *the fourth dimension*. For Kant, that space has three dimensions is the very paradigm of a synthetic but necessary proposition. But for many thinkers since Kant, propositions about the topological structure of space, such as this one, are contingent, and dimensions beyond the familiar three are perfectly possible. How would the possibility of a fourth spatial dimension bear on Kant's argument?

As the reader will know from the preceding selection by Gardner, in a space of four dimensions an object like a hand could be flipped over so as to become its own incongruent counterpart. The point is readily grasped with the help of lower-dimensional analogs. If tokens of the letters "p" and "q" are confined to a two-dimensional sheet of newsprint, neither can be twisted or turned so as to make it occupy the space of the other; but if we are permitted to lift one of the letters out of the plane of the page and turn it over, the feat can be accomplished. Just so in a four-dimensional space: a left hand could be turned around so as to fit in the space now occupied by a right. Glovemakers would no longer have to manufacture separate left and right models!

These facts about what is possible given an extra dimension are relevant to Kant's argument in two ways. First, they furnish a new argument against internalism. For internalism, the rightness of a hand is an intrinsic property of it. (Even though it consists in relations among the parts, it is intrinsic to the hand as a whole, as in our unnamed alternative (i) above.) Intrinsic properties are those that are shared between an object and any duplicate of it, and whether two objects are duplicates is not affected simply by moving one of them about. (The hardware store clerk has complied with my request to make a duplicate of my key even if the original I handed to her was pointing toward the ceiling and the one she returns to me is pointing toward the floor.) Since in four-dimensional space you could convert a right hand into a left simply by moving it about, it follows that rightness and leftness are not intrinsic properties. So internalism is refuted.

Secondly, the possibility of a fourth dimension undercuts the thought-experiment of the solitary hand that Kant used against externalism. Kant

maintained that a hand all alone in the universe would necessarily be left or right, in which case its rightness or leftness could not consist in its relations to any other material thing. Gardner and others have denied this, maintaining that the first created hand would be neither right nor left; it would become one or the other only upon the introduction of a second hand (or a human body for the hand to attach to). Though perhaps implausible at first blush, Gardner's position gains enormously in force once we accept the possibility of a fourth dimension. If four-dimensional spaces are possible (even if not actual), then the difference between a right and a left hand comes to no more than the difference between a "p" and a "q," or between an arrow pointing up and an arrow pointing down. That is to say, it is merely a matter of orientation, and orientation seems inherently to involve a relation to something else. Two objects can be alike or different in their orientation, but an object considered by itself has no orientation at all. So externalism appears to be vindicated.

Another phenomenon relevant to the evaluation of Kant's argument is *the fall of parity*. This refers to the finding, first made in the 1950s, that some laws of nature are sensitive to the distinction between right and left. For example, some particles more often decay into a left-handed configuration than a right, and the outcome of some physical processes can depend on whether the initial conditions assume a right- or a left-handed form. It would not be more surprising in principle if we were to discover that a left glove, when tossed into the air, generally lands palm up, whereas a right lands palm down.

The fall of parity has been adduced as showing that Kant was wrong to maintain that the difference between right and left can only be grasped ostensively, in which case it presumably could not be communicated in binary code to a distant galaxy. We *could* communicate the difference (according to Kant's critics) simply by sending the directions for one of the parity-violating experiments: "Let a bunch of X-particles decay; the decay configuration you will get most often is the one we call left."⁷

However that may be, it seems to me that the fall of parity shows that Kant was correct about something else: namely, that externalism is false. If being right or left is only a matter of being the same or different in orientation as some other object, how can it be a law of nature that certain processes always (or even just usually) have left-handed outcomes? That would be like a law of nature that instructs a seed to grow, not into a watermelon vine or something of a certain intrinsic description, but into the same type of plant that a neighboring seed will grow into.⁸

If the points briefly developed above are correct, a significant conclusion emerges. I have suggested (1) that in response to Kant's argument we must be either internalists, externalists, or absolutists; (2) that externalism is refuted by the fall of parity; and (3) that internalism would be refuted if four-dimensional spaces were possible. It follows either that absolutism is correct (as Kant maintained) or that four-dimensional spaces are impossible (as he also maintained). So Kant was right about at least one thing.

Dialogue on Higher Dimensions

Treios: Let me tell you my latest argument against the fourth dimension. It occurred to me as I was reading what Martin Gardner has to say about the fourth dimension in connection with Kant's puzzle about incongruent counterparts – things like left and right hands. He points out that in a four-dimensional space, you could turn a right hand over so as to make it a left.

Philomath: Yes, mathematicians have known that ever since Moebius. So what's your argument?

Treios: It follows that in four-dimensional space, right and left hands would not be incongruent after all. They would be intrinsically alike, just differently oriented, like arrows pointing in opposite directions. But the difference between right and left is obviously more profound than that. The rightness or leftness of a hand is an intrinsic, recognizable property of it; you can tell that a hand is right or left just by looking at it alone.

Philomath: I disagree. You can't tell that a single hand is right because a single hand can't *be* right. It makes sense to call a hand right or left only in relation to another hand.

Treios: I don't see how you can say that. If I toss you a glove, you can tell me instantly whether it's right or left without comparing it with another glove.

Philomath: Let me qualify my position. You can judge a hand or glove to be right or left without reference to another hand or glove, but you still have to refer to *some* other asymmetrical object, if only your own body.

Treios: I don't see the need for a body, either. A disembodied observer could still tell whether a hand was left or right.

Philomath: I'll grant you that for the sake of argument. I still maintain that your observer would need a vantage point, and could judge a hand to be right or left only in relation to his vantage point.

Treios: I'll grant the need for a vantage point, but I don't see why the observer's judgments of left or right would have to be relative to it. A right hand presents the same recognizable *Gestalt* to *all* vantage points.

Philomath: Not at all. To an observer capable of moving about in four-dimensional space, a hand would appear sometimes as right, sometimes as left.

Treios: But the fourth dimension is just what I am arguing against.

Philomath: Precisely – so your argument begs the question. Don’t you see that a Flatlander could use the same argument against the possibility of three-dimensional space? “There is an intrinsic difference between a p and a q,” he might claim, “and you can tell which sort of letter you are dealing with by looking at it alone.” But viewers outside the plane of Flatland could see the same letter now as a p, now as a q, just as easily as we can go around to the other side of a shop window.

Treios: I see you are going to keep bringing in the Flatlanders. I’ll read *Flatland* before we meet again.

* * *

Treios: I have now read *Flatland*.⁹ It’s a marvelous little book, but it hasn’t convinced me. Do you want to hear my main objection to it?

Philomath: Shoot.

Treios: I think we should distinguish between not seeing the possibility of something and seeing the impossibility of it. I think the condition of the Flatlanders is simply inability to see the possibility of three-dimensional space, which I admit does not give them good reason to disbelieve in it. But I think what we possess in regard to the fourth dimension is something more than that – not just inability to visualize it, but positive insight into its impossibility.

Philomath: A dubious distinction, if you ask me.

Treios: Let me illustrate it for you. Do you see that it is possible for there to be a closed curve no four points of which are the vertices of a square?

Philomath: No; it is an unsolved problem whether that is possible.

Treios: Do you see it to be *impossible*, then?

Philomath: Of course not. I just said that it is an unsolved problem.

Treios: Very well; now let’s change the example. Do you see that it is possible for there to be cube with 13 edges?

Philomath: That’s absurd.

Treios: Just as I expected. In the first case, you don’t see the possibility of a thing, but you don’t see its impossibility, either. In the second case, you positively see that it is impossible for a cube to have 13 edges.

Philomath: Seeing has nothing to do with it; it’s just that if something had 13

edges we wouldn't call it a cube. But let's get back to the fourth dimension. How do you claim to see that it's impossible?

Treios: Try to visualize more than three perpendicular lines meeting at the corner of this desk – one more line coming in at right angles to each of the other three edges. I bet you can't do it.

Philomath: (*Shrugs.*) Agreed.

Treios: Now is it just that you can't see how to fit a fourth line in? Or is it something positive – you see that there is no place for it to go?

Philomath: Whichever it is, I don't set any store by it. Mathematicians can prove all kinds of interesting properties of four-dimensional figures, and the resulting geometry is perfectly consistent. You are trying to set limits on possibility that are narrower than those of logical consistency, but consistency is the only game in town.

* * *

Treios: Let me ask you this. Are you as willing to believe in negative dimensions as you are in higher positive dimensions? Two-dimensional planes are bounded by lines of dimension one and lines are bounded by points of dimension zero; might points be bounded by items of dimension minus-one?

Philomath: I have no use for negative dimensions.

Treios: My question is whether you think them possible.

Philomath: No, but it is not for any reason you can use against the fourth dimension. Points are not bounded by anything at all. As Euclid says, "a point is that which hath no parts."

Treios: That only raises the question whether the zero-dimensional entities we normally call points really are points in Euclid's sense. Maybe items of dimension zero are bounded by items of dimension minus-one, which are in turn bounded by items of dimension minus-two, and so on.

Philomath: I can't make any sense of that.

Treios: But I'm only extending in the downward direction the analogies you are so happy with going up.

Philomath: OK, I take back my opposition to negative dimensions. Though at present I have no conception of them, I don't say they are impossible. Maybe some day the mathematics of negative dimensions will be worked out.

Treios: It is hard to make headway against someone with so open a mind.

Philomath: It is harder against a closed mind, I assure you.

* * *

Treios: Let me try one more time to see if I can't get you to agree that you positively see one of the implications of higher dimensions to be impossible. In three-dimensional space, a one-dimensional loop or an infinite line does not suffice to separate one part of space from the rest, but a two-dimensional surface, such as an infinite plane or the surface of a sphere, *does* separate one part of space from the rest. You can't get from the inside of the sphere to the outside without going through the surface.

Philomath: Yes; Poincaré used the generalization of that fact to define what it is for a space to have dimension n .

Treios: That's just what my argument is going to rely on. In four-dimensional space, a two-dimensional surface would not separate space into two parts. A box or a sphere (cupping his hands) would no longer completely enclose a region of space, just as a circle (making a circle with his thumb and forefinger) does not completely enclose any region of three-dimensional space. So in four-dimensional space, there would be a path by which a beetle could get out of a closed box without going through a wall. And *that*, I hope you will admit, is impossible.

Philomath: If there were a fourth dimension, there would be such a path.

Treios: Yes, that is one of my premises. To which I add, there can be no such a path, so there is no fourth dimension.

Philomath: I wasn't just repeating your premise. My point was that for all we know there could be a fourth dimension, so for all we know there could be such a path.

Treios: Well, I don't know what to say. That the beetle is completely surrounded by the box, so that there's no way out without going through a wall, is as obvious to me as anything ever gets.

Philomath: It's not obvious to everybody. Some people *do* succeed in visualizing the fourth dimension.

Treios: I have my doubts about that. From what I've read, I have the impression that their supposed "seeing" of the fourth dimension is really just a matter of interpreting certain figures or dances of lines on a computer screen as manifestations of something four-dimensional. If we saw a point expand to a sphere

and contract again to a point, we could say “This is how a hypersphere would look as it passed through three-dimensional space.” But it is also how a point would look as it grew to a sphere and shrank again to a point.

Philomath: It is not always just a matter of what you call interpretation. Sometimes something clicks and one sees a configuration as four-dimensional, just as when you see a Necker cube drawn on paper as three-dimensional.

Treios: I’ll believe that when it happens to me. In the meantime, I hope you’ll forgive me for doubting that it happens to others. When I find something inconceivable myself, I am bound to find it inconceivable that others find it conceivable.

Philomath: You just don’t get it. The lesson of Flatland is completely lost on you. There is not one of your arguments that a Flatlander could not use against the third dimension. “There is an intrinsic difference between a p and a q. There cannot be more than two lines meeting at right angles. A dot cannot escape from a square without passing through a side. Therefore, there can be no third dimension – and no one who visualizes it, either.” Every one of your arguments could be used by a Flatlander, and every one of them would be wrong.

Treios: You’re taking for granted just what I questioned, that the Flatlanders’ state of mind in regard to the third dimension could be something positive like ours in regard to the fourth. In that case, isn’t what you are presenting me with a completely generalizable skeptical argument? Isn’t it possible to challenge *any* claim to knowledge, however firm and convincing its grounds, by dreaming up beings who would have similar grounds and be mistaken? As easily as we can imagine Flatlanders who are deluded about the structure of space, can we not imagine beings who are deluded about the basic laws of logic? Who think they see that it is impossible for something to have both color and shape – to be both red and square, for example? And could we not then challenge our own belief in the impossibility of contradictions by saying that for all we know our situation may be like theirs? If this is your position, I have no answer to it. I can only point out that it leaves us knowing practically nothing but the Cartesian *cogito*. If it is *not* your position, I wonder why you are so selective about which of our intuitions you will let us rely upon.

Philomath: I must get to class. I’ll think about that and see you again tomorrow.

Notes

- 1 Immanuel Kant, “Concerning the Ultimate Ground of the Differentiation of Directions in Space,” in *Theoretical Philosophy, 1755–1770*, vol. I in *The Cambridge Edition of the Works of Immanuel Kant*, translated and edited by David Walford in

- collaboration with Ralf Meerbote (Cambridge: Cambridge University Press, 1992), pp. 365–72.
- 2 Ibid., p. 370.
 - 3 Ibid., p. 371, correcting one error.
 - 4 Ibid., p. 371.
 - 5 John Earman, “Kant, Incongruous Counterparts, and the Nature of Space and Space-Time,” *Ratio*, 13 (1971), pp. 1–18; Martin Gardner, this volume; Graham Nerlich, “Hands, Knees, and Absolute Space,” ch. 2 of *The Shape of Space* (Cambridge: Cambridge University Press, 1976). Relevant work by these authors and others is reprinted in *The Philosophy of Right and Left*, edited by James Van Cleve and Robert E. Frederick (Dordrecht: Kluwer Academic Publishers, 1991). Nerlich actually subscribes to externalism as regards the properties of leftness and rightness, but he holds that this is an argument analogous to Kant’s works for a different pair of properties, being enantiomeric and being homomorphic. Roughly, an object is enantiomeric if it could have an incongruent counterpart and homomorphic otherwise.
 - 6 Walford, *Theoretical Philosophy*, p. 365. In many earlier translations, the German word *Gegend* is misleadingly rendered as “region” rather than “direction.” On the reasons for preferring “direction” to “region,” see Walford, pp. 456–7.
 - 7 See Martin Gardner, “The Ozma Problem and the Fall of Parity,” in Van Cleve and Frederick, *Philosophy of Right and Left*, pp. 75–95, or Gardner’s *The New Ambidextrous Universe*, 3rd edn (New York: W. H. Freeman, 1990).
 - 8 For further development of this point, see my “Introduction to the Arguments of 1770 and 1783,” in Van Cleve and Frederick, *Philosophy of Right and Left*, pp. 15–26, especially pp. 20–22.
 - 9 Edward A. Abbott, *Flatland* (New York: Dover Books, 1952), first published in 1884.
-

14 Achilles and the Tortoise*

Max Black

1. Suppose Achilles runs ten times as fast as the tortoise and gives him a hundred yards’ start. In order to win the race Achilles must first make up for his initial handicap by running a hundred yards; but when he has done this and has reached the point where the tortoise started, the animal has had time to advance ten yards. While Achilles runs these ten yards, the tortoise gets one yard ahead; when Achilles has run this yard, the tortoise is a tenth of a yard ahead; and so on, without end. Achilles never catches the tortoise, because the tortoise always holds a lead, however small.

This is approximately the form in which the so-called “Achilles” paradox has come down to us. Aristotle, who is our primary source for this and the other

* From Max Black, “Achilles and the Tortoise,” *Analysis*, 11 (1951), pp. 91–101.

paradoxes attributed to Zeno, summarizes the argument as follows: "In a race the quickest runner can never overtake the slowest, since the pursuer must first reach the point whence the pursued started, so that the slower must always hold a lead" (*Physics*, 239b).

2. It would be a waste of time to prove, by independent argument, that Achilles *will* pass the tortoise. Everybody knows this already, and the puzzle arises because the conclusion of Zeno's argument is known to be absurd. We must try to find out, if we can, exactly what mistake is committed in this argument.

3. A plausible answer that has been repeatedly offered takes the line that "this paradox of Zeno is based upon a mathematical fallacy" (A.N. Whitehead, *Process and Reality* [New York, 1929], p. 107).

Consider the lengths that Achilles has to cover, according to our version of the paradox. They are, successively, a hundred yards, ten yards, one yard, a tenth of a yard, and so on. So the total number of yards he must travel in order to catch the tortoise is

$$100 + 10 + 1 + \frac{1}{10} + \dots$$

This is a convergent geometrical series whose sum can be expressed in decimal notation as 11.1, that is to say exactly $11\frac{1}{9}$. When Achilles has run this number of yards, he will be dead level with his competitor; and at any time thereafter he will be actually ahead.

A similar argument applies to the time needed for Achilles to catch the tortoise. If we suppose that Achilles can run a hundred yards in ten seconds, the number of seconds he needs in order to draw level is

$$10 + 1 + \frac{1}{10} + \frac{1}{100} + \dots$$

This, too, is a convergent geometrical series, whose sum is expressed in decimal notation as 11.1̄, that is to say exactly $11\frac{1}{9}$. This, as we should expect, is one tenth of the number we previously obtained for the length of the race. (For Achilles was running at ten yards per second.)

We can check the calculation without using infinite series at all. The relative velocity with which Achilles overtakes the tortoise is nine yards per second. Now the number of seconds needed to cancel the initial gap of a hundred yards at a relative velocity of pursuit of nine yards per second is 100 divided by 9, i.e., $11\frac{1}{9}$. This is exactly the number we previously obtained by summing the geometrical series representing the times needed by Achilles. Achilles is actually running at ten yards per second, so the actual distance he travels is $10 \times 11\frac{1}{9}$, or $111\frac{1}{9}$, as before. Thus we have confirmed our first calculations by an argument not involving the summation of infinite series.

4. According to this type of solution, the fallacy in Zeno's argument is due to the use of the words "never" and "always." Given the premise that "the pursuer must first reach the point whence the pursued started," it does *not* follow, as alleged, that the quickest runner "never" overtakes the slower: Achilles does catch the tortoise at some time – that is to say at a time exactly $1\frac{1}{9}$ seconds from the start. It is wrong to say that the tortoise is "always" in front: there is a place – a place exactly $1\frac{1}{9}$ yards from Achilles' starting point – where the two are dead level. Our calculations have shown this, and Zeno failed to see that only a finite time and finite space are needed for the infinite series of steps that Achilles is called upon to make.

5. This kind of mathematical solution has behind it the authority of Descartes and Peirce and Whitehead – to mention no lesser names – yet I cannot see that it goes to the heart of the matter. It tells us, correctly, when and where Achilles and the tortoise will meet, *if* they meet; but it fails to show that Zeno was wrong in claiming they *could not* meet.

Let us be clear about what is meant by the assertion that the sum of the infinite series

$$100 + 10 + 1 + \frac{1}{10} + \frac{1}{100} + \dots$$

is $1\frac{1}{9}$. It does not mean, as the naïve might suppose, that mathematicians have succeeded in adding together an infinite number of terms. As Frege pointed out in a similar connection,¹ this remarkable feat would require an infinite supply of paper, an infinite quantity of ink, and an infinite amount of time. If we had to add all the terms together, we could never prove that the series had a finite sum. To say that the sum of the series is $1\frac{1}{9}$ is to say that if enough terms of the series are taken, the difference between the sum of that *finite number* of terms and the number $1\frac{1}{9}$ becomes, and stays, as small as we please. (Or to put it another way: Let n be any number less than $1\frac{1}{9}$. We can always find a finite number of terms of the series whose sum will be less than $1\frac{1}{9}$ but greater than n .)

Since this is all that is meant by saying that the infinite series has a sum, it follows that the "summation" of all the terms of an infinite series is not the same thing as the summation of a finite set of numbers. In one case we can get the answers by working out a finite number of additions; in the other case we *must* "perform a limit operation," that is to say, we must prove that there is a number whose difference from the sum of the initial members of the series can be made to remain as small as we please.

6. Now let us apply this to the race. The series of distances traversed by Achilles is convergent. This means that if Achilles takes enough steps whose sizes are given by the series one hundred yards, ten yards, one yard, one-tenth yard, etc., the distance *still to go* to the meeting point eventually becomes, and stays, as small as we please. After the first step he still has $1\frac{1}{9}$ yards to go; after the second, only $1\frac{1}{90}$ yard; after the third, no more than $\frac{1}{9}$ yard; and so

on. The distance still to go is reduced by nine-tenths at each move.

But the distance, however much reduced, still remains to be covered; and after each step there are infinitely many steps still to be taken. The logical difficulty is that Achilles seems called upon to perform *an infinite series of tasks*; and it does not help to be told that the tasks become easier and easier, or need progressively less and less time in the doing. Achilles may get nearer to the place and time of his rendezvous, but his task remains just as hard, for he still has to perform what seems to be logically impossible. It is just as hard to draw a very small square circle as it is to draw an enormous one: we might say both tasks are infinitely hard. The logical difficulty is not in the extent of the distance Achilles has to cover but in the apparent impossibility of his traveling any distance whatsoever. I think Zeno had enough mathematical knowledge to understand that if Achilles could run $11\frac{1}{2}$ yards – that is to say, keep going for $1\frac{1}{2}$ seconds – he would indeed have caught the tortoise. The difficulty is to understand how Achilles could arrive anywhere at all without first having performed an infinite series of acts.

7. The nature of the difficulty is made plainer by a second argument of Zeno, known as the “Dichotomy” which, according to Aristotle, is “the same in principle” (*Physics*, 239b). In order to get from one point to another, Achilles must first reach a third point midway between the two; similarly, in order to reach this third point he must first reach a fourth point; to reach this point he must first reach another point; and so on, without end. To reach *any* point, he must first reach a nearer one. So, in order to be moving at all, Achilles must already have performed an infinite series of acts – must, as it were, have traveled along the series of points from the infinitely distant and *open* “end.” This is an even more astounding feat than the one he accomplishes in winning the race against the tortoise.

The two arguments are complementary: the “Achilles” shows that the runner cannot reach any place, even if he gets started; while the “Dichotomy” shows that he cannot get started, i.e., cannot leave any place he has reached.

8. Mathematicians have sometimes said that the difficulty of conceiving the performance of an infinite series of tasks is factitious. All it shows, they say, is the weakness of human imagination and the folly of the attempt to make a mental image of mathematical relationships. The line really does have infinitely many points, and there is no logical impediment to Achilles’ making an infinite number of steps in a finite time. I shall try to show that this way of thinking about the race is untenable.

9. I am going to argue that the expression, “infinite series of acts,” is self-contradictory, and that failure to see this arises from confusing a series of acts with a series of numbers generated by some mathematical law. (By an “act” I mean something marked off from its surroundings by having a definite beginning and end.)

In order to establish this by means of an illustration I shall try to make plain

some of the absurd consequences of talking about “counting an infinite number of marbles.” And in order to do this I shall find it convenient to talk about counting an infinite number of marbles as if I supposed it was sensible to talk in this way. But I want it to be understood all the time that I do not think it sensible to talk in this way, and that my aim in so talking is to show how absurd this way of talking is. Counting may seem a very special kind of “act” to choose, but I hope to be able to show that the same considerations apply to an infinite series of any kind of acts.

10. Suppose we want to find out the number of things in a given collection,² presumably identified by some description. Unless the things are mathematical or logical entities it will be impossible to deduce the size of the collection from the description alone; and somebody will have to do the work of taking a census. Of course he can do this without having any idea of how large the collection will turn out to be: his instructions may simply take the form, “Start counting and keep on until there is nothing left in the collection to count.” This implies that there will be a point at which there will be “nothing left to count,” so that the census-taker will then know his task to have been completed.

Now suppose we can know that the collection is infinite. If, knowing this, we were to say, “Start counting, and continue until there is nothing left to count” we should be practicing a deception. For our census-taker would be led to suppose that sooner or later there would be nothing left to count, while all the time we would know this supposition to be false. An old recipe for catching guinea pigs is to put salt on their tails. Since they have no tails, this is no recipe at all. To avoid deception we should have said, in the case of the infinite collection, “Start counting and *never stop*.” This should be enough to tell an intelligent census-taker that the collection is infinite, so that there is no sense in trying to count it.

If somebody says to me, “Count all the blades of grass in Hyde Park,” I might retort, “It’s too difficult; I haven’t enough time.” But if some cosmic bully were to say, “Here is an infinite collection; go ahead and count it,” only logical confusion could lead me to mutter, “Too difficult; not enough time.” The trouble is that, no matter what I do, the result of all my work will not and cannot count as compliance with the instructions. If somebody commands me to obey a certain “instruction,” and is then obliging enough to add that nothing that I can do will count as compliance with that instruction, only confusion could lead me to suppose that any genuine task had been set.

11. However, some writers have said that the difficulty of counting an infinite collection is just a matter of *lack of time*. If only we could count faster and faster, the whole job could be done in a finite time; there would still never be a time at which we were ending, but there would be a time at which we already would have ended the count. It is not necessary to finish counting; it is sufficient that the counting shall have been finished.

Very well. Since the task is too much for human capacity, let us imagine a

machine that can do it. Let us suppose that upon our left a narrow tray stretches into the distance as far as the most powerful telescope can follow; and that this tray or slot is full of marbles. Here, at the middle, where the line of marbles begins, there stands a kind of mechanical scoop; and to the right, a second, but empty tray, stretching away into the distance beyond the farthest reach of vision. Now the machine is started. During the first minute of its operation, it seizes a marble from the left and transfers it to the empty tray on the right; then it rests a minute. In the next half-minute the machine seizes a second marble on the left, transfers it, and rests half-a-minute. The third marble is moved in a quarter of a minute, with a corresponding pause; the next in one-eighth of a minute; and so until the movements are so fast that all we can see is a gray blur. But at the end of exactly four minutes the machine comes to a halt, and now the left-hand tray that was full seems to be empty, while the right-hand tray that was empty seems full of marbles.

Let us call this an *infinity machine*. And since it is the first of several to be described let us give it the name "Alpha."

12. I hope nobody will object that the wear and tear on such a machine would be too severe; or that it would be too hard to construct. We are dealing with the logical coherence of ideas, not with the practicability of mechanical devices. If we can conceive of such a machine without contradiction, that will be enough; and believers in the "actual infinite" will have been vindicated.

13. An obvious difficulty in conceiving of an infinity machine is this. How are we supposed to know that there are infinitely many marbles in the left-hand tray at the outset? Or, for that matter, that there are infinitely many on the right when the machine has stopped? Everything we can observe of Alpha's operations (and no matter how much we slow it down) is consistent with there having been involved only a very large, though still finite, number of marbles.

14. Now there is a simple and instructive way of making certain that the machine shall have infinitely many marbles to count. Imagine the arrangements modified as follows. Let there be only *one* marble in the left-hand tray to begin with, and let some device always return *that same marble* during the time at which the machine is resting. Let us give the name "Beta" to a machine that works in this way. From the standpoint of the machine, as it were, the task has not changed. The difficulty of performance remains exactly the same whether the task, as in Alpha's case, is to transfer an infinite series of qualitatively similar but different marbles; or whether the task, as in Beta's case, is constantly to transfer the *same* marble – a marble that is immediately returned to its original position. Imagine Alpha and Beta set to work side by side on their respective tasks: every time the one moves, so does the other; if one succeeds in its task, so must the other; and if it is impossible for either to succeed, it is impossible for *each*.

15. The introduction of our second machine, Beta, shows clearly that the infinite count really is impossible. For the single marble is always returned, and

each move of the machine accomplishes nothing. A man given the task of filling three holes by means of two pegs can always fill the third hole by transferring one of the pegs; but this automatically creates another empty place, and it won't help in the least to "keep on trying" or to run through this futile series of operations faster and faster. (We don't get any nearer to the end of the rainbow by running faster.) Now our machine, Beta, is in just this predicament: the very act of transferring the marble from left to right immediately causes it to be returned again; the operation is self-defeating and it is logically impossible for its end to be achieved. Now if this is true for Beta, it must be true also for Alpha, as we have already seen.

16. When Hercules tried to cut off the heads of Hydra, two heads immediately grew where one had been removed. It is rumored that the affair has been incorrectly reported: Zeus, the all powerful, took pity on Hercules and eased his labor. It was decreed that only *one* head should replace the head that had been cut off and that Hercules should have the magical power to slash faster and faster in geometrical progression. If this is what really happened, had Hercules any cause to be grateful? Not a bit. Since the head that was sliced off immediately grew back again, Hercules was getting nowhere, and might just as well have kept his sword in its scabbard.

17. Somebody may still be inclined to say that nevertheless when the machine Beta finally comes to rest (at the end of the four minutes of its operation) the single marble might after all be found in the right-hand tray, and this, if it happened, would *prove* that the machine's task had been accomplished. However, it is easy to show that this suggestion will not work.

I said, before, that "some device" always restored the marble to its original position in the left-hand tray. Now the most natural device to use for this purpose is another machine – Gamma, say – working like Beta but *from right to left*. Let it be arranged that no sooner does Beta move the marble from left to right than Gamma moves it back again. The successive working periods and pauses of Gamma are then equal in length to those of Beta, except that Gamma is working while Beta is resting, and vice versa. The task of Gamma, moreover, is exactly parallel to that of Beta, that is, to transfer the marble an infinite number of times from one side to the other. If the result of the whole four minutes' operation by the first machine is to transfer the marble from left to right, the result of the whole four minutes' operation by the second machine must be to transfer the marble from right to left. But there is only one marble and it must end somewhere. If it ought to be found on the right, then by the same reasoning it ought to be found on the left. But it cannot be both on the right and on the left. Hence neither machine can accomplish its task, and our description of the infinity machines involves a contradiction.

18. These considerations show, if I am not mistaken, that the outcome of the infinity machine's work is independent of what the machine is supposed to have done antecedently. The marble might end up on the right, on the left, or

nowhere. When Hercules ended his slashing, Zeus had to decide whether the head should still be in position or whether, after all, Hercules' strenuous efforts to do the impossible should be rewarded.

Hercules might have argued that every time a head appeared, he had cut it off, so no head ought to remain; but the Hydra could have retorted, with equal force, that after a head had been removed another had always appeared in its place, so a head ought to remain in position. The two contentions cancel one another and neither would provide a ground for Zeus' decision.

Even Zeus, however, could not abrogate the continuity of space and motion; and this, if I am not mistaken, is the source of the contradiction in our description of the machine Beta. The motion of the marble is represented, graphically, by a curve with an infinite number of oscillations, the rapidity of the oscillations increasing constantly as approach is made to the time at which the machine comes to rest. Now to say that motion is continuous is to deny that any real motion can be represented by a curve of this character. Yet every machine that performed an infinite series of acts in a finite time would have to include a part that oscillated "infinitely fast," as it were, in this impossible fashion. For the beginning of every spatio-temporal act is marked by a change in the velocity or in some other magnitude characterizing the agent.

19. It might be thought that the waiting intervals in the operations of the three infinity machines so far described have been essential to the argument. And it might be objected that the steps Achilles takes are performed consecutively and without intervening pauses. I will now show that the pauses or "resting periods" are not essential to the argument.

Consider for this purpose two machines, Delta and Epsilon, say, that begin to work with a single marble each, but in opposite directions. Let Delta start with the marble *a* and Epsilon with the marble *b*. Now suppose the following sequence of operations: while Delta transfers marble *a* from left to right in one minute, Epsilon transfers marble *b* from right to left; then Delta moves *b* from left to right in half a minute while Epsilon returns *a* from right to left during the same time; and so on, indefinitely, with each operation taking half the time of its predecessor. During the time that either machine is transporting a marble, its partner is placing the other marble in position for the next move.³ Once again, the total tasks of Delta and Epsilon are exactly parallel: if the first is to succeed, both marbles must end on the right, but if the second is to succeed, both must end on the left. Hence neither can succeed, and there is a contradiction in our description of the machines.

20. Nor will it help to have a machine – Phi, say – transferring marbles that become progressively smaller in geometrical progression.⁴ For, by an argument already used, we can suppose that while Phi is performing its operations, one of the machines already described is going through its paces at the same rates and at the same times. If Phi could complete its task, Alpha, Beta, Gamma, Delta and Epsilon would have to be able to complete their respective tasks. And we have already seen that this is not possible. The sizes of the successive tasks have

nothing to do with the logical impossibility of completing an infinite series of operations. Indeed it should be clear by this time that the logical possibility of the existence of any one of the machines depends upon the logical possibility of the existence of all of them or, indeed, of any machine that could count an infinite number of objects. If the idea of the existence of any one of them is self-contradictory, the same must be true for each of them. The various descriptions of these different hypothetical devices simply make it easier for us to see that one and all are logically impossible. And though a good deal more might be said about this, I hope I have said enough to show why I think this notion of counting an infinite collection is self-contradictory.

21. If we now reconsider for a moment the arguments that have been used in connection with our six infinity machines, we can easily see that no use was made of the respects in which counting differs from any other series of acts. Counting differs from other series of acts by the conventional assignment of numerals to each stage of the count, and in other respects, too. But every series of acts is like counting in requiring the successive doing of things, each having a beginning and end in space or time. And this is all that was used or needed in our arguments. Since our arguments in no way depended upon the specific peculiarities of counting they would apply, as I said at the outset, to any infinite series of acts.

22. And now let us return to Achilles. If it really were necessary for him to perform an infinite number of *acts* or, as Aristotle says “to pass over or severally to come in contact with infinite things” (*Physics*, 233a), it would indeed be logically impossible for him to pass the tortoise. But all the things he really does are finite in number; a finite number of steps, heart beats, deep breaths, cries of defiance, and so on. The track on which he runs has a finite number of pebbles, grains of earth, and blades of grass,⁵ each of which in turn has a finite, though enormous, number of atoms. For all of these are things that have a beginning and end in space or time. But if anybody says we must imagine that the atoms themselves occupy space and so are divisible “in thought,” he is no longer talking about spatio-temporal things. To divide a thing “in thought” is merely to halve the numerical interval which we have assigned to it. Or else it is to suppose – what is in fact physically impossible beyond a certain point – the actual separation of the physical thing into discrete parts. We can of course choose to say that we shall represent a distance by a numerical interval, and that every part of that numerical interval shall also count as representing a distance; then it will be true a priori that there are infinitely many “distances.” But the class of what will then be called “distances” will be a series of pairs of numbers, not an infinite series of spatio-temporal things. The infinity of this series is then a feature of one way in which we find it useful to *represent* the physical reality; to suppose that therefore Achilles has to *do* an infinite number of things would be as absurd as to suppose that because I can attach two numbers to an egg I must make some special effort to hold its halves together.

23. To summarize: I have tried to show that the popular mathematical refutation of Zeno's paradoxes will not do, because it simply assumes that Achilles can perform an infinite series of acts. By using the illustration of what would be involved in counting an infinite number of marbles, I have tried to show that the notion of an infinite series of acts is self-contradictory. For any material thing, whether machine or person, that set out to do an infinite number of acts would be committed to performing a motion that was discontinuous and therefore impossible. But Achilles is not called upon to do the logically impossible; the illusion that he must do so is created by our failure to hold separate the finite number of real things that the runner has to accomplish and the infinite series of numbers by which we describe what he actually does. We create the illusion of the infinite tasks by the kind of mathematics that we use to describe space, time, and motion.

Notes

- 1 *Grundgesetze der Arithmetik*, 2 (1903), §124. Or see my translation in *Translations from the Philosophical Writings of Gottlob Frege* (Oxford, 1952), p. 219.
 - 2 Or class or set or aggregate, etc.
 - 3 An alternative arrangement would be to have three similar machines constantly circulating three marbles.
 - 4 Somebody might say that if the marble moved by Beta eventually shrank to nothing there would be no problem about its final location.
 - 5 Cf. Peirce: "I do not think that if each pebble were broken into a million pieces the difficulty of getting over the road would necessarily have been increased; and I don't see why it should if one of these millions – or all of them – had been multiplied into an infinity" (*Collected Papers* [Cambridge, Mass., 1931], 6.182).
-

15 A Contemporary Look at Zeno's Paradoxes: An Excerpt from *Space, Time, and Motion**

Wesley C. Salmon

The Paradoxes of Motion

Our knowledge of the paradoxes of motion comes from Aristotle who, in the course of his discussions, offers a paraphrase of each. Zeno's original formulations have not survived.¹

* From Wesley C. Salmon, *Space, Time, and Motion* (Minneapolis: University of Minnesota Press, 1980). Reprinted by permission of the author.

- (1) *Achilles and the Tortoise.* Imagine that Achilles, the fleetest of Greek warriors, is to run a footrace against a tortoise. It is only fair to give the tortoise a head start. Under these circumstances, Zeno argues, Achilles can never catch up with the tortoise, no matter how fast he runs. In order to overtake the tortoise, Achilles must run from his starting point A to the tortoise's original starting point T_0 (see figure 3). While he is doing that, the tortoise will have moved ahead to T_1 . Now Achilles must reach the point T_1 . While Achilles is covering this new distance, the tortoise moves still farther to T_2 .

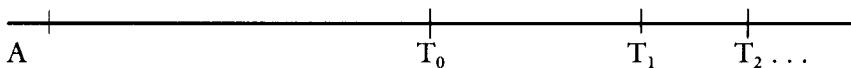


Figure 3

Again, Achilles must reach this new position of the tortoise. And so it continues; whenever Achilles arrives at a point where the tortoise *was*, the tortoise has already moved a bit ahead. Achilles can narrow the gap, but he can never actually catch up with him. This is the most famous of all of Zeno's paradoxes. It is sometimes known simply as "The Achilles."

- (2) *The Dichotomy.* This paradox comes in two forms, progressive and regressive. According to the first, Achilles cannot get to the end of any racecourse, tortoise or no tortoise; indeed, he cannot even reach the original starting point T_0 of the tortoise in the previous paradox. Zeno argues as follows. Before the runner can cover the whole distance he must cover the first half of it (see figure 4).

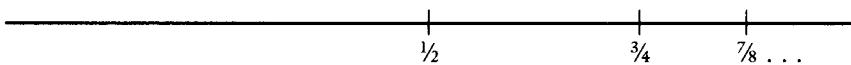


Figure 4

Then he must cover the first half of the remaining distance, and so on. In other words, he must first run one-half, then an additional one-fourth, then an additional one-eighth, etc., always remaining somewhere short of his goal. Hence, Zeno concludes, he can never reach it. This is the progressive form of the paradox, and it has very nearly the same force as Achilles and the Tortoise, the only difference being that in the Dichotomy the goal is stationary, while in Achilles and the Tortoise it moves, but at a speed much less than that of Achilles.

The regressive form of the Dichotomy attempts to show, worse yet, that the runner cannot even get started. Before he can complete the full

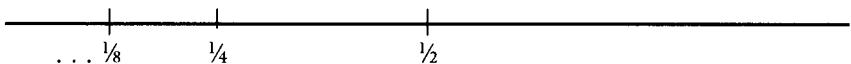


Figure 5

distance, he must run half of it (see figure 5). But before he can complete the first half, he must run half of that, namely, the first quarter. Before he can complete the first quarter, he must run the first eighth. And so on. In order to cover any distance no matter how short, Zeno concludes, the runner must already have completed an infinite number of runs. Since the sequence of runs he must already have completed has the form of a regression,

$$\dots \frac{1}{16}, \frac{1}{8}, \frac{1}{4}, \frac{1}{2},$$

it has no first member, and hence, the runner cannot even get started.

- (3) *The Arrow.* In this paradox, Zeno argues that an arrow in flight is always at rest. At any given instant, he claims, the arrow is where it is, occupying a portion of space equal to itself. During the instant it cannot move, for that would require the instant to have parts, and an instant is by definition a minimal and indivisible-element of time. If the arrow did move during the instant it would have to be in one place at one part of the instant, and in a different place at another part of the instant. Moreover, for the arrow to move during the instant would require that during the instant it must occupy a space larger than itself, for otherwise it has no room to move. As Russell says, "It is never moving, but in some miraculous way the change of position has to occur *between* the instants, that is to say, not at any time whatever."² This paradox is more difficult to understand than Achilles and the Tortoise or either form of the Dichotomy, but another remark by Russell is apt: "The more the difficulty is meditated, the more real it becomes."
- (4) *The Stadium.* Consider three rows of objects A, B, and C, arranged as in the first position of figure 6. Then, while row A remains at rest, imagine rows B and C moving in opposite directions until all three rows are lined up as shown in the second position. In the process, C₁ passes twice as many B's as A's; it lines up with the first A to its left, but with the second B to its left. According to Aristotle, Zeno concluded that "double the time is equal to half."

<i>First Position</i>			<i>Second Position</i>		
A ₁	A ₂	A ₃	A ₁	A ₂	A ₃
B ₁	B ₂	B ₃	B ₁	B ₂	B ₃
C ₁	C ₂	C ₃	C ₁	C ₂	C ₃

Figure 6

Some such conclusion would be warranted if we assume that the time it takes for a C to pass to the next B is the same as the time it takes to

pass to the next A, but this assumption seems patently false. It appears that Zeno had no appreciation of relative speed, assuming that the speed of C relative to B is the same as the speed of C relative to A. If that were the only foundation for the paradox we would have no reason to be interested in it, except perhaps as a historical curiosity. It turns out, however, that there is an interpretation of this paradox which gives it serious import.

Suppose, as people occasionally do, that space and time are atomistic in character, being composed of space-atoms and time-atoms of non-zero size, rather than being composed of points and instants whose size is zero.³ Under these circumstances, motion would consist in taking up different discrete locations at different discrete instants. Now, if we suppose that the As are not moving, but the Bs move to the right at the rate of one place per instant while the Cs move to the left at the same speed, some of the Cs get past some of the Bs without ever passing them. C₁ begins at the right of B₂ and it ends up at the left of B₂, but there is no instance at which it lines up with B₂; consequently, there is no time at which they pass each other – it never happens.

It has been suggested that Zeno's arguments fit into an overall pattern.⁴ Achilles and the Tortoise and the Dichotomy are designed to refute the doctrine that space and time are continuous, while the Arrow and the Stadium are intended to refute the view that space and time have an atomic structure. The paradox of plurality [not discussed here], also fits into the total schema. Thus, it has been argued, Zeno tries to cut off all possible avenues to escape from the conclusion that space, time, and motion are not real but illusory.

It is extremely tempting to suppose, at first glance, that the first three of these paradoxes at least arise from understandable confusions on Zeno's part about concepts of the infinitesimal calculus. It was in this spirit that the American philosopher C. S. Peirce, writing early in the twentieth century, said of Achilles that "this ridiculous little catch presents no difficulty at all to a mind adequately trained in mathematics and logic."⁵ There is no reason to think he regarded any of Zeno's other paradoxes more highly.

We should begin by noting that, although the calculus was developed in the seventeenth century, its foundations were beset with very serious logical difficulties until the nineteenth century – when Cauchy clarified such fundamental concepts as functions, limits, convergence of sequences and series, the derivative, and the integral; and when his successors Dedekind, Weierstrass, et al., provided a satisfactory analysis of the real number system and its connections with the calculus. I am firmly convinced that Zeno's various paradoxes constituted insuperable difficulties for the calculus in its pre-nineteenth-century form, but that the nineteenth-century achievements regarding the foundations of the calculus provide means which go far toward the resolution of Zeno's paradoxes. Let us see what light these purified concepts can throw on the paradoxes of motion.⁶

The Sum of an Infinite Series

It is hard to guess how deep or subtle Zeno's actual reasoning was; experts differ on the point.⁷ It may have been that Zeno's original version of Achilles and the Tortoise involved the following sort of argument: since Achilles must traverse an infinite number of distances, each greater than zero, in order to catch up with the tortoise, he can never do so, for such a process would take an infinite amount of time. Against this form of the argument Aristotle quite appropriately pointed out that the time span during which Achilles chases after the tortoise can likewise be subdivided into infinitely many non-zero intervals, so Achilles has infinitely many non-zero time intervals in which to traverse the infinitely many non-zero space intervals. But this response can hardly be adequate, for the question still remains: how can infinitely many positive intervals of time OR space add up to anything less than infinity? The answer to this question was not provided until Cauchy offered a satisfactory treatment of convergent series in the first half of the nineteenth century.

The first concept we need is the *limit* of an infinite sequence. An infinite sequence is simply an ordered set of terms $\{S_n\}$ which correspond in a one-to-one fashion with the positive integers – each term of the sequence being coordinated by the subscript n to a positive integer. The sequence is said to be *convergent* if it has a limit. To say that such a sequence has a limit means that there is some number L (the limit) such that the terms of the sequence become and remain arbitrarily close to that value as we run through the successive terms. More precisely, for any number ϵ greater than 0, there is some positive integer N such that for every term S_n with $n > N$, the difference between S_n and L is less than ϵ . In the sequence

$$\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \dots, \frac{1}{2n}, \dots$$

the limit is 0, since the difference between the terms of the sequence and 0 is arbitrarily small for sufficiently large values of n . If, for example, we choose $\epsilon = \frac{1}{10}$, by the time we reach the fourth term $S_4 = \frac{1}{16}$, the difference between that term and $L (= 0)$ is less than $\frac{1}{10}$, and the difference remains less than $\frac{1}{10}$ for every subsequent member of the sequence. For $\epsilon = \frac{1}{100}$, $|S_n - 0|$ is less than ϵ for $n = 7$, and the difference remains less than $\frac{1}{100}$ for every subsequent term. Similarly, ϵ may be chosen as small as we like, say $\frac{1}{1,000,000}$ or $\frac{1}{1,000,000,000}$, provided it is greater than zero, and there is some point in this sequence beyond which all remaining terms differ from L by less than ϵ . It is easy to show, by completely parallel reasoning, that the sequence

$$\frac{1}{2}, \frac{3}{4}, \frac{7}{8}, \dots, 1 - \frac{1}{2n}, \dots$$

converges to the limiting value of 1.

After the concept of the limit of a sequence has been defined, it can be used to define the sum of an infinite *series*. An infinite series is simply an infinite sequence of terms which are related to one another by addition; for example,

$$\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \dots + \frac{1}{2^n}, + \dots$$

Such a sum is not defined in elementary arithmetic, for ordinary addition is restricted to sums of finite numbers of terms, but this operation can be extended very naturally to an infinite series. In order to define the sum of an infinite *series*

$$s_1 + s_2 + s_3 + \dots$$

we form the *sequence* of partial sums,

$$\begin{aligned} S_1 &= s_1 \\ S_2 &= s_1 + s_2 \\ S_3 &= s_1 + s_2 + s_3 \\ \text{etc.} & \end{aligned}$$

Each of these partial sums is a sum with a finite number of terms, and it involves only the familiar operation of addition from elementary arithmetic. We have already defined the limit of an infinite sequence. If the *sequence* of partial sums,

$$S_1, S_2, S_3, \dots$$

has a limit, we say that the infinite *series*

$$s_1 + s_2 + s_3 + \dots$$

is convergent, and we define its sum as the limit of the sequence of partial sums. This amounts to saying, intuitively, that the sum of a convergent infinite series is a number that can be approximated arbitrarily closely by adding up a sufficient (finite!) number of terms. Given this definition of the sum of an infinite series, it becomes perfectly meaningful to say that the infinitely many terms of a convergent series have a finite sum.

Both the first form of the Dichotomy and the Achilles paradoxes present us with infinite series to be summed. In the Dichotomy, for instance, it is shown that the runner, to cover a racecourse that is one mile in length, must cover the following series of non-overlapping distances:

$$\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \dots$$

Each term of this series is greater than zero. We form the sequence of partial sums

$$\begin{aligned} S_1 &= \frac{1}{2} \\ S_2 &= \frac{1}{2} + \frac{1}{4} = \frac{3}{4} \\ S_3 &= \frac{1}{2} + \frac{1}{4} + \frac{1}{8} = \frac{7}{8} \\ \text{etc.} & \end{aligned}$$

As we noted above, this sequence converges to the limit 1; that is the sum of this convergent infinite series. Achilles and the Tortoise is quite analogous. If Achilles can run twice as fast as the tortoise, and the tortoise has a head start of one-half of the course, the infinite series generated by Achilles running to each subsequent starting point of the tortoise is precisely the one we have just summed. To whatever extent these paradoxes raised problems about the intelligibility of adding up infinitely many positive terms, the nineteenth-century theory of convergent sequences and series resolved the problem.

Instantaneous Velocity

An initial reaction to the paradox of the Arrow might be the suspicion that it hinges on a confusion between the concepts of instantaneous motion and instantaneous rest. Perhaps Zeno did feel that the only way for an arrow to be at a particular place was to be at rest – that the notion of instantaneous non-zero velocity was illegitimate. If Zeno argued – we have no way of knowing whether he did or not – that at every moment of its flight the arrow is at some place in its trajectory, and hence at every moment of its flight it has velocity zero, then he would have been correct in concluding that its velocity during the whole course of its flight would be zero, rendering the arrow motionless. Nineteenth-century mathematics showed, however, that one of these assumptions is incorrect. It is entirely intelligible to attribute non-zero instantaneous velocities to moving objects when an instantaneous velocity is understood as a derivative – namely, the rate of change of position with respect to time. This derivative is defined as the limit of the average velocity during decreasing non-zero intervals of time. Suppose, for example, that the arrow flies at a uniform speed. We find that in one second it covers ten feet, in one-tenth of a second it covers one foot, in one-hundredth of a second it covers one-tenth of a foot, and so on. As we take these *average* velocities over decreasing finite time intervals which converge to an instant t_1 , the average velocities approach a limit of ten feet per second, and this is, by definition, the instantaneous velocity of the arrow at t_1 . The same can be said for every moment during its flight; it travels its whole course at ten feet per second, and its velocity at each moment is ten feet per second. If Zeno felt that the only intelligible instantaneous velocity is zero, nineteenth-century mathematics proved him wrong.

The infinitesimal calculus was, of course, developed in the seventeenth century, and it made use of instantaneous velocities. These were, unfortunately, considered to be infinitesimal distances covered in infinitesimal times. It was against such notions that Berkeley leveled his broadside in *The Analyst*,⁸ characterizing infinitesimals as “ghosts of recently departed quantities.” It is possible that Zeno’s Arrow paradox was also directed against just such a conception. If we try to conceive of finite motion over a finite distance during a finite time as being composed of a large number of motions over infinitesimal distances during infinitesimal times, enormous confusion is likely to ensue. How much space does an arrow occupy during an infinitesimal time? Is it just as large as the arrow, or is it a wee bit larger? If it is larger, then how does the arrow get from

one part of that space to another? And if not, then how can the arrow be moving at all? And how long is an infinitesimal time span? Does it have parts or not? If so, how can we characterize motion during its parts? If not, how can motion occur during this infinitesimal time? These are questions that Zeno and his fellow Greeks could not answer, and to which modern calculus prior to Cauchy had no satisfactory answer either. This is why I remarked earlier that nineteenth-century – not seventeenth-century – mathematics held an important key, in the concept of the derivative, to the resolution of Zeno's Arrow paradox.

Mathematical Functions

There is, however, still an underlying problem about instantaneous velocity. We have seen how such a concept can be defined intelligibly, but this definition makes essential reference to what is happening at neighboring instants. Instantaneous velocity is defined as a limit of a sequence of average velocities over finite time intervals; without some information about what happens in these intervals we can say nothing about the instantaneous velocity. If we know simply that the center of the arrow was at the point s_1 at time t_1 we can draw no conclusion whatever about its velocity at that instant. Unless we know what the arrow was doing at other times close to t_1 we *cannot* distinguish instantaneous motion from instantaneous rest. It was just this consideration, I believe, which led the philosopher Henri Bergson to say that Zeno's Arrow paradox calls attention to the absurd proposition "... that movement is made of immobilities."⁹ Bergson concluded that the Arrow paradox proves that the standard mathematical characterization of motion must be wrong. We must look at this argument a little more closely.

In modern physics, motion is treated as a functional relationship between points of space and instants of time. The formula for the motion of a freely falling body, for example, is

$$x = f(t) = \frac{1}{2}gt^2.$$

Such formulas make it possible, by employing the function f , to compute the position x given a value of time t . But to understand this treatment of motion fully, it is necessary to have a clear conception of mathematical functions. Before the nineteenth century there was no satisfactory treatment of functions; functions were widely regarded as things which moved or flowed. Such a conception is of no help in attempting to resolve Zeno's paradoxes; on the contrary, Zeno's paradoxes of motion constitute severe difficulties for any such notion of mathematical functions. The situation was dramatically improved when Cauchy defined a function as simply a pairing of numbers from one set with numbers from another set. The numbers of the first set are the *values of the argument*, sometimes called the *independent variable*; the numbers of the second set (which need not be a different set) are the *values of the function*, sometimes called the *dependent variable*. For example, the function $F(x) = y = x^2$ pairs real numbers with non-negative real numbers. With the number 2 it associates

the number 4, with the number -1 it associates the number 1, with the number $\frac{1}{2}$ it associates the number $\frac{1}{4}$, and so forth. Now according to Cauchy, the mathematical function F simply is the set of all such pairs of numbers [namely, that shown in Table 1].

Table 1

x	$F(x) = x^2$
1	1
2	4
3	9
$\frac{1}{2}$	$\frac{1}{4}$
$\frac{1}{3}$	$\frac{1}{9}$
-2	4
-1	1
etc.	etc.

Similarly, the function f used to describe the motion of a falling body is nothing more or less than a pairing of the values of the position variable x with values of the time variable t . At $t = 0$, $x = 0$; at $t = 1$, $x = 16$; at $t = 2$, $x = 64$. This is how we say, in mathematical language, that a body starting from rest in the vicinity of the surface of the earth and falling freely travels 16 feet in the first second, 48 feet in the next second, and so on.

Let us now apply this conception of a mathematical function to the motion of an arrow; to keep the arithmetic simple, let it travel at the uniform speed of ten feet per second in a straight line, starting from $x = 0$ at $t = 0$. At any subsequent time t , its position $x = 10t$. Accordingly, part of what we mean by saying that the arrow moved from point A ($x = 10$) to point B ($x = 30$) is simply that it was *at A* when $t = 1$, and it was *at B* when $t = 3$. When we ask how it got from A to B , the answer is that it occupied each of the intervening points x ($10 < x < 30$) at suitable times t ($1 < t < 3$) — that is, satisfying the equation $x = 10t$. For example, when $t = 2$, the arrow was at the point C ($x = 20$). When we ask how it got from A to C , the answer is again: by occupying the intervening positions at suitable times. Notice that this answer is *not*: by zipping through the intervening points at ten feet per second. The requirement is that the arrow be *at* the appropriate point *at* the appropriate time — nothing is said about the instantaneous velocity of the arrow as it occupies each of these points. This approach has been appropriately dubbed “the at-at theory of motion.” Once the motion has been described by a mathematical function that associates positions with times, it is then possible to differentiate the function and find its derivative, which in turn provides the instantaneous velocities for each moment of travel. But the motion itself is described by the pairing of positions with times alone. Thus, Russell was led to remark, “Weierstrass, by strictly banishing all infinitesimals, has at last shown that we live in an unchanging world, and that the arrow, at every moment of its flight, is truly at rest. The only point where

Zeno probably erred was in inferring (if he did infer) that, because there is no change, therefore the world must be in the same state at one time as at another. This consequence by no means follows. . . ”¹⁰

What Russell is saying is basically sound, although he does perhaps phrase it overdramatically. It is not that the arrow is “truly at rest” during its flight; rather, the motion consists in being *at* a particular point *at* a particular time, and regarding each individual position at each particular moment, there is no distinction between being at rest at the point and being in motion at the point. The distinction between rest and motion arises only when we consider the positions of the body at a number of different moments. This means that, aside from *being at* the appropriate places at the appropriate times, there is no *additional* process of *moving* from one to another. In this sense, there is no absurdity at all in supposing motion to be composed of immobilities.¹¹

Although this way of viewing motion is, I believe, logically impeccable, it may be psychologically difficult to accept. Perhaps the problem can best be seen in connection with the regressive form of the Dichotomy paradox. Here we have Achilles at the starting point at the very moment at which the race begins. What, we ask, must he do first? Well, someone might say, first he has to run to the starting point of the tortoise. But that answer cannot be correct, for before he can do that, he must run to a point halfway between his and the tortoise’s respective starting points. Before he can do that, however, he must get to a point halfway to the halfway point. And so on. We are off on the infinite regress. It seems that there is no first thing for him to do; whatever we suppose his first task to be, there is another that must be completed before he can finish it. There is, in other words, no first interval for him to cross. This conclusion is true. But it does *not* follow that Achilles cannot get started.

Consider the arrow once more. Suppose it is at point *C* midway in its flight path. When we ask how it gets from *C* to *B* we may be wondering, consciously or unconsciously, where it goes next – how it gets to the next point. But this question is surely illegitimate, for we are thinking of the arrow’s path as a continuous one. Since the points in a continuum are densely ordered, there is no next point. Between any two distinct points there is another (and, hence, infinitely many). The question about Achilles, which we just considered in connection with the regressive Dichotomy, may arise from the same psychological source. We may feel that his first act must be to get to the point next to his starting point, but no such point exists. According to the *at-at* theory of motion, this fact is no obstacle to motion. Both space and time are regarded as continuous, and hence, densely ordered. True, there is no next point of space for Achilles to occupy, but also there is no next moment of time in which he must do so. For each moment of time there is a corresponding point, and for each spatial point there is a corresponding moment; nothing more is required.

The psychological compulsion to demand a next point or a next moment may arise from the fact that we do not experience time as a continuum of instants without duration, but rather, as a discrete series of specious presents, each of which lasts perhaps a few milliseconds. Aside from anthropomorphism, however, there is no reason to try to impose the discrete structure of psycho-

logical time upon the mathematical notion of time as a continuum, since the continuous conception has proved itself such an extremely fruitful tool for the description of physical motion.¹²

Limits of Functions

There is one final issue, arising out of the paradoxes of motion, that was significantly clarified by nineteenth-century foundations of mathematics. During the preceding two centuries, while the calculus floated on vague spatial and temporal intuitions, there was considerable controversy about the ability of a function to reach its limit. Some functions seemed to do so; others did not. It was all quite baffling. This puzzle relates directly to Zeno's paradoxes of Achilles and the Tortoise and the progressive form of the Dichotomy. Achilles seems capable of chasing the tortoise right up to the point of overtaking him, but can he reach that limiting point? Likewise, on the track by himself, Achilles seems capable of traversing the various fractional parts of the course right up to the finish line, but can he achieve that limit? Again, the definitions of functions and limits provided in the nineteenth century come nicely to the rescue. A limit is simply a number. A function is simply a pairing of two sets of numbers. If the limit happens to be one of the numbers in the set of values of the function, then the function does assume the limiting value for some value of its argument variable. If not, then the function never assumes the limiting value. No further question about the ability of a function to "reach" its limit can properly arise.

There can be no serious doubt that the aforementioned nineteenth-century mathematical developments went a long way in resolving the problems Zeno raised about space, time, and motion. The only question is whether there are any remaining problems associated with the paradoxes of motion. Beginning about 1950, a number of mathematically sophisticated writers, who were fully aware of the foregoing considerations, felt that an important problem still remained. One of the most articulate was Max Black, who argued that the analysis of Achilles' attempt to catch the tortoise into an infinite sequence of distinct runs introduces a severe logical difficulty.¹³ The problem, specifically, is whether it even makes sense to suppose that anyone has completed an infinite sequence of runs. Black puts the matter forcefully and succinctly when he says that the mathematical operation of summing an infinite series will tell us where and when Achilles will catch the tortoise if he can catch the tortoise at all, but that is a big "if." There is, Black argues, a fundamental difficulty in supposing that he can catch the tortoise, for, he maintains, "the expression, 'infinite series of acts,' is self-contradictory."¹⁴

Black's argument is based upon consideration of a number of imaginary machines that transfer balls from one tray to another.¹⁵ Suppose, for instance, that there are two machines, Hal and Pal, each equipped with a tray in front. When a ball is placed in Hal's tray, he moves it to Pal's tray; when a ball is placed in Pal's tray, he moves it to Hal's tray. They have a sort of friendly rivalry about getting rid of the balls. Suppose, further, that they are programmed in such a

way that each successive transfer of the ball takes a shorter time; in particular, when the ball is first put into either tray, the machine takes $\frac{1}{2}$ minute to move it to the other tray, next time it takes $\frac{1}{4}$ minute, next time $\frac{1}{8}$ minute, and so forth. (Actually, it is more like a frantic compulsion to get rid of the ball; they carry the maxim "It is more blessed to give than to receive" to a ridiculous extreme.) We begin by putting a ball in Hal's tray, and he takes $\frac{1}{2}$ minute to move it to Pal's tray. Pal then takes $\frac{1}{2}$ minute to put it back in Hal's tray, during which time Hal is resting. Then Hal takes $\frac{1}{4}$ minute to transfer it to Pal's tray, while Pal is resting; in the next $\frac{1}{4}$ minute Pal returns it to Hal's tray while Hal rests. As the process goes on, the pace increases until we see just a blur, but at the end of two minutes it is over, and both machines come to rest. The ball has been transferred infinitely many times; in fact, each machine has made infinitely many transfers (and enjoyed infinitely many rest periods) during the two minutes.

Now, we must ask, where is the ball? Is it in Hal's tray? No, it cannot be in Hal's tray, because every time it was put in, Hal removed it. Is it in Pal's tray? No, because every time it was put there Pal removed it. Black concludes that the supposition that this infinite sequence of tasks has been completed leads to an absurdity.

Another hypothetical infinity machine – perhaps the simplest – is the Thomson lamp.¹⁶ This lamp is of a common variety; it has a single push-button switch on its base. If the lamp is off and you push the switch, the lamp turns on; if the lamp is on and you push the switch, the lamp turns off. Now suppose that someone pushes the switch an infinite number of times; he accomplishes this by completing the first thrust in $\frac{1}{2}$ minute, the second in $\frac{1}{4}$ minute, the third in $\frac{1}{8}$ minute, much as the runner in the Dichotomy is supposed to cover the infinite sequence of distances in decreasing times. Consider the final state of the lamp after the infinite sequence of switchings. Is the lamp on or off? It cannot be on, for each time it was on it was switched off. It cannot be off, for each time it was off it was switched on.

The speed of switching demanded is, of course, beyond human capability, but we are concerned with logical possibilities, not "medical" limitations. Moreover, there are mechanical difficulties inherent in the speed required of Hal and Pal as well as Thomson's lamp, but we are not concerned with problems of engineering. Further, there is no use trying to evade the question by saying that the bulb would burn out or the switch would wear out. Even if we could cover such eventualities by technological advances, there remains a logical problem in supposing that an infinite sequence of switching (or ball transfers) has been achieved. The lamp must be both on and off, and also, neither on nor off. This is a thoroughly unsatisfactory state of affairs.

Black and Thomson are *not* maintaining that Achilles cannot overtake the tortoise and finish the race. We all know that he can, and to argue otherwise would be silly. Black is arguing that it is incorrect to *describe* either feat as "completing an infinite sequence of tasks," and Thomson draws a similar moral. They are suggesting that the paradoxes arise because of a misdescription of the situation.

These authors have focused upon a fundamental point. We must begin by

realizing that no definition, by itself, can provide the answer to a *physical* problem. Take the simplest possible case, the familiar definition of arithmetical addition of two terms. We find, *by experience*, that it applies in some situations and not in others. If we have m apples in one basket and n oranges in another, then we will have $m + n$ pieces of fruit if we put them together in the same container. (Popular folklore notwithstanding, we obviously can "add" apples and oranges.) However, as is well known, if we have m quarts of alcohol in one bucket, and n quarts of water in another, we will not have $m + n$ quarts of solution if we put them together in the same container. The situation is simply another instance of the relation between pure and applied mathematics discussed in the preceding chapter [not included here]. We can define various mathematical operations within pure mathematics, but that is no guarantee of their applicability to the physical world. If such operations are to be applied in the description of physical facts we must determine empirically whether a given physical operation is an admissible interpretation of a given mathematical operation. We have just seen that the combining of apples and oranges in fruit baskets is a suitable counterpart of arithmetical addition, while the mixing of alcohol and water is not. A more significant example occurs in Einstein's special theory of relativity, where composition of velocities is seen not to be a physical counterpart of standard vector addition.

The same sort of question arises when we consider applying the (now standard) definition of the sum of an infinite series. Does a given physical situation correspond to a particular mathematical operation, in this case, the operation of summing an infinite series? Black concludes that the running of a race does not correspond to the summing of an infinite series, for the completion of an infinite sequence of tasks is a logical impossibility. Thus, the running of a race cannot correctly be described as completing an infinite sequence of tasks. This conclusion has far-reaching implications for modern science. If it is right, the usual scientific description of the racecourse as an infinitely divisible mathematical continuum is fundamentally incorrect. It may be a useful idealization for some purposes, but Zeno's paradoxes show that the description cannot be literally correct. The inescapable consequence of this view would seem to be that mathematical physics needs a radically different mathematical foundation if it is to deal adequately with physical reality.

Before accepting any such result, we must examine the infinity machines more closely. They do involve difficulties, but Black and Thomson have not identified them accurately. Consider Thomson's lamp. (The same considerations will apply to Black's infinity machines or any of the others.) Thomson has described a physical switching process that occupies one minute. Given that we begin at t_0 with the lamp off, and given that a switching occurs at $t_1 = \frac{1}{2}$, $t_2 = \frac{3}{4}$, and so on, we have a description that tells, for any moment *prior to* the time $T = 1$ (that is, one minute after t_0), whether the lamp is on or off. For $T = 1$, and subsequent times, it tells us nothing. For any time *prior to* T that the lamp is on, there is a subsequent time *prior to* T that the lamp is off, and conversely. But this does not imply that the lamp is both on and off at T ; we can make any supposition we like without logical conflict. We have, in effect, a function defined over a half-open

interval $0 \leq t < 1$, and we are asked to infer its value at $t = 1$. Obviously, there is no definite answer to such a question. If the function approached a limit at $t = 1$, it would be natural to extend the definition of the function by making that limit the value of the function at the end point. But the “switching function” describing Thomson’s lamp has no such limit, so any extension we might choose would seem arbitrary.¹⁷ The same goes for the position of the ball Hal and Pal pass back and forth. In the Dichotomy and the Achilles paradoxes, by contrast, the “motion function” of the runner does approach a limit, and this limit provides a suitably appealing answer to the question about the location of the runner at the conclusion of his sequence of runs.¹⁸

One cannot escape the feeling, however, that there are significant and as yet unmentioned differences between the infinite sequence of runs Achilles must make to catch the tortoise and the infinite sequence of ball transfers executed by Black’s machines (or the infinite sequence of switch pushes required by the Thomson lamp). And there is at least one absolutely crucial difference. Consider the motion of the ball as it is passed back and forth between Hal and Pal. Say that the trays are three inches apart. Then the ball is made to traverse this *fixed* positive distance infinitely many times. In order to do so, it must travel an *infinite* distance in a finite length of time. Now, no one is interested in showing that Achilles can run an infinite distance in a finite amount of time – he is fast, but not that fast. The problem is to show how he can run a *finite* distance that can be subdivided into an infinite number of subintervals.

Achilles can make his run if he can achieve a fixed positive velocity; the ball which travels back and forth over the fixed distance between Hal and Pal must achieve velocities that increase without any bound. This difficulty could, of course, be repaired. Suppose we stipulate that the distances covered by the ball, like the distances Achilles must cover, decrease as the time available for each transit decreases. This can be done by making the trays of Hal and Pal move closer and closer during the two-minute interval, so that they coincide in the middle at the end of the infinite sequence of transfers. But now there is no problem at all about the position of the ball at the end – it is right in the middle in both trays! Similar considerations apply to the Thomson lamp. In order to accomplish a switching, the button must be moved a certain finite distance, say $\frac{1}{8}$ inch. If this is done infinitely many times, the finger which pushes the button and the button itself must traverse an infinite total distance. A necessary, though not sufficient, condition for the convergence of an infinite series is that the terms converge to zero. In order to overcome this difficulty, the switch would have to be modified in some suitable way, in which case an answer can be given to the question regarding the final on-off state of the lamp.¹⁹

In the literature on Zeno’s paradoxes of motion, especially that concerned with the infinity machines, a good deal of emphasis has been placed on the question of whether Achilles can be said to perform an infinite series of *distinct* tasks. When we divide up the racecourse into an infinite series of positive subintervals, it is often claimed, we are artificially breaking up what is properly considered one motion into infinitely many parts which – so the allegation goes – cannot be considered as individual tasks. In order to clarify this question,

Adolf Grünbaum has given Achilles a fictitious twin – a doppelgänger – who runs a parallel racecourse, starting and finishing at the same time as the original Achilles.²⁰

The new Achilles is a jerky runner. He starts out and runs the first half of the course twice as fast as his counterpart, and then stops and waits for him. When the slower one reaches the midpoint, the interloper runs twice as fast to the three-quarter mark, and again waits for the slower to catch up. He repeats the same performance for each of the remaining infinite series of subintervals. Grünbaum calls the original Achilles, who runs smoothly from start to finish, the *legato runner*; his new twin, who starts and stops, is called the *staccato runner*. The important facts about the staccato runner are: (1) He reaches the end of the course at the same time as the legato runner; if the original Achilles can run the course, so can the staccato runner. (2) The staccato runner takes a rest of finite (non-zero) duration between each of his infinite succession of runs; hence, there can be no question that he performs an infinite sequence of *distinct* runs. (3) The staccato runner (while he is running) runs at a fixed velocity which is simply twice that of his legato mate, so he is not involved in the kinds of ever-increasing velocities that were required in the unmodified Black and Thomson devices.

There is just one final feature of the staccato Achilles which might be a source of worry. Although he is not required to achieve indefinitely increasing velocities, he is required to do a lot of sudden stopping and starting, shifting instantaneously from velocity zero to velocity $2v$ (where v is the legato runner's velocity) and back again. This clearly involves infinite accelerations – and infinitely many of them. One could reasonably doubt the possibility of this degree of jerkiness. It turns out, however, that even the discontinuity in velocity is not a necessary feature of the staccato runner. The physicist Richard Friedberg has shown, by means of a complicated mathematical function, how to describe the motion of a more sophisticated (and less jerky!) staccato runner who covers *each* of the infinite sequence of subintervals by starting from rest, accelerating continuously to a maximum finite velocity, decelerating smoothly to rest, and remaining at rest for the required interval between runs. This staccato runner executes a motion conforming to a continuous function; his velocity (first derivative) and acceleration (second derivative) are continuous, as are all of the higher time-derivatives as well. Moreover, the peak velocities that occur in the successively shorter runs also decrease, converging to zero as the length of the run also converges to zero.²¹ It is hard to see what kind of logical (or conceptual) objection can be raised against this kind of motion. But if the sophisticated staccato runner's series of tasks is feasible, so would be the motions of any of the appropriately modified infinity machines. The motion of the ball passed between Hal and Pal, for example, could be described by a combination of two such functions – the first would describe a sequence of motions from left to right with interspersed periods of rest; the second would consist of a similar sequence, but with the motions from right to left. The second set of motions would be executed during the periods of rest granted by the first function, and the first set of motions would occur during the rest periods granted by the second function.

It therefore appears that a suitably designed Hal-Pal pair of infinity machines are logically possible if the *legato* Achilles – the one we all granted from the beginning – can complete his ordinary garden-variety run.

The Discrete vs the Continuous

The infinitesimal calculus has long been – and still is – the basic mathematical tool in the description of physical reality. It employs variables that range over continuous sets of values, and the functions it deals with are continuous. Although the calculus has been completely “arithmetized,” so that its *formal* development does not demand any geometrical concepts, it is still applied to phenomena that occur in physical space. Its applicability to spatial occurrences is achieved through analytic geometry, which begins with a one-to-one correspondence between the points on a line and the set of real numbers. The set of real numbers constitutes a continuum in the strict mathematical sense; consequently, the order-preserving one-to-one correspondence between the real numbers and the points of the geometrical line renders the line a continuum as well. If, moreover, the geometrical line is a correct representation of lines in physical space, then physical space is likewise continuous. Motion is treated, moreover, as a function of a continuous time variable, and the function itself is continuous. The continuity of the motion function is essential, for velocity is regarded as the first derivative of such a function, and acceleration as the second derivative. Functions which are not continuous are not differentiable, and hence they do not even have derivatives. Continuity is buried deep in standard mathematical physics. It is for this reason that we have concerned ourselves at length with the problems continuity gives rise to.²²

A serious objection might be raised, however, to the view that the mathematical continuum provides a precise and literal representation of physical reality. Since physics customarily uses such idealizations as frictionless planes, point-masses, and ideal gases, the argument could go, it might be reasonable to suppose that the mathematical continuum is another idealization that is convenient for some purposes, but does not provide a *completely* accurate description of space, time, and motion. There is, in addition, ample precedent for treating magnitudes that are known to be discrete as if they were continuous. The law of radioactive decay, for example, employs a continuous exponential function even though it is universally acknowledged that the phenomenon it describes involves discrete disintegrations of individual atoms. Where very large finite numbers of entities are involved, the fiction of an infinite collection is often a convenient one which yields good approximations to what actually happens. In electromagnetic theory, for another example, the infinitesimal calculus is used extensively in dealing with charges, even though all the evidence points to the quantization of charges. It has sometimes been suggested that these considerations hold the solution to Zeno’s paradoxes. For instance, the physicist P. W. Bridgman has said, “With regard to the paradoxes of Zeno . . . if I literally thought of a line as consisting of an assemblage of points of zero length

and of an interval of time as the sum of moments without duration, paradox would then present itself."²³

Although I am in complete agreement with the claim that physics uses idealizations to excellent advantage, it does not seem to me that this provides any basis for an answer to Zeno's paradoxes of plurality or motion. The first three paradoxes of motion purport to show a priori that motion, if it occurs, must be discontinuous. Indeed, Zeno's intention, as far as we can tell, seems to have been to prove a priori that motion cannot occur. With the exception of a very few metaphysicians of the stripe of F. H. Bradley, most philosophers would admit that the question of whether anything moves must be answered on the basis of empirical evidence, and that the available evidence seems overwhelmingly to support the affirmative answer. Given that motion is a fact of the physical world, it seems to me a further empirical question whether it is continuous or not. It may be a very difficult and highly theoretical question, but I do not think it can be answered a priori. Other philosophers have disagreed. Alfred North Whitehead believed that Zeno's paradoxes support the view that motion is atomistic in character, while Henri Bergson seemed to hold an a priori commitment to the continuity of motion.²⁴ It seems to me that considerable importance attaches to the analysis of Zeno's paradoxes for just this reason. Space and time may, as some physicists have suggested, be quantized, just as some other parameters, such as charge, are taken to be.²⁵ If this is so, it must be a conclusion of sophisticated physical investigation of the spatio-temporal structure of the atomic and subatomic domains. A priori arguments, such as Zeno's paradoxes, cannot sustain any such conclusion. The fine structure of space-time is a matter for theoretical physics, not for a priori metaphysics, physicists and philosophers alike notwithstanding. The result of our attempts to resolve Zeno's paradoxes of motion is not a proof that space, time, and motion are continuous; the conclusion is rather that for all we can tell a priori it is an open question whether they are continuous or not.

Before we finally leave Zeno's paradoxes, something should be said about the view of space, time, and motion as discrete quantities. The historical evidence suggests that some of Zeno's arguments were directed against this alternative; that is a plausible interpretation of the Stadium paradox at any rate. Zeno seems to have realized that, if space and time both have discrete structure, there is a standard type of motion that must always occur at a fixed velocity. If, for instance, an arrow is to fly from position *A* to position *B* in as nearly continuous a fashion as is possible in discrete space and time, then it must occupy adjacent space atoms at adjoining atoms of time. In other words, the standard velocity would be one atom of space per atom of time. To travel at a lesser speed, the arrow would have to occupy at least some of the space atoms for more than one time atom; to travel at a greater speed, the arrow would have to skip some of the intervening space atoms entirely, never occupying them in the course of the trip. All of this sounds a bit strange, perhaps, but surely not logically contradictory; this is the way the world might be. Moreover, it is possible, as Zeno's original Stadium paradox shows, for two arrows to pass one another traveling in opposite directions without ever being located next to one another. Imagine

two paths, located as close together as possible in our discrete space, between *A* and *B*. Let one arrow travel one of these paths from *A* to *B*, while the other travels the other path from *B* to *A* (see figure 7). Suppose that the arrow traveling the upper track leaves *A* and occupies the first square on the left, while the arrow traveling the lower track leaves *B* at the same (atomic) moment of time, occupying the first square on the right end of his path. Let each arrow move along its track at the rate of one square for each atom of time. At the fourth moment, the upper arrow is just to the left of the lower arrow; at the next moment, the upper arrow is just to the right of the lower arrow. At no moment are they side-by-side – they get past one another, but there is no event which qualifies as the passing (if we mean being located side-by-side traveling in the opposite directions). This is strange perhaps, but again, it is hardly logically impossible.

	1	2	3	4	5	6	7	8	
A	8	7	6	5	4	3	2	1	B

Figure 7

The mathematician Hermann Weyl has, however, posed a basic difficulty for those who would like to quantize space.²⁶ If we think of a two-dimensional space as being made up of a large number of tiles (something like figure 7), we get into immediate trouble over certain geometrical relations. Suppose for example, that we have a right triangle *ABC* in such a space (see figure 8). Consider, first, the tiles drawn with solid lines. If the positions *A*, *B*, and *C* represent the respective corner tiles, then we see that the side *AB* is four units long, the side *AC* is four units long, and the hypotenuse *BC* is also four units long. The

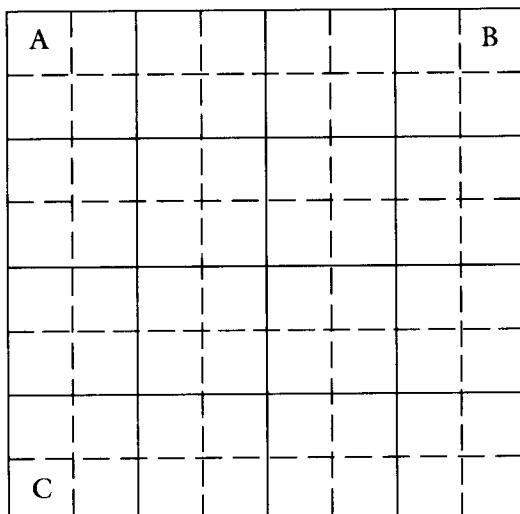


Figure 8

Pythagorean theorem says, however, that the square of the hypotenuse equals the sum of the squares of the other two sides. This means that a right triangle with two legs of four units each should have a hypotenuse about $5\frac{1}{3}$ units long. The Pythagorean theorem is at least approximately true in physical space, as we have found by much experience. The result based upon tile-counting does not begin to approximate the correct result.

This example shows something important about approximations. It is easy to see that discontinuous motion in discrete space and time would be difficult to distinguish from continuous motion if our space and time atoms were small enough. It might be tempting to suppose that our geometrical relations would approach the accustomed ones if we make our tiles small enough. This, unfortunately, is not the case, as you can see by taking the finer grid in Figure 8 given by the broken and solid lines together. Instead of 16 tiles, we now have 64 tiles covering the same region of space. But looking at our triangle ABC once more, we see that all three sides are now 8 units long. No matter how small we make the squares, the hypotenuse remains equal in length to the other two sides. No wonder this is sometimes called the "Weyl tile" argument!²⁷ This is one case in which transition to very small atoms does not help at all to produce the needed approximation to the obvious features of macroscopic space. It shows the danger of assuming that such approximation will automatically occur as we make the divisions smaller and smaller.

It is important to resist any temptation to account for the difficulty by saying that the diagonal distance across a tile is longer than the breadth or height of a tile, and that we must take that difference into account in ascertaining the length of the hypotenuse of the triangle. Such considerations are certainly appropriate if we are thinking of the tiles as subdivisions of a continuous background space possessing the familiar Euclidean characteristics. But the basic idea behind the tiles in the first place was to do away with continuous space and replace it by discrete space. In discrete space, a space atom constitutes one unit, and that is all there is to it. It cannot be regarded as properly having a shape, for we cannot ascribe sizes to parts of it – it has no parts.

Now, I do not mean to argue that there is no consistent way of describing an atomic space or time. It would be as illegitimate to try to prove the continuity of space and time *a priori* as it would be to try to prove their discreteness *a priori*. But, in order to make good on the claim that space and time are genuinely quantized, it would be necessary to provide an adequate geometry based on these concepts. I am not suggesting that this is impossible, but it is no routine mathematical exercise, and I do not know that it has actually been done.²⁸

Notes

- 1 These formulations are taken from Wesley C. Salmon, *Zeno's Paradoxes* (Indianapolis: Bobbs-Merrill 1970), pp. 8–12.
- 2 Bertrand Russell, *Our Knowledge of the External World* (New York: W. W. Norton, 1929), p. 189.
- 3 See J. O. Wisdom, "Achilles on a Physical Racecourse," reprinted in Salmon, *Zeno's Paradoxes*.

- 4 See G. E. L. Owen, "Zeno and the Mathematicians," reprinted in Salmon, *Zeno's Paradoxes*.
- 5 Charles Hartshorne and Paul Weiss, eds, *The Collected Papers of Charles Sanders Peirce* (Cambridge, Mass.: Harvard University Press, 1935), § 6. 177–184.
- 6 For an excellent discussion of these developments see Carl B. Boyer, *The History of the Calculus and its Conceptual Development* (New York: Dover Publications, 1959).
- 7 Zeno's paradoxes pose enormous problems in historical scholarship; for some of the details see Gregory Vlastos, "Zeno of Elea," in the *Encyclopedia of Philosophy*, ed. Paul Edwards (New York: Macmillan and Free Press, 1967).
- 8 Reprinted in James R. Newman, ed., *The World of Mathematics* (New York: Simon and Schuster, 1956).
- 9 Henri Bergson, *Creative Evolution*, trans. Arthur Mitchell (New York: Holt, Rinehart and Winston, 1911), relevant passages reprinted in Salmon, *Zeno's Paradoxes*, quotation from p. 63.
- 10 Bertrand Russell, *The Principles of Mathematics*, 2nd edn (New York: W. W. Norton, 1943), p. 347.
- 11 The contrary view, that this is indeed an absurdity, is based upon the elementary fallacy of composition. This is the only non-trivial, non-artificial instance of this fallacy I have ever encountered.
- 12 For detailed and enlightening discussions of the relations between "physical time" and "psychological time," see Adolf Grünbaum, "Relativity and the Atomicity of Becoming," *Review of Metaphysics*, iv (1950–1), pp. 143–86.
- 13 Max Black, "Achilles and the Tortoise," *Analysis*, xi (1950–1), pp. 91–101; reprinted in Salmon, *Zeno's Paradoxes*.
- 14 Ibid., p. 72 in Salmon.
- 15 The idea of an infinity machine was first suggested by Hermann Weyl, *Philosophy of Mathematics and Natural Science* (Princeton, N.J.: Princeton University Press, 1949). See Salmon, *Zeno's Paradoxes*, p. 201, for relevant quotation.
- 16 James Thomson, "Tasks and Super-Tasks," *Analysis*, xv (1954–5), pp. 1–13; reprinted in Salmon, *Zeno's Paradoxes*.
- 17 The "switching function" may be defined as follows: let 1 represent the "on-state" of the lamp, and let 0 represent the "off-state." This function has a determinate value for each value of $t < 1$, but it fluctuates infinitely often between 0 and 1 in any neighborhood of $t = 1$; hence, it has no limit at $t = 1$.
- 18 The arguments of this paragraph were given by Paul Benacerraf, "Tasks, Super-Tasks, and the Modern Eleatics," *Journal of Philosophy*, lix (1962), pp. 765–84; reprinted in Salmon, *Zeno's Paradoxes*.
- 19 This analysis of infinity machines and their modifications is due to Adolf Grünbaum, "Modern Science and Zeno's Paradoxes of Motion," Part II, in Salmon, *Zeno's Paradoxes*, pp. 218–44.
- 20 Ibid., Part I, "The Zenonian Runners," pp. 204–18.
- 21 See Salmon, *Zeno's Paradoxes*, pp. 215–16, for the details.
- 22 A continuous function is, intuitively, one that can be plotted by means of a line that has no gaps in it – one that can be drawn without lifting the pencil from the paper on which the function is being plotted. For a respectable mathematical treatment of the concept of continuity, in terms requiring no previous mathematical training beyond high school, see Richard Courant and Herbert Robbins, *What is Mathematics?* (New York: Oxford University Press, 1941), Ch. vi.
- 23 P. W. Bridgman, "Some Implications of Recent Points of View in Physics," *Revue Internationale de Philosophie*, iii (1949), p. 490; quoted by Grünbaum, see Salmon,

- Zeno's Paradoxes, p. 177.
- 24 See Salmon, *Zeno's Paradoxes*, pp. 16–20, for discussion of metaphysical interpretations of these paradoxes; see pp. 59–66 for a famous passage from Bergson.
- 25 See Grünbaum, "Modern Science and Zeno's Paradoxes of Motion," Part III, in Salmon, *Zeno's Paradoxes*, pp. 244–60, for an assessment of the extent to which the quantization of space and time has been accomplished.
- 26 Weyl, *op. cit.*, p. 43. See Salmon, *Zeno's Paradoxes*, p. 175, for the relevant quotation.
- 27 "Weyl" is pronounced like "vile."
- 28 See Peter D. Asquith, *Alternative Mathematics and Their Status*, Ph.D. dissertation, Indiana University, 1970.
-

16 Grasping the Infinite*

José A. Benardete

Once upon a time, long ago, a great controversy broke out among the Gumquats, plunging that ancient, benighted people into a state of confusion perilously close to civil war. A young hero had arisen to challenge some of the most deeply cherished beliefs of the tribe. With all the truculence of youth and ambition, he insisted that, contrary to received opinion, it must be admitted that there are a definite number of leaves in the jungle, a definite number of fish in the ocean, a definite number of stones in the valley. It was a profound mistake, he argued, to suppose that the stones in the valley were really innumerable, uncountable, numberless, indeed so plentiful as to be quite *without number*. The Gumquats at this time were fortunate enough to possess a decimal system of counting, but they rarely had any occasion to count beyond 100. Ancient records were on hand to prove that the highest that anyone had ever counted, in all the recorded history of the tribe, was to the number 488. This number was popularly regarded with almost sacred awe, it was chanted during the holy festivals, and it was held highly unlikely that anyone would ever count beyond it. It seemed to represent the very limit of human achievement.

To the horror of the old, to the delight of the young, our hero gathered the whole tribe together and undertook to break the spell of superstition under which they languished. In full view of all, he proceeded to count up to 200, then on to 300, 400, and as he moved on to 486, 487, 488! a great hush fell upon the tribe, 489! – the young burst forth with cheers, the old clapped their hands upon their ears, refusing to listen to this transgression. Our hero was unable to reach 500. Spears and rocks were being hurled in all directions. It was

* Portions of this paper originally appeared in José A. Benardete, *Infinity: an Essay in Metaphysics* (Oxford: Clarendon Press, 1964). Reprinted by permission of the author.

war. Happily enough, a venerable high priest intervened to compose the passions of the contending factions. He was wise and judicious. "I am prepared to overlook this frivolous trespass," he said charitably. "A mere aberration of youth. But I cannot countenance the dangerous heresy which is being urged upon us. I stand by the faith of our fathers that the leaves in the jungle, the fish in the ocean, the stones in the valley are all really innumerable, uncountable, numberless, indeed so plentiful as to be quite without number. If there be a definite number of stones in the valley, then they must be countable. Do you agree?" The high priest addressed his question to our hero. "Yes, they must be countable," answered our hero with mock humility. "Our fathers," continued the priest, "have bequeathed to us the conviction that the stones in the valley are without number, they have taught us that they are uncountable. It is now being urged that they are really countable after all. Very well, young man: prove it! Let us see you count the stones in the valley. I am afraid," added the priest with a supercilious smile, "you will find this task somewhat more difficult than merely counting beyond 488." Our hero was not unprepared for this challenge. "I am not able to count the stones in the valley," he said simply. This admission threw his followers into a state of wild dismay. They had been so much awed by his earlier performance that they had come to believe that when he insisted that the stones in the valley were really countable, this could only mean that he possessed the power actually to count them. "I am not able to count the stones in the valley," our hero repeated. "But they are countable nonetheless. God is able to count them. Do you deny that God is able to count the stones in the valley?" All eyes were fixed upon the high priest. How would he answer this damaging question? "God is able to do all things," replied the priest unctuously. "God then is able to count the stones in the valley," our hero pressed on. "But then there must certainly exist a definite number of stones in the valley if God is able to count them!" He paused to let this point sink in. "I do not deny that in a very loose and lax manner of speaking the stones in the valley may perhaps be said to be innumerable. Certainly, they cannot be counted by man. But they can be counted by God, and it is God, not man, who is the measure of all things. From God's point of view, there is a definite number of stones in the valley." So beguilingly persuasive were our hero's words that not only did they succeed in restoring the young to confidence but even the old were visibly shaken in their faith. Murmurs rippled through the crowd. Of course, there is a definite number of stones in the valley! They could almost see them all in their mind's eye, God having attached a number to each stone by a little tag.

Surely the ancestral faith was utterly exploded. What was there left for the old priest to say? "Young man," he spoke with surprising calm. "You are doubtless familiar with the Great Rapids to the north. Are these Rapids navigable or unnavigable?" "Everyone knows that they are savagely unnavigable," replied our hero, puzzled by this odd turn in the controversy. "And what of toadstools?" asked the priest. "Are they edible or inedible?" "They are inedible," replied our hero uneasily. "And what of tigers? Are they rideable or unrideable?" "It is impossible to ride a tiger," said the young man testily. "Really?" murmured the priest with evident irony. "God is able to ride the tiger. Tigers must be rideable. God is

free to dine on the toadstool. Toadstools must be edible. God is able to navigate the Great Rapids. The Rapids must be navigable." He paused to allow his young adversary an opportunity to speak, but our hero could only stammer in confusion. "I am not surprised by your hesitation," the old man said. "When we say that the Great Rapids are unnavigable, are we so impious as to deny that God is able to navigate them? Certainly not. We mean simply that they cannot be navigated by man. When we say that toadstools are inedible, are we denying that they are edible for God? Certainly not. We mean merely that they are inedible for man. When we say that tigers are unridable . . . But I need scarcely continue. When we say that the stones in the valley are innumerable and uncountable, are we denying that God is able to number them and count them? How absurd! We mean merely that they are innumerable and uncountable – for man! You do not suppose that it is a loose manner of speaking to say that the Great Rapids are unnavigable?" "No," said our hero weakly. "No more is it a loose manner of speaking to say that the stones in the valley, the fish in the ocean, the leaves in the forest are all innumerable, uncountable, numberless, indeed so plentiful as to be quite without number," boomed the old man. "But I have already counted almost to 500," protested our hero desperately. "If I were to continue counting on and on, I would eventually reach a number equal to the number of stones in the valley." "Of course!" replied the priest with disdain. "If! If! If you were to continue counting on and on . . . ! If you were to succeed in navigating the Great Rapids, then you would prove that they are navigable after all. What good is this 'if'? The 'if' doesn't make the Rapids navigable, nor does your 'if' make the stones in the valley numberable and countable."

The old man grew suddenly gentle. "My boy, you have allowed yourself to be transported by a fit of divine enthusiasm unsuitable to a mere mortal. From God's point of view, the stones in the valley are indeed countable, the Great Rapids are indeed navigable, toadstools are indeed edible, tigers are indeed ridable. All things are possible for God. But God's point of view is suitable only for God. Man is truly the measure of all things, not as they are in themselves (Heaven forbid!), but as they are for man. The great numbers that you envisage exist only in the mind of God; they do not exist for man; they are divine and holy; they are not to be profaned by human presumption. My boy, you must rest content with speaking the language of men: the language of God is his alone. You must speak as our fathers have always spoken. You must say that the Great Rapids are unnavigable, that toadstools are inedible, that tigers are unridable; above all, you must say that the stones in the valley are innumerable, that the fish in the ocean are uncountable, that the leaves in the forest are numberless, and that they are all so plentiful as to be quite without number."

The young man bowed his head in a spirit of abject contrition, and the tribe of Gumquats, restored to their ancestral faith, returned to their dogmatic slumbers.

This fable may not be uninstructive; it is designed, above all, to illuminate the Gumquat concept of the infinite. Although we have neglected to mention that

concept expressly, it will not be difficult to reconstruct on the basis of the evidence presented. The Gumquats believe not only that the stones in the valley are *literally* uncountable and hence *literally* without number, they also believe that they are infinitely many. They are persuaded that they are so plentiful as to be *literally* infinite. Are they mistaken in their conviction? No. When they insist that there are infinitely many stones in the valley, they mean merely that if one (i.e., a human being) were to be so foolish as to attempt to count all the stones in the valley, he would never reach the *end* of his task; he would die in the process. In that sense the stones are infinite, endless. If the literal meaning of a word is admitted to be the non-metaphorical meaning that it bears in common discourse, then it must be confessed that the stones in the valley are not only infinite but *literally* infinite, and not merely in Gumquat but in vernacular English as well. It will have become only too evident that the Gumquats are no alien tribe of savages living remote in the jungle. They live directly in our midst: we are the Gumquats. Their language is our language. Their concept of number is our concept of number – as it is found in *mufti* in common discourse. The only difference is that above and beyond this vulgar concept of number we have grown accustomed to another, rather more elevated concept that supplies the underpinning to our very simplest formal arithmetic. We wish to say that the vulgar (or “proto-”) concept is scarcely more than a vestigial remnant, quite without ontological import. Surely it is a hard fact that there are a definite number of leaves in the jungle. In a very crude sense, they are doubtless uncountable; but how can it be denied that they are countable *in principle*? And yet we do not wish to say that toadstools are in principle edible or the Great Rapids in principle navigable. The Gumquats fail to note any difference in all these cases. What is our justification for adopting a divine perspective in regard to number and only in regard to number, being quite content to preserve our human perspective in regard to the other concepts? Whatever the reason, it is evident that a divine dignity attaches to our standard concept of number which is altogether absent elsewhere.

One must not suppose that the Gumquats are ignorant of the distinction between what is possible in principle and what is merely feasible in practice. That distinction governs all their thought. The Gumquats are persuaded that there are a definite number of leopards in the forest, but no one supposes that it is at all possible, practically speaking, to count them. Slinking about most elusively, the leopard is believed to be almost extinct in those parts. In the event of famine all of the leopards roaming the vast forest might well be driven by their hunger to descend *en masse* upon the Gumquats and, being but few in number (if popular opinion is to be credited in this matter), they would then be readily available for counting. Never has a Gumquat been heard to say that the leopards in the forest are innumerable. Countable they are believed to be – in principle, though it is obvious enough that no one could be expected to count them in the ordinary course of things. How different the leaves of the jungle and the stones of the valley: these are all accessible and hence uncountable even in principle; and though the fish in the ocean are in large measure unavailable, it is not for *that* reason that they are believed to be innumerable. The distinc-

tion between what is possible in principle and what is merely feasible in practice, is but one of many distinctions that effect subtle modulations in the Gumquat language. Sometimes a Gumquat may be heard to say that the wives of the Great King (he has forty-six wives) are innumerable, uncountable, numberless, even infinite. But this is merely hyperbole. What is meant is that it must be difficult for the Great King, as for his subjects, to keep track of them – they are so many. Thus a distinction is drawn between what is infinite in the strict sense, the literally infinite, and what is infinite in a loose sense, the metaphorically infinite. Whereas the leaves of the jungle are supposed to be infinite in the strict sense, *literally* infinite, it is only in a loose, metaphorical sense that the wives of the Great King are ever said to be infinite. It ought to come as no surprise that the Gumquats insist that the leaves of the jungle are literally infinite. If the distinction between finite and infinite, between the countable and the uncountable, is to be recognized in vulgar discourse, if these concepts are to arise at all in *mufti*, it is imperative that they have a *use* in the language: they must be negotiable and cashable in terms of the actual experience of the Gumquats. What possible use could they have for the distinction between finite and infinite, between the countable and the uncountable, as that distinction is enshrined in our standard concept? The standard concept of the infinite simply has no application in the form of life that constitutes vulgar practice. Not that a Gumquat may not be moved to remark that the stars in the heavens are infinite, but he has no occasion to mean anything more by his remark than that they are infinite in the sense that the leaves of the jungle are known to be infinite.

It would be a great mistake to appeal to *logical* possibility in an effort to prove that (the Gumquats notwithstanding) it is a hard fact that there are a definite number of stones in the valley. It may not be humanly possible, but it is surely logically possible, to count them all. Certainly. It is also logically possible for a man to navigate the Great Rapids, to dine on the toadstool, to ride the tiger. It is logically possible for gold to cure measles. But this is to say no more than that the statement “Gold has the power to cure measles” is free of self-contradiction. From this mere logical possibility we are not entitled to conclude that gold has the power to cure measles. So, too, though it is logically possible that the stones of the valley might all be counted, we have no right to suppose that they are really countable after all. We might as well say that toadstools are *really* edible. Logical possibility is no measure of what is really the case.

The vitality of the proto-concept comes through to us most strongly when it is seen to satisfy all of the Peano postulates.¹ Not certainly in the precise sense that Peano intended, for the Peano postulates are designed to express our standard concept of number; but at least *nominally*, these postulates are satisfied by the proto-concept of numbers as well. Consider the five Peano postulates. (1) 0 is a number. (2) The immediate successor of any number is a number. (3) There are no two numbers with the same successor. (4) 0 is not the successor of any number. (5) Every property of 0, which belongs to the successor of every number with this property, belongs to all numbers. All five postulates are accepted by the Gumquats, but they refuse to admit that these five postulates logically (i.e., in terms of their proto-logic) entail an infinite progression –

infinite being here understood in the standard sense. That there are infinitely many numbers, they admit, but only in the proto-sense of infinite: no man can exhaust all the numbers by counting, just as no man can count all the leaves of the jungle.²

Very close to being a Gumquat himself, Wittgenstein could write,

Suppose that children are taught that . . . God created an infinite number of stars. . . . Queer: when one takes something of this sort as a matter of course, as if it were in one's stride, it loses its whole paradoxical aspect. It is as if I were to be told: Don't worry, this series, or movement, goes on without ever stopping. We are as it were excused the labor of thinking of an end. "We won't bother about an end." It might also be said: "for us the series is infinite." "We won't worry about an end to this series; for us it is always beyond our ken."³

No wonder that Kripgenstein (Saul Kripke's version of Wittgenstein) insists that what *we* designate by "+" can as readily be taken to be quaddition as addition since the quum of *n* and *m* is the same as the sum of *n* and *m* when the sum of *n* and *m* belongs to the domain of Gumquat numbers; otherwise (bizarrely enough) *n* + *m* (i.e., *n* quus *m*) = 5. So "*n* + *m*" even for us is really only determinate in the Gumquat domain, with "+" being neutral as between plus and quus.⁴ Quaddition then is not to be distinguished from addition except when it comes to *large* numbers, though we are encouraged to believe that even the simple sentence "5 quus 7 = 12" fails to mean the same thing as "5 plus 7 = 12", since "quus" and "plus" designate different mathematical functions.

Granted that the substratum of our mathematical practice is supplied by Gumquat arithmetic, what justifies us in supposing that beyond that substratum lies anything other than sheer mythology, as I take Kripke's Wittgensteinian skeptic to be in effect insisting? Kripke's skeptic poses his challenge in terms of both constitution and justification. Kripke writes, "An answer to the skeptic must satisfy two conditions. First, it must give an account of what fact it is (about my mental state) that constitutes my meaning plus, not quus. But further, there is a condition that any putative candidate for such a fact must satisfy. It must, in some sense, show how I am justified in supposing that I am engaged in addition rather than quaddition."⁵ Where are we to look for that problematic "fact" of which Kripgenstein despairs?

Because the "quus" paradox is designed by Kripgenstein merely to point up a difficulty about meaning across the board, one paradox may help defuse another. Grass in the past having always been green, philosophers argue as against David Hume that we do have reason to believe that it will be green tomorrow. Not so, says Nelson Goodman's skeptic, who argues that since grass has always been grue in the past we have equal reason to suppose that it will be grue tomorrow, where the word "grue" is defined as applying to something if it is green prior to tomorrow and otherwise blue. So empirical induction from past experience urges us to believe that grass will be grue (i.e., blue) tomorrow. Shifting now to meaning, Kripgenstein argues that the "quus" paradox can be

generalized to apply to any predicate, in particular to “green.” The word “green” applies to any one of these grue things that we have been observing, right? And never to anything observed by us to be not grue. So grue has no less right to be what the noise “green” expresses in our language.⁶

If I do mean grue by “green” then my evident projection of “green” into the future entails my expecting grass to be blue tomorrow. But surely I know myself to be free of any such expectation . . . unless indeed Kripgenstein is simply playing fast and loose with a skepticism that was selectively designed to apply only to meaning. Actually, there is a deep issue here, arising precisely from Wittgenstein’s having convinced philosophers, by means of rich scenarios of almost novelistic detail, that someone’s expecting (or not expecting) the telephone to ring in the next hour consists in a battery of multigrade dispositions to behave in various ways. More generally, it is only against the background of a dispositionalist theory of mind, now widely accepted, that Kripgenstein can have any bite at all; in its absence he can only appear dotty. Quite properly then a peculiarly Kripgensteinian skepticism as to my knowing, on a privileged first-person basis, that I do not expect grass to be blue tomorrow, can be seen to be astringently activated. How something (anything, myself included) is disposed to behave in various counterfactual circumstances (e.g., an elastic rubber ball’s being prone to bounce on rolling off a table) can only be known – so goes the powerful argument – by means of familiar inductive procedures.

The trick now in rebutting Kripke’s skeptic lies in carrying the dispositionalist account one step further. Knowing itself, for Wittgenstein, is also to be construed dispositionally, for example, as regards the belief component in knowledge. So my knowing that in applying “green” to grass I do not expect it to be blue tomorrow consists in one disposition riding piggy-back on another.⁷ Therein lies my justification for insisting that I do not mean grue by “green,” a move all the more attractive to those already impressed by the recent turn to reliabilist accounts of justification in theory of knowledge. No need then to posit an internal scanning device (e.g., introspection) whereby I monitor my belief and desire states.⁸ More important still, no need to posit such a device in regard to what I mean by “green” or “+.”

Granted now that much of the sting has been removed from Kripgenstein’s *general* challenge, the special one posed by large numbers can only be aggravated by my own heavy-duty dispositionalism. Assume (absurdly) that grasp of the concept of addition consists in the disposition to give the right answer to *any* addition problem ranging over all of the standard numbers. How can that mere disposition, albeit error-free, be justified in any concrete case if one may not even know that one is engaged in addition? The point is well taken. More is required. Grasp of the concept of addition requires an ability (to add numbers) whose exercise (in normal conditions anyway) involves knowing what one is doing – where this piggy-backed knowing may also be taken dispositionally, for example, in a readiness to provide an informal proof of one’s answer. This enhanced dispositionality provides justification enough.

The fact that we are prone to make mistakes in adding long strings of numerals provides Kripke with another, independent objection to dispositionalism.

This “error” version Christopher Peacocke undertakes to dispel, and in no merely *ad hoc* fashion through a systematic account of what it is to grasp a concept anyway, though he leaves it very much open whether his theory can accommodate large numbers.⁹ Error aside, does my ability to add extend to astronomically large numbers? Hardly, any more than my ability to ride a bicycle (on a horizontal surface) extends to steeper inclines. Thus my shift from dispositions to abilities, while welcome in its own right, merely serves to exacerbate large numbers *per se*. Nor can I simply rely on my knowing the truth of the following proposition. The equation “ $n + m = r$ ” is true for arbitrary n , m and r if and only if $n + m = r$.¹⁰ What proposition is that supposed to be? That will depend first on whether n , m and r are allowed to range beyond the Gumquat domain. Secondly, it will depend on whether the expression “+” means quus or plus or neither (being neutral between them). How then to convince myself that I do grasp the standard proposition to which the science of mathematics is presumed to be committed? For no book has ever explained how to execute the shift from Gumquat to standard arithmetic. It is not as if one can be expected to use the construction of the (positive and negative) integers out of (sets of ordered pairs of) the standard natural numbers as a model for antecedently constructing the standard natural numbers in their turn out of Gumquat numbers.

Seeking reassurance, I propose to design a scientific experiment which will verify, albeit by indirection, my prediction that “5” would certainly not be my answer to any relevant “+ one” query addressed to me on the counterfactual supposition of my counting well beyond the domain of Gumquat numbers. Posing the issue in its sharpest form, I shall even pretend that any such extra-Gumquat counting by me violates the very laws of nature, which laws I further assume to be robust enough to sustain counterfactuals themselves. If I have somewhat pretentiously mentioned conducting a scientific experiment it is largely the literary imagination that will preside, quite in accord with the Gumquat fable itself, by providing such novelistic detail as might ensure psychological and sociological verisimilitude. Genuine enough, however, the experiment is modeled on one draughted by me a few years ago¹¹ in connection with Glaucon’s hypothesis about human nature in the second book of Plato’s *Republic*. Formulated expressly as a “subjunctive” counterfactual, Glaucon’s hypothesis urges that if any human being could only make himself invisible with the ring of Gyges, thereby becoming free to commit murder with impunity, he would wade through blood in his drive to become tyrant of his country. Assume with me that becoming relevantly “invisible” is also contrary to the laws of nature, thereby converting Glaucon’s hypothesis into another contralegal. No matter. To the rescue comes one of Boccaccio’s Calandrino stories in which the credulous fellow is duped into believing that invisibility is his to assume at will. In order then to verify how someone *would* behave if – contrary to the laws of nature – he could pass invisibly among people committing murder undetected, it is enough to verify how he *will* behave tomorrow after being hoaxed, with weighty “evidence” that exploits the marvels of recent technology, into believing himself (even rationally) to be so endowed. Enhanced dispositionalism can thus call even upon contralegal dispositions as well as the more familiar sort. With our

Calindrino model firmly in hand, it is easy now to verify how anyone (Gumquat or not) would answer the question: “Is the number of leaves in the jungle + 1 equal to 5?” simply by verifying how he will in fact answer it after coming to believe, again through weighty “evidence,” that someone has succeeded in counting them one by one.

Not that we have any incentive to verify (or falsify) the hypothesis that he would of course answer in the negative. The point is just that thanks to the hypothesis *being* verifiable (or falsifiable) it is not to be doubted that there is a higher-order dispositional fact of the matter as to whether one means by “+” quus rather than plus; assuming indeed that Kripke’s second challenge about justification can be met in the present case. For I agree with Kripke that the higher-order dispositional fact specified by me (i.e., being disposed to answer in the negative the question posed) fails all by itself to provide justification for one’s answer. Why suppose that one’s answer is the right one? Perhaps one has committed a computational error. Put quite so baldly as that, the suggestion is readily seen to be absurd. No long string of numerals here regarding which we are liable, for obvious reasons, to go astray. But perhaps someone believes that the number of leaves in the jungle is just 4. So that sort of case will also have to be banished. Granted now that a combination of such banal facts rules out “quus” as being what one means by “+,” we can at least entertain the hope that by continuing in the same vein we might succeed, more positively, in pinning down plus as what “+” means in our language.

Anticipating a favorable outcome, I can even indicate, still more ambitiously, how it can be expected to bear on the infinity machines that Zeno’s negative arguments have unexpectedly inspired in our time, by being restyled in an affirmative mode. Along that line we can suppose with Crispin Wright that “in my sleep, say, a genie granted me appropriately boundless computational powers” to test Zenonically every even number (for being the sum of two prime numbers), all to be completed in one minute, thereby inviting Wright to produce an omega proof of Goldbach’s Conjecture. Even allowing with Adolph Grünbaum the *kinematic* coherence of the scenario, Wright can “see grounds for nothing but scepticism” when it comes to “explaining” how “it could be clear to me on waking that that was what” the genie “had done,” how more particularly one could “discover that one has ceased to have *any* limitations of speed, or accuracy, of computation.”¹² Well, let’s allow the occasional mistake to be made here also, though double checking and even omega checking can remedy such lapses. More to the point, once it is recalled that being “clear” about something, for example, the non-denumerability of the real numbers, is to be understood in richly dispositional terms, Wright’s genie should have no difficulty in endowing him with suitable cognitive capacities to ride piggy-back on his computational powers. As to how one might “discover” in oneself these preternatural capacities and powers, there is no great difficulty in principle. Suppose that on pain of death you are ordered to jump fifty, even a thousand feet. Suppose further that, unbeknownst to you, you have been invested with the strength to fulfill that command. No surprise now, surely, if you succeed in saving your life, though the fashionable view today that insists on explaining

behavior as resulting from a combination of belief and desire (forgetting the need for the relevant capacity) may well be perplexed by the outcome. Although the desire is there in full force, the relevant belief is absent. Yet jump you will.

Notes

- 1 Peano showed that all the mathematics of the natural numbers could be derived from three primitive notions (0, number, successor) and the five postulates discussed in the text. – [Eds]
- 2 In his 1970 paper “Wang’s Paradox,” in his *Truth and other Enigmas* (Cambridge, Mass: Harvard University Press, 1978), Michael Dummett in effect charges the “strict finitism” of Gumquat arithmetic with being infected with a semantic incoherence that is already evident in the vagueness of observational terms like “green” with their liability to the Sorites paradox. Satisfied that such terms are validated by linguistic practice, Crispin Wright in his 1982 “Strict Finitism,” in his *Realism, Meaning and Truth*, 2nd edition (Oxford: Blackwell, 1993), argues that “strict finitism remains the natural outcome” of Dummett’s own deflationary approach to mathematics, by being “powerfully buttressed by the ideas of the later Wittgenstein” (p. 166).
- 3 Ludwig Wittgenstein, *Remarks on the Foundations of Mathematics*, ed. G. H. von Wright et al. (Oxford: Blackwell; 1956), part iv, section 14, p. 141.
- 4 Saul Kripke, *Wittgenstein on Rules and Private Language* (Cambridge, Mass: Harvard University Press, 1982), pp. 7–9. See also p. 27 where quaddition is “redefined.”
- 5 Ibid., p. 11.
- 6 Kripke links “quus” to grue on pp. 58–9.
- 7 A refinement on Wittgenstein was supplied by Gilbert Ryle in his classic chapter “Dispositions and Occurrences,” *The Concept of Mind* (London: Hutchinson, 1949) when he wrote, “‘Know’ is a capacity verb . . . signifying that the person described can bring things off, or get things right” (p. 133). Merely “to believe that the ice is dangerously thin is . . . to be prone to skate warily.” But “to say that he keeps to the edge because he knows that the ice is thin is to . . . give quite a different sort of ‘explanation’” (pp. 134–5).
- 8 This piggy-back effect will be especially relished when viewed in connection with Sydney Shoemaker’s critique of “self-blindness” whereby someone can – absurdly – only learn of his own beliefs and desires by means of the standard third-person gathering of evidence that he exercises in regard to others. See his *The First-Person Perspective and other Essays* (Cambridge, UK: Cambridge University Press, 1996).
- 9 Christopher Peacocke, *A Study in Concepts* (Cambridge, Mass: MIT Press, 1992), p. 137.
- 10 Featured by the dominant school in philosophy of language today (associated above all with Donald Davidson), disquotational propositions of this sort cease to be unproblematic when viewed in terms of Kripkenstein. See Kripke’s discussion of Davidson on pp. 71–2.
- 11 See my “The Ring of Gyges: an Aristotelian Approach to Ethics,” *Proceedings of the Creighton Club*, April 1992 Meeting of the New York State Philosophical Association.
- 12 Crispin Wright, *Realism, Meaning and Truth*, 2nd edition (Oxford: Blackwell, 1993), p. 147.

17 The Paradoxes of Time Travel*

David Lewis

Time travel, I maintain, is possible. The paradoxes of time travel are oddities, not impossibilities. They prove only this much, which few would have doubted: that a possible world where time travel took place would be a most strange world, different in fundamental ways from the world we think is ours.

I shall be concerned here with the sort of time travel that is recounted in science fiction. Not all science fiction writers are clear-headed, to be sure, and inconsistent time travel stories have often been written. But some writers have thought the problems through with great care, and their stories are perfectly consistent.¹

If I can defend the consistency of some science fiction stories of time travel, then I suppose parallel defenses might be given of some controversial physical hypotheses, such as the hypothesis that time is circular or the hypothesis that there are particles that travel faster than light. But I shall not explore these parallels here.

What is time travel? Inevitably, it involves a discrepancy between time and time. Any traveler departs and then arrives at his destination; the time elapsed from departure to arrival (positive, or perhaps zero) is the duration of the journey. But if he is a time traveler, the separation in time between departure and arrival does not equal the duration of his journey. He departs; he travels for an hour, let us say; then he arrives. The time he reaches is not the time one hour after his departure. It is later, if he has traveled toward the future; earlier, if he has traveled toward the past. If he has traveled far toward the past, it is earlier even than his departure. How can it be that the same two events, his departure and his arrival, are separated by two unequal amounts of time?

It is tempting to reply that there must be two independent time dimensions; that for time travel to be possible, time must be not a line but a plane.² Then a pair of events may have two unequal separations if they are separated more in one of the time dimensions than in the other. The lives of common people occupy straight diagonal lines across the plane of time, sloping at a rate of exactly one hour of time₁ per hour of time₂. The life of the time traveler occupies a bent path, of varying slope.

On closer inspection, however, this account seems not to give us time travel as we know it from the stories. When the traveler revisits the days of his childhood, will his playmates be there to meet him? No; he has not reached the part of the plane of time where they are. He is no longer separated from them along

* From *American Philosophical Quarterly*, 13 (1976), pp. 145–52. Reprinted by permission of the author and *American Philosophical Quarterly*.

one of the two dimensions of time, but he is still separated from them along the other. I do not say that two-dimensional time is impossible, or that there is no way to square it with the usual conception of what time travel would be like. Nevertheless I shall say no more about two-dimensional time. Let us set it aside, and see how time travel is possible even in one-dimensional time.

The world – the time traveler’s world, or ours – is a four-dimensional manifold of events. Time is one dimension of the four, like the spatial dimensions except that the prevailing laws of nature discriminate between time and the others – or rather, perhaps, between various timelike dimensions and various spacelike dimensions. (Time remains one-dimensional, since no two timelike dimensions are orthogonal.) Enduring things are timelike streaks: wholes composed of temporal parts, or *stages*, located at various times and places. Change is qualitative difference between different stages – different temporal parts – of some enduring thing, just as a “change” in scenery from east to west is a qualitative difference between the eastern and western spatial parts of the landscape. If this paper should change your mind about the possibility of time travel, there will be a difference of opinion between two different temporal parts of you, the stage that started reading and the subsequent stage that finishes.

If change is qualitative difference between temporal parts of something, then what doesn’t have temporal parts can’t change. For instance, numbers can’t change; nor can the events of any moment of time, since they cannot be subdivided into dissimilar temporal parts. (We have set aside the case of two-dimensional time, and hence the possibility that an event might be momentary along one time dimension but divisible along the other.) It is essential to distinguish change from “Cambridge change,” which can befall anything. Even a number can “change” from being to not being the rate of exchange between pounds and dollars. Even a momentary event can “change” from being a year ago to being a year and a day ago, or from being forgotten to being remembered. But these are not genuine changes. Not just any old reversal in truth value of a time-sensitive sentence about something makes a change in the thing itself.

A time traveler, like anyone else, is a streak through the manifold of space-time, a whole composed of stages located at various times and places. But he is not a streak like other streaks. If he travels toward the past he is a zig-zag streak, doubling back on himself. If he travels toward the future, he is a stretched-out streak. And if he travels either way instantaneously, so that there are no intermediate stages between the stage that departs and the stage that arrives and his journey has zero duration, then he is a broken streak.

I asked how it could be that the same two events were separated by two unequal amounts of time, and I set aside the reply that time might have two independent dimensions. Instead I reply by distinguishing time itself, *external time* as I shall also call it, from the *personal time* of a particular time traveler: roughly, that which is measured by his wristwatch. His journey takes an hour of his 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, if he travels toward the future; or the arrival is before the departure in external time (or less than an hour after), if he travels toward the past.

That is only rough. I do not wish to define personal time operationally, making wristwatches infallible by definition. That which is measured by my own wristwatch often disagrees with external time, yet I am no time traveler; what my misregulated wristwatch measures is neither time itself nor my personal time. Instead of an operational definition, we need a functional definition of personal time: it is that which occupies a certain role in the pattern of events that comprise the time traveler's life. If you take the stages of a common person, they manifest certain regularities with respect to external time. Properties change continuously as you go along, for the most part, and in familiar ways. First come infantile stages. Last come senile ones. Memories accumulate. Food digests. Hair grows. Wristwatch hands move. If you take the stages of a time traveler instead, they do not manifest the common regularities with respect to external time. But there is one way to assign coordinates to the time traveler's stages, and one way only (apart from the arbitrary choice of a zero point), so that the regularities that hold with respect to this assignment match those that commonly hold with respect to external time. With respect to the correct assignment properties change continuously as you go along, for the most part, and in familiar ways. First come infantile stages. Last come senile ones. Memories accumulate. Food digests. Hair grows. Wristwatch hands move. The assignment of coordinates that yields this match is the time traveler's personal time. It isn't really time, but it plays the role in his life that time plays in the life of a common person. It's enough like time so that we can – with due caution – transplant our temporal vocabulary to it in discussing his affairs. We can say without contradiction, as the time traveler prepares to set out, "Soon he will be in the past." We mean that a stage of him is slightly later in his personal time, but much earlier in external time, than the stage of him that is present as we say the sentence.

We may assign locations in the time traveler's personal time not only to his stages themselves but also to the events that go on around him. Soon Caesar will die, long ago; that is, a stage slightly later in the time traveler's personal time than his present stage, but long ago in external time, is simultaneous with Caesar's death. We could even extend the assignment of personal time to events that are not part of the time traveler's life, and not simultaneous with any of his stages. If his funeral in ancient Egypt is separated from his death by three days of external time and his death is separated from his birth by three score years and ten of his personal time, then we may add the two intervals and say that his funeral follows his birth by three score years and ten and three days of *extended personal time*. Likewise a bystander might truly say, three years after the last departure of another famous time traveler, that "he may even now – if I may use the phrase – be wandering on some plesiosaurus-haunted oolitic coral reef, or beside the lonely saline seas of the Triassic Age."³ If the time traveler does wander on an oolitic coral reef three years after his departure in his personal time, then it is no mistake to say with respect to his extended personal time that the wandering is taking place "even now."

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 traveler's

life is like a mountain railway. The place two miles due east of here may also be nine miles down the line, in the west-bound direction. Clearly we are not dealing here with two independent dimensions. Just as distance along the railway is not a fourth spatial dimension, so a time traveler's personal time is not a second dimension of time. How far down the line some place is depends on its location in three-dimensional space, and likewise the location of events in personal time depend on their locations in one-dimensional external time.

Five miles down the line from here is a place where the line goes under a trestle; two miles further is a place where the line goes over a trestle; these places are one and the same. The trestle by which the line crosses over itself has two different locations along the line, five miles down from here and also seven. In the same way, an event in a time traveler's life may have more than one location in his personal time. If he doubles back toward the past, but not too far, he may be able to talk to himself. The conversation involves two of his stages, separated in his personal time but simultaneous in external time. The location of the conversation in personal time should be the location of the stage involved in it. But there are two such stages; to share the locations of both, the conversation must be assigned two different locations in personal time.

The more we extend the assignment of personal time outwards from the time traveler's stages to the surrounding events, the more will such events acquire multiple locations. It may happen also, as we have already seen, that events that are not simultaneous in external time will be assigned the same location in personal time – or rather, that at least one of the locations of one will be the same as at least one of the locations of the other. So extension must not be carried too far, lest the location of events in extended personal time lose its utility as a means of keeping track of their roles in the time traveler's history.

A time traveler who talks to himself, on the telephone perhaps, looks for all the world like two different people talking to each other. It isn't quite right to say that the whole of him is in two places at once, since neither of the two stages involved in the conversation is the whole of him, or even the whole of the part of him that is located at the (external) time of the conversation. What's true is that he, unlike the rest of us, has two different complete stages located at the same time at different places. What reason have I, then, to regard him as one person and not two? What unites his stages, including the simultaneous ones, into a single person? The problem of personal identity is especially acute if he is the sort of time traveler whose journeys are instantaneous, a broken streak consisting of several unconnected segments. Then the natural way to regard him as more than one person is to take each segment as a different person. No one of them is a time traveler, and the peculiarity of the situation comes to this: all but one of these several people vanish into thin air, all but another one appear out of thin air, and there are remarkable resemblances between one at his appearance and another at his vanishing. Why isn't that at least as good a description as the one I gave, on which the several segments are all parts of one time traveler?

I answer that what unites the stages (or segments) of a time traveler is the same sort of mental, or mostly mental, continuity and connectedness that unites anyone else. The only difference is that whereas a common person is connected

and continuous with respect to external time, the time traveler is connected and continuous only with respect to his own personal time. Taking the stages in order, mental (and bodily) change is mostly gradual rather than sudden, and at no point is there sudden change in too many different respects all at once. (We can include position in external time among the respects we keep track of, if we like. It may change discontinuously with respect to personal time if not too much else changes discontinuously along with it.) Moreover, there is not too much change altogether. Plenty of traits and traces last a lifetime. Finally, the connectedness and the continuity are not accidental. They are explicable; and further, they are explained by the fact that the properties of each stage depend causally on those of the stages just before in personal time, the dependence being such as tends to keep things the same.⁴

To see the purpose of my final requirement of causal continuity, let us see how it excludes a case of counterfeit time travel. Fred was created out of thin air, as if in the midst of life; he lived a while, then died. He was created by a demon, and the demon had chosen at random what Fred was to be like at the moment of his creation. Much later someone else, Sam, came to resemble Fred as he was when first created. At the very moment when the resemblance became perfect, the demon destroyed Sam. Fred and Sam together are very much like a single person: a time traveler whose personal time starts at Sam's birth, goes on to Sam's destruction and Fred's creation, and goes on from there to Fred's death. Taken in this order, the stages of Fred-*cum*-Sam have the proper connectedness and continuity. But they lack causal continuity, so Fred-*cum*-Sam is not one person and not a time traveler. Perhaps it was pure coincidence that Fred at his creation and Sam at his destruction were exactly alike; then the connectedness and continuity of Fred-*cum*-Sam across the crucial point are accidental. Perhaps instead the demon remembered what Fred was like, guided Sam toward perfect resemblance, watched his progress, and destroyed him at the right moment. Then the connectedness and continuity of Fred-*cum*-Sam has a causal explanation, but of the wrong sort. Either way, Fred's first stages do not depend causally for their properties on Sam's last stages. So the case of Fred and Sam is rightly disqualified as a case of personal identity and as a case of time travel.

We might expect that when a time traveler visits the past there will be reversals of causation. You may punch his face before he leaves, causing his eye to blacken centuries ago. Indeed, travel into the past necessarily involves reversed causation. For time travel requires personal identity – he who arrives must be the same person who departed. That requires causal continuity, in which causation runs from earlier to later stages in the order of personal time. But the orders of personal and external time disagree at some point, and there we have causation that runs from later to earlier stages in the order of external time. Elsewhere I have given an analysis of causation in terms of chains of counterfactual dependence, and I took care that my analysis would not rule out causal reversal *a priori*.⁵ I think I can argue (but not here) that under my analysis the direction of counterfactual dependence and causation is governed by the direction of other *de facto* asymmetries of time. If so, then reversed causation and

time travel are not excluded altogether, but can occur only where there are local exceptions to these asymmetries. As I said at the outset, the time traveler's world would be a most strange one.

Stranger still, if there are local – but only local – causal reversals, then there may also be causal loops: closed causal chains in which some of the causal links are normal in direction and others are reversed. (Perhaps there must be loops if there is reversal; I am not sure.) Each event on the loop has a causal explanation, being caused by events elsewhere on the loop. That is not to say that the loop as a whole is caused or explicable. It may not be. Its inexplicability is especially remarkable if it is made up of the sort of causal processes that transmit information. Recall the time traveler who talked to himself. He talked to himself about 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 by the causal processes that constitute recording, storage, and retrieval of memory traces. His younger self knew, after the conversation, because his older self had known and the information had been preserved by the causal processes that constitute telling. But where did the information come from in the first place? Why did the whole affair happen? There is simply no answer. The parts of the loop are explicable, the whole of it is not. Strange! But not impossible, and not too different from inexplicabilities we are already inured to. Almost everyone agrees that God, or the Big Bang, or the entire infinite past of the universe, or the decay of a tritium atom, is uncaused and inexplicable. Then if these are possible, why not also the inexplicable causal loops that arise in time travel?

I have committed a circularity in order not to talk about too much at once, and this is a good place to set it right. In explaining personal time, I presupposed that we were entitled to regard certain stages as comprising a single person. Then in explaining what united the stages into a single person, I presupposed that we were given a personal time order for them. The proper way to proceed is to define personhood and personal time simultaneously, as follows. Suppose given a pair of an aggregate of person-stages, regarded as a candidate for personhood, and an assignment of coordinates to those stages, regarded as a candidate for his personal time. If⁶ the stages satisfy the conditions given in my circular explanation with respect to the assignment of coordinates, then both candidates succeed: the stages do comprise a person and the assignment is his personal time.

I have argued so far that what goes on in a time travel story may be a possible pattern of events in four-dimensional space-time with no extra time dimension; that it may be correct to regard the scattered stages of the alleged time traveler as comprising a single person; and that we may legitimately assign to those stages and their surroundings a personal time order that disagrees sometimes with their order in external time. Some might concede all this, but protest that the impossibility of time travel is revealed after all when we ask not what the time traveler *does*, but what he *could do*. Could a time traveler change the past? It seems not: the events of a past moment could no more change than numbers

could. Yet it seems that he would be as able as anyone to do things that would change the past if he did them. If a time traveler visiting the past both could and couldn't do something that would change it, then there cannot possibly be such a time traveler.

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 traveled to 1920, suddenly he realizes that he is not too late after all. He buys a rifle; he spends long hours in target practice; he shadows Grandfather to learn the route of his daily walk to the munitions works; he rents a room along the route; and there he lurks, one winter day in 1921, rifle loaded, hate in his heart, as Grandfather walks closer, closer, . . .

Tim can kill Grandfather. He has what it takes. Conditions are perfect in every way: the best rifle money could buy, Grandfather an easy target only twenty yards away, not a breeze, door securely locked against intruders, Tim a good shot to begin with and now at the peak of training, and so on. What's to stop him? The forces of logic will not stay his hand! No powerful chaperone stands by to defend the past from interference. (To imagine such a chaperone, as some authors do, is a boring evasion, not needed to make Tim's story consistent.) In short, Tim is as much able to kill Grandfather as anyone ever is to kill anyone. Suppose that down the street another sniper, Tom, lurks waiting for another victim, Grandfather's partner. Tom is not a time traveler, but otherwise he is just like Tim: same make of rifle, same murderous intent, same everything. We can even suppose that Tom, like Tim, believes himself to be a time traveler. Someone has gone to a lot of trouble to deceive Tom into thinking so. There's no doubt that Tom can kill his victim; and Tim has everything going for him that Tom does. By any ordinary standards of ability, Tim can kill Grandfather.

Tim cannot kill grandfather. Grandfather lived, so to kill him would be to change the past. But the events of a past moment are not subdivisible into temporal parts and therefore cannot change. Either the events of 1921 timelessly do include Tim's killing of Grandfather, or else they timelessly don't. We may be tempted to speak of the "original" 1921 that lies in Tim's personal past, many years before his birth, in which Grandfather lived; and of the "new" 1921 in which Tim now finds himself waiting in ambush to kill Grandfather. But if we do speak so, we merely confer two names on one thing. The events of 1921 are doubly located in Tim's (extended) personal time, like the trestle on the railway, but the "original" 1921 and the "new" 1921 are one and the same. If Tim did not kill Grandfather in the "original" 1921, then if he does kill Grandfather in the "new" 1921, he must both kill and not kill Grandfather in 1921 – in the one and only 1921, which is both the "new" and the "original" 1921. It is logically impossible that Tim should change the past by killing Grandfather in 1921. So Tim cannot kill Grandfather.

Not that past moments are special; no more can anyone change the present or the future. Present and future momentary events no more have temporal parts than past ones do. You cannot change a present or future event from what

it was originally to what it is after you change it. What you *can* do is to change the present or the future from the unactualized way they would have been without some action of yours to the way they actually are. But that is not an actual change: not a difference between two successive actualities. And Tim can certainly do as much; he changes the past from the unactualized way it would have been without him to the one and only way it actually is. To “change” the past in this way, Tim need not do anything momentous; it is enough just to be there, however unobtrusively.

You know, of course, roughly how the story of Tim must go on if it is to be consistent: he somehow fails. Since Tim didn’t kill Grandfather in the “original” 1921, consistency demands that neither does he kill Grandfather in the “new” 1921. Why not? 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. His failure by no means proves that he was not really able to kill Grandfather. We often try and fail to do what we are able to do. Success at some tasks requires not only ability but also luck, and lack of luck is not a temporary lack of ability. Suppose our other sniper, Tom, fails to kill Grandfather’s partner for the same reason, whatever it is, that Tim fails to kill Grandfather. It does not follow that Tom was unable to. No more does it follow in Tim’s case that he was unable to do what he did not succeed in doing.

We have this seeming contradiction: “*Tim doesn’t, but can, because he has what it takes*” versus “*Tim doesn’t, and can’t, because it’s logically impossible to change the past*.” I reply that there is no contradiction. Both conclusions are true, and for the reasons given. They are compatible because “can” is equivocal.

To say that something can happen means that its happening is compossible with certain facts. *Which* facts? That is determined, but sometimes not determined well enough, by context. An ape can’t speak a human language – say, Finnish – but I can. Facts about the anatomy and operation of the ape’s larynx and nervous system are not compossible with his speaking Finnish. The corresponding facts about my larynx and nervous system are compossible with my speaking Finnish. But don’t take me along to Helsinki as your interpreter: I can’t speak Finnish. My speaking Finnish is compossible with the facts considered so far, but not with further facts about my lack of training. What I can do, relative to one set of facts, I cannot do, relative to another, more inclusive, set. Whenever the context leaves it open which facts are to count as relevant, it is possible to equivocate about whether I can speak Finnish. It is likewise possible to equivocate about whether it is possible for me to speak Finnish, or whether I am able to, or whether I have the ability or capacity or power or potentiality to. Our many words for much the same thing are little help since they do not seem to correspond to different fixed delineations of the relevant facts.

Tim’s killing Grandfather that day in 1921 is compossible with a fairly rich set of facts: the facts about his rifle, his skill and training, the unobstructed line of fire, the locked door and the absence of any chaperone to defend the past, and so on. Indeed it is compossible with all the facts of the sorts we would ordinarily count as relevant in saying what someone can do. It is compossible

with all the facts corresponding to those we deem relevant in Tom's case. Relative to these facts, Tim can kill Grandfather. But his killing Grandfather is not compossible with another, more inclusive set of facts. There is the simple fact that Grandfather was not killed. Also there are various other facts about Grandfather's doings after 1921 and their effects: Grandfather begat Father in 1922 and Father begat Tim in 1949. Relative to these facts, Tim cannot kill Grandfather. He can and he can't, but under different delineations of the relevant facts. You can reasonably choose the narrower delineation, and say that he can; or the wider delineation, and say that he can't. But choose. What you mustn't do is waver, say in the same breath that he both can and can't, and then claim that this contradiction proves that time travel is impossible.

Exactly the same goes for Tom's parallel failure. For Tom to kill Grandfather's partner also is compossible with all facts of the sorts we ordinarily count as relevant, but not compossible with a larger set including, for instance, the fact that the intended victim lived until 1934. In Tom's case we are not puzzled. We say without hesitation that he can do it, because we see at once that the facts that are not compossible with his success are facts about the future of the time in question and therefore not the sort of facts we count as relevant in saying what Tom can do.

In Tim's case it is harder to keep track of which facts are relevant. We are accustomed to exclude facts about the future of the time in question, but to include some facts about its past. Our standards do not apply unequivocally to the crucial facts in this special case: Tim's failure, Grandfather's survival, and his subsequent doings. If we have foremost in mind that they lie in the external future of that moment in 1921 when Tim is almost ready to shoot, then we exclude them just as we exclude the parallel facts in Tom's case. But if we have foremost in mind that they precede that moment in Tim's extended personal time, then we tend to include them. To make the latter be foremost in your mind, I chose to tell Tim's story in the order of his personal time, rather than in the order of external time. The fact of Grandfather's survival until 1957 had already been told before I got to the part of the story about Tim lurking in ambush to kill him in 1921. We must decide, if we can, whether to treat these personally past and externally future facts as if they were straightforwardly past or as if they were straightforwardly future.

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. I am not going to vote Republican next fall. The fatalist argues that, strange to say, I not only won't but can't; for my voting Republican is not compossible with the fact that it was true already in the year 1548 that I was not going to vote Republican 428 years later. My rejoinder is that this is a fact, sure enough; however, it 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. We are unlikely to be fooled by the fatalist's methods of disguise in this case, or other ordinary cases. But in cases of time travel,

precognition, or the like, we're on less familiar ground, so it may take less of a disguise to fool us. Also, new methods of disguise are available, thanks to the device of personal time.

Here's another bit of fatalist trickery. Tim, as he lurks, already knows that he will fail. At least he has the wherewithal to know it if he thinks, he knows it implicitly. For he remembers that Grandfather was alive when he was a boy, he knows that those who are killed are thereafter not alive, he knows (let us suppose) that he is a time traveler who has reached the same 1921 that lies in his personal past, and he ought to understand – as we do – why a time traveler cannot change the past. What is known cannot be false. So his success is not only not compossible with facts that belong to the external future and his personal past, but also is not compossible with the present fact of his knowledge that he will fail. I reply that the fact of his foreknowledge, at the moment while he waits to shoot, is not a fact entirely about that moment. It may be divided into two parts. There is the fact that he then believes (perhaps only implicitly) that he will fail; and there is the further fact that his belief is correct, and correct not at all by accident, and hence qualifies as an item of knowledge. It is only the latter fact that is not compossible with his success, but it is only the former that is entirely about the moment in question. In calling Tim's state at that moment knowledge, not just belief, facts about personally earlier but externally later moments were smuggled into consideration.

I have argued that Tim's case and Tom's are alike, except that in Tim's case we are more tempted than usual – and with reason – to opt for a semi-fatalist mode of speech. But perhaps they differ in another way. In Tom's case, we can expect a perfectly consistent answer to the counterfactual question: what if Tom had killed Grandfather's partner? Tim's case is more difficult. If Tim had killed Grandfather, it seems offhand that contradictions would have been true. The killing both would and wouldn't have occurred. No Grandfather, no Father; no Father, no Tim; no Tim, no killing. And for good measure: no Grandfather, no family fortune; no fortune, no time machine; no time machine, no killing. So the supposition that Tim killed Grandfather seems impossible in more than the semi-fatalistic sense already granted.

If you suppose Tim to kill Grandfather and hold all the rest of his story fixed, of course you get a contradiction. But likewise if you suppose Tom to kill Grandfather's partner and hold the rest of his story fixed – including the part that told of his failure – you get a contradiction. If you make *any* counterfactual supposition and hold all else fixed you get a contradiction. The thing to do is rather to make the counterfactual supposition and hold all else as close to fixed as you consistently can. That procedure will yield perfectly consistent answers to the question: what if Tim had not killed Grandfather? In that case, some of the story I told would not have been true. Perhaps Tim might have been the time-traveling grandson of someone else. Perhaps he might have been the grandson of a man killed in 1921 and miraculously resurrected. Perhaps he might have been not a time traveler at all, but rather someone created out of nothing in 1920 equipped with false memories of a personal past that never was. It is hard to say what is the least revision of Tim's story to make it true that Tim kills

Grandfather, but certainly the contradictory story in which the killing both does and doesn't occur is not the least revision. Hence it is false (according to the unrevised story) that if Tim had killed Grandfather then contradictions would have been true.

What difference would it make if Tim travels in branching time? Suppose that at the possible world of Tim's story the space-time manifold branches; the branches are separated not in time, and not in space, but in some other way. Tim travels not only in time but also from one branch to another. In one branch Tim is absent from the events of 1921; Grandfather lives; Tim is born, grows up, and vanishes in his time machine. The other branch diverges from the first when Tim turns up in 1921; there Tim kills Grandfather and Grandfather leaves no descendants and no fortune; the events of the two branches differ more and more from that time on. Certainly this is a consistent story; it is a story in which Grandfather both is and isn't killed in 1921 (in the different branches); and it is a story in which Tim, by killing Grandfather, succeeds in preventing his own birth (in one of the branches). But it is not a story in which Tim's killing of Grandfather both does occur and doesn't: it simply does, though it is located in one branch and not in the other. And it is not a story in which Tim changes the past. 1921 and later years contain the events of both branches, coexisting somehow without interaction. It remains true at all the personal times of Tim's life, even after the killing, that Grandfather lives in one branch and dies in the other.⁷

Notes

- 1 I have particularly in mind two of the time travel stories of Robert A. Heinlein: "By His Bootstraps," in R. A. Heinlein, *The Menace from Earth* (Hicksville, N.Y., 1959), and "– All You Zombies –," in R. A. Heinlein, *The Unpleasant Profession of Jonathan Hoag* (Hicksville, N.Y., 1959).
- 2 Accounts of time travel in two-dimensional time are found in Jack W. Meiland, "A Two-Dimensional Passage Model of Time for Time Travel," *Philosophical Studies*, vol. 26 (1974), pp. 153–73; and in the initial chapters of Isaac Asimov, *The End of Eternity* (Garden City, N.Y., 1955). Asimov's denouement, however, seems to require some different conception of time travel.
- 3 H. G. Wells, *The Time Machine, An Invention* (London 1895), epilogue. The passage is criticized as contradictory in Donald C. Williams, "The Myth of Passage," *The Journal of Philosophy*, vol. 48 (1951), p. 463.
- 4 I discuss the relation between personal identity and mental connectedness and continuity at greater length in "Survival and Identity," in *The Identities of Persons*, ed. Amélie Rorty (Berkeley and Los Angeles, 1976).
- 5 "Causation," *The Journal of Philosophy*, vol. 70 (1973), pp. 556–67; the analysis relies on the analysis of counterfactuals given in my *Counterfactuals* (Oxford, 1973).
- 6 "Iff" is short for "if and only if". – [Eds]
- 7 The present paper summarizes a series of lectures of the same title, given as the Gavin David Young Lectures in Philosophy at the University of Adelaide in July, 1971. I thank the Australian-American Educational Foundation and the American Council of Learned Societies for research support. I am grateful to many friends for comments on earlier versions of this paper; especially Philip Kitcher, William Newton-Smith, J. J. C. Smart, and Donald Williams.

How do Things Persist through Changes of Parts and Properties?

18 Of Confused Subjects which are Equivalent to Two Subjects: an Excerpt from *The Port-Royal Logic** ---

Antoine Arnauld and Pierre Nicole

It is important, in order to understand better the nature of what is called the subject in propositions, to add here a remark which has been made in more important works than this, but which, since it belongs to logic, may find a place here.

It is, that when two or more things which have some resemblance succeed each other in the same place, and, principally, when there does not appear any obtrusive difference between them, although men may distinguish them in speaking metaphysically, they nevertheless do not distinguish them in their ordinary speech; but, embracing them under a common idea, which does not exhibit the difference, and denotes only what they have in common, they speak of them as if they were the same thing.

Thus, though we change the air every moment, nevertheless we consider the air which surrounds us as being always the same; and we say that from being cold it has become warm, as if it were the same, whereas, often that air which we feel cold is not the same as that which we find warm.

This water, we also say, in speaking of a river, was turbid two days ago, and, behold, now it is clear as crystal; while it is impossible it could be the same water. *In idem flumen* (says Seneca), *bis non descendimus, manet idem fluminis nomen, aqua transmissa est.*¹ [*Epistola*, lviii.]

We consider the bodies of animals, and speak of them, as being always the same, though we are assured, that at the end of a few years there remains no part of the matter which at first composed them; and not only do we speak of them as the same body, without considering what we say, but we do so also when we reflect expressly on the subject. For common language allows us to say, — *The body of this animal was composed ten years ago of certain parts of matter, and now it is composed of parts altogether different.* There appears to be some contradiction in speaking thus; for if the parts were altogether different, then is

* From Thomas S. Baynes, LL.D, trans., *The Port-Royal Logic* (Edinburgh: William Blackwood and Sons, 1851).

it not the same body. This is true; but we speak of it, nevertheless, as the same body. And what renders these propositions true is, that the same term is taken for different subjects in this different application.

Augustus said that he had found the city of Rome of brick, and had left it of marble. In the same way we say of a town, of a mansion, of a church, that it was destroyed at such a time, and rebuilt at such another time. What then is this *Rome*, which was at one time of brick, and at another time of marble? What are these towns, these mansions, and churches, which are destroyed at one time, and rebuilt at another? Is the *Rome* of brick the same as the *Rome* of marble? No; but the mind, nevertheless, forms to itself a certain confused idea of *Rome*, to which it attributes these two qualities – being of brick at one time, and of marble at another. And when it afterwards forms propositions about it, and says, for example, that *Rome*, which was brick before the time of Augustus, was marble when he died, – the word *Rome*, which appears to be only one subject, denotes, nevertheless, two, which are really distinct, but united under the confused idea of *Rome*, which prevents the mind from perceiving the distinction of these subjects.

It is in this way that we have cleared up, in the work² whence we have borrowed this remark, the affected perplexity which the (Calvinist) ministers delight to find in that proposition – *This is my body*, which no one would ever find, following the light of common sense. For, as we should never say that it was a proposition very perplexed, and very difficult to be understood, if we said of a church which had been burned and rebuilt – *This church was built ten years ago, and has been rebuilt in a twelvemonth*; in the same way, we could not reasonably say there was any difficulty in understanding this proposition, – *That which is bread at this moment is my body at this other moment*. It is true that it is not the same *this* in these different moments, as the burned church and the rebuilt church are not really the same church; but the mind conceiving the bread and the body of Jesus Christ under the common idea of a present object, which it expresses by *this*, attributes to that object, which is really twofold, and only a unity of confusion, the being bread at one moment, and the body of Jesus Christ at another, just as, having formed of that church burned and rebuilt, the common idea of a church, it gives to that confused idea two attributes, which cannot belong to the same subject.

Hence it follows that, taken in the sense of the Catholics, there is no difficulty in the proposition, *This is my body*, since it is only an abridgment of this other proposition, which is perfectly clear, – *That which is bread at this moment is my body at this other moment* – and since the mind supplies all that is not expressed. For as we have remarked at the end of the First Part, when we use the demonstrative pronoun *hoc* to denote something which is presented to our senses, the precise idea formed by the pronoun remaining confused, the mind adds thereto the clear and distinct ideas obtained from the-senses, in the form of an incidental proposition. Thus, when Jesus Christ pronounced the word *this*, the minds of the apostles added to it, *which is bread*, and as they conceived that it was bread at that moment, they made also the addition of time, and thus the word *this* formed also this idea, – *This which is bread at this moment*. In the same way,

when Christ said *that it was his body*, they conceived that *this was his body at that moment*. Thus the expression, *This is my body*, formed in them that total proposition, *This which is bread at this moment is my body at this other moment*, and the expression being clear, an abridgment of the proposition which diminishes nothing of the idea, is so also.

And as to the difficulty proposed by the ministers, that the same thing cannot be bread and the body of Jesus Christ, since it belongs equally to the extended proposition – *This which is bread at this moment is my body at this other moment* – and to the abridged proposition – *This is my body*; it is clear that this is no better than a frivolous wrangling, which might be alleged equally against these propositions – *This church was burned at such a time, and rebuilt at such another time*. They must all be disintegrated, through this way of conceiving many separate subjects under a single idea, which occasions the same term to be taken sometimes for one subject and sometimes for another, while the mind does not perceive this transition from one subject to another.

After all, we do not here profess to decide the important question touching the way in which we ought to understand these words – *This is my body* – whether in a figurative or in a literal sense; for it is not enough to show that a proposition *may* be taken in a certain sense, it ought to be further proved that it *must* be so taken. But as there are some ministers who, on the principles of a false logic, obstinately maintain that the words of Jesus Christ cannot bear the catholic sense, it is not out of place to show here, briefly, that the catholic sense has in it nothing but what is clear, reasonable, and conformed to the common language of all mankind.

Notes

- 1 “We do not go into the same river twice; the name of the river remains the same, but the water has passed by.”
 - 2 *Traité de la Perpétuité de la Foi*, by Arnauld and Nicole (Paris, 1672).
-

19 Identity through Time*

Roderick M. Chisholm

According to Bishop Butler, when we say of a physical thing existing at one time that it is identical with or the same as a physical thing existing at some

* Reprinted from *Language, Belief and Metaphysics*, ed. H. Keifer and M. Munitz; by permission of the author and the State University of New York Press; © 1970 by State University of New York, Albany, New York. All rights reserved.

other time ("this is the same ship we traveled on before"), we are likely to be using the expression "same" or "identical" in a "loose and popular sense." But when we say of a person existing at one time that he is identical with or the same as a person existing at some other time ("the ship has the same captain it had before"), we are likely to be using "same" or "identical" in a "strict and philosophical sense."¹ I shall attempt to give an interpretation to these two theses; and I shall suggest that there is at least an element of truth in each.

To illustrate the first of the two theses – that it is likely to be only in a loose and popular sense that we may speak of the identity of a physical thing through time – let us recall the traditional problem of the ship of Theseus, in a somewhat updated version. The ship, when it came to be, was made entirely of wood. One day a wooden plank was replaced by an aluminum one (this is the updating) and the wooden plank was cast off. But we still had the same ship, it was said, since the change was only slight. Somewhat later, another wooden plank was cast off and also replaced by an aluminum one. Still the same ship, of course, since, once again, the change was only slight. The changes continue, but they are always sufficiently slight so that the ship on any given day can be said to be the same as the ship on the day before. Finally, of course, the ship is made entirely of aluminum. Some will feel certain that the aluminum ship is the same ship as the one that was once made entirely of wood. After all, it preserved its identity from one change to the next, and identity is transitive. Consider, however, this possibility, suggested by Thomas Hobbes: ". . . if some man had kept the old planks as they were taken out, and by putting them afterwards together in the same order, had again made a ship of them, this, without doubt, had also been the same numerical ship with that which was at the beginning; and so there would have been two ships numerically the same, which is absurd."² To compound the problem, let us imagine that the captain of the original ship had taken a vow to the effect that if his ship were ever to go down, then he would go down with it. What now, if the two ships collide at sea and he sees them start to sink together? Where does his duty lie – with the aluminum ship or with the reassembled wooden ship?

Putting the problem schematically, we may suppose that on Monday a simple ship, "The USS *South Dakota*," came into being, composed of two principle parts, *A* and *B*. On Tuesday, part *A* is replaced by a new part *C*. (We may imagine that the replacement was accomplished with a minimum of disturbance: as *A* was eased off, *C* was pushed on immediately behind and in such a way that one could not say at any time during the process that there was only half a ship in the harbor.) On Wednesday, there was fission, with *B* going off to the left and annexing itself to *F* as it departed from *C*, and with *C* going off to the right and annexing itself to *J* as it departed from *B*. On Thursday, over at the left, *B* is replaced by *L*, while, over at the right, *C* is replaced by *H*. And now the captain of the original USS *South Dakota* sees *FL* and *JH* in equal distress.

One of his advisers tells him: "The ship on the left is the one that took the maiden voyage on Monday, and the ship on the right, therefore, is not." But another of his advisers tells him: "No, it's just the other way around. The ship

on the right is the one that took the maiden voyage on Monday, and the ship on the left, therefore, is not." Agreeing on the need for philosophical assistance, the two advisers appeal to a metaphysician who instructs them in the following way: "First of all," he says, "we must make a technical distinction between what I shall call an intactly persisting temporal object and what I shall call a nonintactly persisting temporal object. A thing is an intactly persisting temporal object if it exists during a period of time and is such that, at any moment of its existence, it has the same parts it had at any other moment of its existence. We may suppose that *AB*, the object that came into being on Monday and passed away on Tuesday, was such an intactly persisting object. So, too, for *BC*, for *FB*, for *CJ*, for *FL*, and for *JH*. Thus a nonintactly persisting temporal object will be a temporal object that is composed of one set of parts at one time and of another set of parts at another time. If we can say of a ship, that it is composed of *A* and *B* on Monday and composed of *B* and *C* on Tuesday, then a ship is such a nonintactly persisting temporal object."³

Mon	<i>AB</i>
Tue	<i>BC</i>
Wed	<i>FB CJ</i>
Thu	<i>FL JH</i>

Figure 9

Appealing now to our diagram (figure 9), the metaphysician continues: "I assume that the situation you disagree about involves the six intact temporal objects you have labeled. It also involves a number of nonintact temporal objects. Thus (i) there is that total object, having the temporal shape of an upside down Y – *that* object is composed of *AB* on Monday, of *BC* on Tuesday, of *FB* and *CD* on Wednesday, and of *FL* and *JH* on Thursday; (ii) there is that object composed of the stem and the left fork of the Y – that object is composed of *AB* on Monday, of *BC* on Tuesday, of *FB* on Wednesday, and of *FL* on Thursday; and (iii) there is that object composed of the stem and of the right fork of the Y – the object that is composed of *AB* on Monday, of *BC* on Tuesday, of *CJ* on Wednesday, and of *JH* on Thursday. The second and third of these temporal objects thus have certain parts in common, and the first one includes both the second and the third among *its* parts."

"Given such distinctions as these," our metaphysician now concludes, "you can see that there is really nothing for you to dispute about. Just consider the question: Is the ship on the left the one that made the maiden voyage on Monday? If you are asking whether *FL* is identical with *AB*, then the answer is obviously *no*, for *FL* didn't come into being until Thursday and *AB* ceased to exist on Tuesday. On the other hand, if you are asking whether *FL* and *AB* are both parts of our second temporal object, the one composed of the stem and of the left fork of the Y, the answer is clearly *yes*, and *JH* is not a part of that object. And if you are asking whether *JH* and *AB* are both parts of our third temporal object, the one that is composed of the stem and of the right fork of the Y, then the answer, once again, is clearly *yes*; and *FL* is not a part of *that* object. All you

need to do then, is to distinguish these various objects and make sure you know *which* ones you are talking about. Then everything will be clear."

I think we might go along with the metaphysician – up to the very last point. Consider the reaction that his sort of instructions might produce: "You say that everything will be clear. Things were *far* more clear before you entered the picture. We couldn't agree as to which of these two ships was the one that set sail on Monday. But we were clear, at least, that only two ships were involved. Now, with all your intact and nonintact temporal objects, we have *no* idea how many ships there were. We have learned from Webster that a ship is a structure used for transportation in water. Your intact temporal objects satisfy *that* definition; so they yield at least six ships. What of the nonintact temporal objects? Is the one having the temporal shape of the Y a ship? That would make seven. The stem would give us eight, the two forks would bring it up to ten; the stem plus the left fork makes eleven, and the stem plus the right one makes it *twelve*. Conceivably we might countenance the presence of twelve ships in this situation if by so doing we could solve our problem. But you haven't solved the problem. Consider the poor captain. He wants to go down with his ship and he *still* doesn't know which way to go."

Our metaphysician, I suggest, did not succeed in locating the source of the dispute.

Consider the problem as it pertained to the relation between *FL* (the object that came to be, on the left, on Thursday) and *AB* (the object that had ceased to be by Tuesday). It was agreed that Webster's definition of "*x* is a ship" would do. It was also agreed that *FL*, *AB*, and the other intact objects satisfied that definition. The question was whether *FL* constituted the same ship as did *AB*. And the question whether *FL* constituted the same ship as did *AB* must be distinguished from the closely related question whether *FL* was identical with *AB*; for, as Locke saw, at least in principle, "*FL* constitutes the same ship as does *AB*" does not imply "*FL* is identical with *AB*".⁴

Railroad trains may provide a more perspicuous example of the distinction between "*x* constitutes the same so-and-so as does *y*" and "*x* is identical with *y*." Suppose we ask: "Is this the same train we rode on last year?" We are not concerned to know whether the set of objects that makes up today's train is identical with the set of objects that made up the train of a year ago. ("I'm not asking whether we rode on *precisely these same cars* a year ago!") The following three statements tell us three quite different things: (1) This set of cars constitutes a train today and it also constituted a train a year ago; (2) This set of cars constitutes the same train as did that set of cars and that set of cars constituted a train a year ago; (3) This set of cars constitutes the same train that that set of cars constituted a year ago. By going to the dictionary we may find a definition or criterion of "*x* is a train"; but we do not thereby find a definition or criterion of "*x* constitutes the same train as does *y*." A definition of the latter expression would be much more complex and would doubtless say something about roadbeds, schedules, and cities. Possibly, for example, if we can agree that the present aggregate of cars leaves Hoboken at 7:30 P.M. for Chicago via Scranton and the Poconos, we may be willing to concede that this is the same train that

we took a year ago, even if all the cars are different. (We may note, in passing, that in this case applicability of “*x* is the same train as *y*” will presuppose applicability of some such expression as “*x* is the same roadbed as was *y*” or “*x* is the same city as was *y*.”)

“The same ship” would seem to require a kind of continuity that “the same railroad train” does not. That is to say, if this is to be the same ship that that was, then this ship must be *evolved* in some clear-cut way from that. The requisite sense of “evolves” is illustrated by our diagram. Thus *BC* is continuous with *AB* in that they have a part in common; we may say, therefore, that the latter object *BC* “directly evolved” from the earlier object *AB*. Analogously for the relation of *FB* to *BC*, of *FL* to *FB*, of *CJ* to *BC*, of *JH* to *CJ*, and of *FB* to *AB*. And since *FL* directly evolved from something that directly evolved from *AB*, we may say simply that *FL* evolved from *AB*.⁵

What more is needed for this to be the same ship that that was? The best we can do, I believe, is to formulate various additional criteria which are such that, if they are satisfied, then this is the same ship that that was. Let us consider only one such criterion – one involving reference to sameness of sailing schedule. Suppose we know, with respect to each object, that it satisfies Webster’s definition of a ship: each object is a structure that is used for transportation in water. Suppose we also know that everything that evolved from that and into this was also a structure used for transportation in water (none of these things was ever towed on land and used there as a dwelling-place or as a restaurant). Suppose we know, moreover, that they all followed the same sailing schedule (they were used, say, to ferry passengers between Hoboken and lower Manhattan). And suppose we know, finally, that if at any time one of these objects underwent fission at that time and evolved into more than one structure that was used for transportation in water, then only one of those structures kept to the original schedule. If we know all these things, then, I think, we may say with confidence, that this is the same ship as that – or, more accurately, that this constitutes now the same ship that that constituted then.

Hence one possible criterion (as distinguished from a definition) of “*x* constitutes now the same ship that *y* constituted then” would be this: *x* evolved from *y*; everything that evolved from *y* and into *x* was a structure used for transportation in water and followed the same sailing schedule that *y* does; and if at any time more than one such structure evolved at that time from *y*, then only one of them followed the same sailing schedule that *y* does.

If we should be fortunate enough to find that Wednesday’s left-hand object followed the same sailing schedule as did those of Monday and Tuesday, and that Wednesday’s right-hand object took off on a course of its own, then we may conclude that the one on the left, and not the one on the right constitutes the same ship as the one that came to be on Monday.

Reverting to the terminology of our metaphysician, we may say that the situation we have been concerned with involved at least six different intactly persisting objects and at least six different nonintactly persisting objects. Does this mean, then, that the situation involved at least a dozen ships? No, for if we speak in a strict and philosophical sense, we will say that counting ships through

a given period of time is not the same as counting structures that are used for transportation in water during that time; it is, rather, to count sets of objects that constitute the same ship during that time. For example, to say that there is *one* ship is to say that there is one set of things all constituting the same ship. To say that there are two ships is to say that there are two sets of things, all the members of the one set constituting the same ship, all the members of the other set constituting the same ship, and no member of the one set constituting the same ship as any member of the other set. And so on, for any number of ships. If, as we are supposing, the *AB*, *BC*, *FB*, and *FL* of our example all follow the same sailing schedule, then they constitute one ship. *CJ*, we said, took off on its own. Hence if *JH* follows the same sailing schedule as did *CJ*, then the situation will involve at most two ships.

We could put the matter paradoxically, therefore, by saying that counting ships is not the same, merely, as counting objects that happen to *be* ships. But if we speak strictly and philosophically, we may avoid any such appearance of paradox. We may say that ships are “logical constructions.” The things that they are constructed upon are things that satisfy Webster’s definition of the loose and popular sense of “ship” – they are structures used for transportation in water. We will not say, therefore, that *AB*, *BC*, and the other intact structures we discussed *are* ships. We will say, instead, that each of these things constitutes a ship. Given the concept of “*x* constitutes the same ship as does *y*,” we could define “*x* constitutes a ship” by saying “*x* is a member of a set of things all constituting the same ship.” The *USS South Dakota*, therefore, would be a logical construction upon one such set of things. If we continue to speak strictly and philosophically, we will not say of the two different things, *AB* and *FL*, that each of them *is*, on its particular day, the *USS South Dakota*. We will say instead that each of them *constitutes*, on its particular day, the *USS South Dakota*. The statements we ordinarily use to describe the ship (e.g., “It weighs more now than it did then”) will be reducible to statements about the things that constitute it (“the thing that constitutes it now weighs more than the thing that constituted it then”).

We now have an obvious interpretation for the first of the theses I have attributed to Bishop Butler – namely, that it is only in a loose and popular sense and not in a strict and philosophical sense that we may speak of the identity of such things as ships through time. He could be construed as telling us, first, that the expression “*x* constitutes at *t* the same ship that *y* constitutes at *t'*” does *not* imply “*x* is identical with *y*"; and analogously for “constituting the same tree,” “constituting the same carriage,” and so on. Then he could be construed as telling us, second, that if we express the fact that *x* constitutes at one time the same ship that *y* constitutes at another time by saying “*x* is identical with *y*” or “*x* is the same as *y*,” then we are speaking only in a loose and popular sense and not in a strict and philosophical sense. And perhaps he could be construed as telling us, finally, that our criteria for *x* constituting the same ship as *y* are pretty much in our own hands, after all, and that once we have determined that a given *x* and *y* do satisfy our criteria for constituting the same ship, or that they do not, then no possible ground for doubt remains.

But there are points of clarification to be made: . . .

(i) Finding an acceptable definition of “*x* is a ship” is a problem for dictionary makers. Finding an acceptable definition of “*x* constitutes the same ship as does *y*” is more likely to be a problem for jurists. It should be noted that we may be in agreement with respect to the proper interpretation of one of those expressions and in disagreement with respect to the proper interpretation of the other; or we may be rigid with respect to the one and latitudinarian with respect to the other.

Assuming we have agreed upon our interpretation of “*x* is a ship,” consider the latitude that yet remains with respect to the interpretation of “*x* constitutes the same ship as does *y*.” According to the particular criterion of constituting the same ship that was satisfied by our example of the USS *South Dakota*, today’s object and tomorrow’s object “constitute the same ship” provided, among other things, that every object that evolves out of today’s object and tomorrow’s is a ship. And for there to be such evolution, each object, we said, must have some part in common with the object from which it directly evolved. We could say, quoting Hume, that with each step it is “in a manner requisite, that the change of parts be not . . . entire,”⁶ but it is very possible that we will find it convenient to relax these criteria. Thus it may be useful to be able to say, on occasion, that a certain object of last year constitutes the same ship as does a certain object of this year even though one of the objects, into which last year’s object evolved and out of which this year’s object evolved, was itself not a ship. Perhaps the ship was partially dismantled and used for a while as a tool shed or as a restaurant; yet, when it was reconverted, we found it convenient, and pleasing to count the result as the same ship that we had before.⁷ We may even find it convenient to say on occasion that though a certain object of last year constitutes the same ship as does a certain object of this year, there was no evolution as defined – the change of parts at one stage was entire. Switching for the moment from ships to rivers, consider this situation: We swim in the upper Rio Grande in the early spring; the river dries up in the summer; new waters then flow in and we swim there once again in the fall. Surely we will want our criterion of “*x* constitutes the same river as does *y*” to allow us to say that we swam in the same river twice.

(ii) The expression “*x* constitutes the same ship as does *y*,” like “*x* is a ship,” allows for borderline cases. We can readily imagine situations in which the only appropriate answer to the question “Is that a ship?” is “Yes and no” – or, better, situations in which “Yes”, is no better an answer than “No,” and “No” is no better an answer than “Yes.”⁸ A hydrofoil that is also a hovercraft may serve as an example. We can readily imagine situations in which to the question “Is this the same ship as that?”, i.e., “Does this constitute the same ship that that did?”, the only answer is “Yes and no.”

It may well happen that when we encounter such a borderline case, we must have an answer other than “Yes and no.” The captain, as we have seen, may well need a more definite answer, and we may need a definite answer to the ques-

tion, “Is the combination hydrofoil and hovercraft a ship?”, for it may be necessary to decide whether such things are to be subject to the regulations that govern ships or to the regulations that govern aircraft. Similarly, for the question “Does this constitute the same ship that that did?”

When the existence of such a borderline case does thus require us to make a choice between “Yes” and “No” the decision is entirely a pragmatic one, simply a matter of convenience. Which ship is to be called “the Ship of Theseus” – the one that evolved step by step from the original ship, or the one that was assembled from the discarded planks of the original ship? Here we have such a borderline question. The question calls for a convention with respect to the interpretation of “constituting the same ship” (or of “is the same ship as,” in its loose and popular sense). We can have it pretty much as we wish, provided we agree. Which ship should the captain go down with? Here, too, we have a borderline question. Perhaps you and I cannot decide, but the courts, or the ships’ courts, can decide. If the captain has agreed to go down with the USS *South Dakota*, and if the court decides that the aluminum ship and not the wooden one is the one that constitutes the USS *South Dakota*, then down with the aluminum ship he ought to go. Or down with it he ought to go unless the authorities decide subsequently (and in time) to “defeat” the convention they have adopted – for any such convention is defeasible and may be altered or defeated if unexpected circumstances show that it will turn out to be inconvenient. The important thing here is this: The convention of the courts, or of the proper authorities will settle the matter. You and I may object to their decision on the ground that some other decision would have been more convenient. But it would make no sense for us to say: “Well, it just might be, you know, that they are mistaken. It just might be that, unknown to them, the wooden ship and not the aluminum one is the USS *South Dakota*.”

(iii) There is also a philosophical point to make about our treatment of the problems of the Ship of Theseus and the USS *South Dakota*.

We may be certain of at least this much: If there is an individual thing x which is such that, through a certain period of time, everything that is part of x at any given moment of that time is also a part of x at any other moment of that time, then what constitutes x at any moment of that time may be said to be identical in the strict and philosophical sense with what constitutes x at any other moment of that time. In such a case, x would satisfy the concept of *intact persistence* that was introduced above. For it was suggested that an individual thing x could be said to *persist intactly* through a given period of time, provided that, at any subperiod of that time, x has the same parts that x has at any other subperiod of that time. In other words, an individual x persists intactly through a given period of time, provided that, for every z , if z is part of x during any subperiod of that time, then z is part of x during every subperiod of that time.⁹ Thus we may say that if just one part of our ship is removed or replaced at a certain time, then other parts of the ship, unlike the ship itself, persist intactly through that time.

We formulated above one possible criterion for saying, of different objects at different times, that they constitute one and the same ship. This criterion, it

should be noted, presupposes intact persistence, though not intact persistence of the ship; for if the criterion is applicable in the case of a given ship, then, with each step of evolution, some part of the ship remains behind, for the change of parts is "not entire." Hence, with each step, some part persists intactly; some part will be such that it keeps all of *its* parts. But though the evolution of our ship from Monday through Thursday involved intact persistence of some part of the ship at some time during each change that took place, it does not presuppose intact persistence of any part of the ship from Monday through Thursday. We are thus more liberal in our interpretation of "*x* constitutes the same ship as does *y*" than we are, say, in our interpretation of "*x* constitutes the same bar of metal, or the same piece of wood, or the same hunk of clay as does *y*." For we are not likely to say of *x* that it constitutes the same bar of metal, or the same piece of wood, or the same hunk of clay as does *y*, unless we think that, throughout the changes from *x* to *y* most of the parts have persisted intactly. But "*x* constitutes the same body of water as does *y*" need not imply that most of the parts have thus persisted intactly. Indeed it need not even imply that the body of water has undergone the type of evolution we described in the case of the ship. Thus, as we have noted, a body of water *x* may constitute in the spring the same river that a body of water *y* constitutes in the fall, even though the river has dried up in the summer and *y*, therefore, has not evolved in the requisite sense from *x*.¹⁰ We might say what St Thomas said of the river Seine: ". . . the Seine river is not 'this particular river' because of 'this flowing water,' but because of 'this source' and 'this bed,' and hence is always called the same river, although there may be other water flowing down it."¹¹ Suppose, then, we say that the river of the spring is the same river as the river of the fall in virtue of the fact that the river of the spring flows through *the same river bed* as does the river of the fall. What, then, would be our criterion for saying that something *x* in the spring is the same river bed as something *y* in the fall? It might be the fact that the river bed in the fall has evolved in the manner I have attempted to describe from the river bed in the spring. Or it might be that the material that constitutes the river bed in the spring is found between *the same river banks* as is the material that constitutes the river bed in the fall. We might then say that *x* in the spring constitutes the same river bank as does *y* in the fall if, once again, *y* has evolved from *x* in the manner I have described.¹²

In other words, persistence, in the loose and popular sense, through time would seem to presuppose such evolution; and such evolution, in turn, presupposes persistence, in the strict and philosophical sense, through time. For it presupposes what I have called intact persistence. It is not implausible to say, therefore, that if there is anything that persists, in the loose and popular sense, through any given period of time, then there is something (perhaps not the same thing) that persists intactly through some subperiod of that time.

What now of Bishop Butler's second thesis: the thesis according to which, when we say of a *person* existing at one time that he is identical with a person existing at another time, we are likely to be using "identical" in a strict and philosophical sense and not merely in a loose and popular sense?

I have suggested a possible interpretation of the expression “loose and popular sense of the *same*.” Putting the point schematically, we may say that “*x* is the same *F* as *y*” is used in a loose and popular sense if it is used in such a way that it does not imply “*x* is identical with *y*.” (The expression “*x* constitutes the same *F* as does *y*” would thus be less misleading for such a use.) I have also suggested that when “*x* is the same *F* as *y*” is used in this loose and popular sense, then it is possible to imagine conditions under which a question of the form “Is *x* the same *F* as *y*? ” has no definite answer – conditions under which we may say both “Yes” and “No,” for “Yes” will be as good an answer as “No,” and “No” will be as good an answer as “Yes.”

Such an interpretation of the expression “loose and popular sense of *same*” suggests at once a possible interpretation of the expression “strict and philosophical sense of *same*.” For example, we are using the expression “*x* is the same person as *y*” in a strict and philosophical sense if we are using it in such a way that it implies “*x* is identical with *y*.” In this case “*x* is the same person as *y*” will be logically equivalent to “*x* is a person and *x* is identical with *y*.” I wish to suggest that “*x* is the same person as *y*,” where the expression in the place of “*x*” is taken to designate a certain person as existing at one time and where the expression in the place of “*y*” is taken to designate a certain person existing at a different time, does have this strict and philosophical use.

When we use “the same person” in this strict way, then, although cases may well arise in which we have no way of *deciding* whether the person *x* is the same person as the person *y*, nevertheless the question “Is *x* the same person as *y*? ” will *have* an answer and that answer will be either “Yes” or “No.” If we know that *x* is a person and if we also know that *y* is a person, then it is not possible to imagine circumstances under which the question “Is *x* the same person as *y*? ” is a borderline question – a question admitting only of a “Yes and no” answer. . . .

It will be instructive to elaborate upon an example that C. S. Peirce suggests.¹³ Let us assume that you are about to undergo an operation and that you still have a decision to make. The utilities involved are, first, financial – you wish to avoid any needless expense – and, secondly, the avoidance of pain, the avoidance, however, just of *your* pain, for pain that is other than yours, let us assume, is of no concern whatever to you. The doctor proposes two operating procedures – one a very expensive procedure in which you will be subjected to total anaesthesia and no pain will be felt at all, and the other of a rather different sort. The second operation will be very inexpensive indeed; there will be no anaesthesia at all and therefore there will be excruciating pain. But the doctor will give you two drugs: first, a drug just before the operation which will induce complete amnesia, so that while you are on the table you will have no memory whatever of your present life; and secondly, just after the agony is over, a drug that will make you completely forget everything that happened on the table. The question is: Given the utilities involved, namely the avoidance of needless expense and the avoidance of pain that *you* will feel, other pains not mattering, is it reasonable for you to opt for the less expensive operation?

My own belief is that it would *not* be reasonable, even if you could be completely certain that both amnesia injections would be successful. I think that *you*

are the one who would undergo that pain, even though you, Jones, would not know at the time that it is Jones who is undergoing it, and even though you would never remember it. Consider after all, the hypothesis that it would *not* be you. What would be your status, in such a case, during the time of the operation? Would you be waiting in the wings somewhere for the second injection, and if so, where? Or would you have passed away? That is to say, would you have *ceased to be*, but with the guarantee that you – you, yourself – would come into being once again when the *agony* was over? And what about the person who *would* be feeling the pain? Who would he be? . . .

Suppose that others come to you – friends, relatives, judges, clergymen – and they offer the following advice and assurance. “Have no fear,” they will say, “Take the cheaper operation and we will take care of everything. We will lay it down that the man on the table is not you, Jones, but is Smith. We will not allow this occasion to be mentioned in your biography. And during the time that you lie there on the table – excuse us (they will interject), we mean to say, during the time that *Smith* lies there on the table – we will say, ‘poor Smith’ and we will not say, even in our hearts, ‘poor Jones.’” What *ought* to be obvious to you, it seems to me, is that the laying down of this convention should have no effect at all upon your decision. For you may still ask, “But won’t that person be I?” and, it seems to me, the question has an answer.

Suppose you know that your body, like that of an amoeba, would one day undergo fission and that you would go off, so to speak, in two different directions. Suppose you also know, somehow, that the one who went off to the left would experience the most wretched of lives and that the one who went off to the right would experience a life of great happiness and value. If I am right in saying that one’s question “Will that person be I?” or “Will I be he?” always has a definite answer, then, I think, we may draw these conclusions. There is no possibility whatever that *you* would be *both* the person on the right and the person on the left. Moreover, there *is* a possibility that you would be one or the other of those two persons. And, finally, *you* could be one of those persons and yet have no memory at all of your present existence.¹⁴ It follows that it would be reasonable of you, if you are concerned with *your* future pleasures and pains, to hope that you will be the one on the right and not the one on the left – also that it would be reasonable of you, given such self-concern, to have this hope even if you know that the one on the right would have no memory of your present existence. Indeed it would be reasonable of you to have it even if you know that the one on the *left* thought he remembered the facts of your present existence.¹⁵ And it seems to me to be absolutely certain that no fears that you might have, about being the half on the left, could reasonably be allayed by the adoption of a convention, or by the formulation of a criterion, even if our procedure were endorsed by the highest authorities.

Notes

I am indebted to John Wisdom, Sydney Shoemaker, and Fred Feldman for criticisms of earlier versions of this paper. Certain paragraphs have been adapted from my “The Loose

and Popular and the Strict and Philosophical Senses of Identity," in Norman S. Care and Robert H. Grimm, eds, *Perception and Personal Identity*, by permission of Case Western Reserve University Press, Cleveland.

- 1 "Of Personal Identity," Dissertation I, in *The Whole Works of Joseph Butler*, LL.D. (London: Thomas Tegg, 1839), pp. 263–270. The dissertation is reprinted in Antony Flew, ed., *Body, Mind and Death* (New York: Macmillan, 1964), pp. 166–72.
- 2 Thomas Hobbes, *Concerning Body*, ch. ii ("Of Identity and Difference"), Section 7.
- 3 But a nonintactly persisting temporal object should be distinguished from what I shall call an "Edwardian" temporal object (after Jonathan Edwards). An Edwardian temporal object would be a temporal object which is such that, for each moment during which it exists, there is a set of parts which are what make up that object at that moment and which exist only at that moment. Hence if x is an Edwardian temporal object, then for any two times, t and t' , at which x exists, there is one set of objects which make up x at t , and another set of objects which make up x at t' , and no member of the first set has any part in common with any member of the second set. If, as some philosophers have supposed, all temporal objects are Edwardian, then no object which persists through a period of time could be said to persist intactly, as this term was defined by our metaphysician above; for no object would be such that it has the same parts at any moment of its existence that it has at any other moment of its existence. This extreme Edwardian view was defended by J. H. Woodger in *The Axiomatic Method in Biology* (Cambridge: Cambridge University Press, 1937) and by Rudolf Carnap in *Introduction to Symbolic Logic* (New York: Dover 1958), see pp. 213–16. Jonathan Edwards took this extreme view to be implied by "God's upholding created substance, or causing its existence in each successive moment." For, he reasoned, "if the existence of created substance, in each successive moment, be wholly the effect of God's immediate power, in that moment, without any dependence on prior existence, as much as the first creation out of nothing, then what exists at this moment, by this power, is a new effect, and simply and absolutely considered, not the same with any past existence. . . ." From this he was able to deduce that it is as reasonable and just to impute Adam's original sin to me now as it is to impute any sin which I may seem to remember having committed myself. (See the *Doctrine of Original Sin Defended*, part 4, ch. 2.) But this extreme view, when considered separately from the doctrine of divine re-creation, has at least the disadvantage of multiplying entities beyond necessity. . . .
- 4 See Locke's *Essay*, Book II, ch. 27, Sections 5, 6, and 8. The point made above does not, of course, imply the more extreme thesis, according to which a statement of the form, " x is identical with y ," is always elliptical for one of the form, " x is the same F as y ."
- 5 These concepts might be defined as follows: x evolves directly from y , provided: either x is identical with y , or there is no time at which x and y both exist but there is a z such that z is part of y at one time and z is part of x at a later time. (Possibly we should add that, during any subperiod between the earlier and the later time, z has the same parts that it has during any other such subperiod.) And, more generally, x evolves from y , provided: x is a member of every class C such that (i) y is a member of C and (ii) whatever directly evolves from anything that is a member of C is also a member of C . (If the definition of "evolves directly" were intended to explicate

- the ordinary use of this expression, it would doubtless be too broad; but it is not so intended.)
- 6 *Treatise*, Book I, part 4, Section vi.
 - 7 An Aristotelian who took ships seriously might say that in such a case two “substantial changes” had occurred.
 - 8 I owe this way of putting the matter to Professor John Wisdom who criticized an earlier version of this paper at Lewis and Clark College in October 1967.
 - 9 What if the parts of a thing are “simply re-arranged” – say, from ABC to CAB? If we take the term “part” in its ordinary sense, as I propose that we do, then we must say that the thing will not have persisted intact, for it will have lost some parts. If the thing changes from ABC to CAB, then it will lose BC, as well as that part that consists of the right half of B and the left half of C. . . .
 - 10 For further possibilities, see Helen M. Cartwright, “Heraclitus and the Bath Water,” *Philosophical Review*, 74 (1965), pp. 466–84.
 - 11 *De Spiritualibus Creaturis*, Article IX, ad. 16; translated as “On Spiritual Creatures,” by M. C. Fitzpatrick and J. J. Wollmuth (Milwaukee: Marquette University Press, 1949). See p. 109 of the translation.
 - 12 One might be tempted to define sameness of river bank or of river bed, say, in terms of *sameness of place*. If sameness of place is not then defined in terms of a relation to things that are said to exist in space, such a definition would seem to presuppose intact persistence of substantival space through time. (It may be noted, incidentally, that our account of “evolving” allows us to say that a thing at a later date evolves from a thing at an earlier date even though there has been no change of parts.)
 - 13 . . . “Now if we had a drug which would abolish memory for a while, and you were going to be cut for the stone, suppose the surgeon were to say, ‘You will suffer damnably, but I will administer this drug so that you will during that suffering lose all memory of your previous life. Now you have, of course, no particular interest in your suffering as long as you will not remember your present and past life, you know, have you?’” *Collected Papers*, vol. 5 (Cambridge, Mass.: Harvard University Press, 1935), p. 355.
 - 14 In this case, there might well be no *criterion* by means of which you or anyone else could decide which of the two halves was in fact yourself. . . .
 - 15 I would endorse, therefore, the following observation that Bayle makes in his article on Lucretius (see Note Q of “Lucretius,” in Pierre Bayle, *A General Dictionary, Historical and Critical*): “The same atoms which compose water, are in ice, in vapours, in clouds, in hail and snow; those which compose wheat, are in the meal, in the bread, the blood, the flesh, the bones etc. Were they unhappy under the figure or form of water, and under that of ice, it would be the same numerical substance that would be unhappy in those two conditions; and consequently all the calamities which are to be dreaded, under the form of meal, concern the atoms which form corn; and nothing ought to concern itself so much about the state or lot of the meal, as the atoms which form the wheat, though they are not to suffer these calamities, under the form of wheat.” Bayle concludes that “there are but two methods a man can employ to calm, in a rational manner, the fears of another life. One is, to promise himself the felicities of Paradise; the other, to be firmly persuaded that he shall be deprived of sensations of every kind.”

20 Identity, Ostension, and Hypostasis

W. V. O. Quine

Identity is a popular source of philosophical perplexity. Undergoing change as I do, how can I be said to continue to be myself? Considering that a complete replacement of my material substance takes place every few years, how can I be said to continue to be I for more than such a period at best?

It would be agreeable to be driven, by these or other considerations, to belief in a changeless and therefore immortal soul as the vehicle of my persisting self-identity. But we should be less eager to embrace a parallel solution of Heraclitus's parallel problem regarding a river: "You cannot bathe in the same river twice, for new waters are ever flowing in upon you."

The solution of Heraclitus's problem, though familiar, will afford a convenient approach to some less familiar matters. The truth is that you *can* bathe in the same *river* twice, but not in the same river stage. You can bathe in two river stages which are stages of the same river, and this is what constitutes bathing in the same river twice. A river is a process through time, and the river stages are its momentary parts. Identification of the river bathed in once with the river bathed in again is just what determines our subject matter to be a river process as opposed to a river stage.

Let me speak of any multiplicity of water molecules as a *water*. Now a river stage is at the same time a water stage, but two stages of the same river are not in general stages of the same water. River stages are water stages, but rivers are not waters. You may bathe in the same river twice without bathing in the same water twice, and you may, in these days of fast transportation, bathe in the same water twice while bathing in two different rivers.

We begin, let us imagine, with momentary things and their interrelations. One of these momentary things, called *a*, is a momentary stage of the river Caÿster, in Lydia, around 400 BC. Another, called *b*, is a momentary stage of the Caÿster two days later. A third, *c*, is a momentary stage, at this same latter date, of the same multiplicity of water molecules which were in the river at the time of *a*. Half of *c* is in the lower Caÿster valley, and the other half is to be found at diffuse points in the Aegean Sea. Thus *a*, *b*, and *c* are three objects, variously related. We may say that *a* and *b* stand in the relation of river kinship, and that *a* and *c* stand in the relation of water kinship.

Now the introduction of rivers as single entities, namely, processes or time-consuming objects, consists substantially in reading identity in place of river kinship. It would be wrong, indeed, to say that *a* and *b* are identical; they are

* From W. V. O. Quine, "Identity, Ostension, and Hypostasis," *Journal of Philosophy*, XLVII, 22 (1950), pp. 621–33. Reprinted by permission of the author and *Journal of Philosophy*.

merely river-kindred. But if we were to point to *a*, and then wait the required two days and point to *b*, and affirm identity of the objects pointed to, we should thereby show that our pointing was intended not as a pointing to two kindred river stages but as a pointing to a single river which included them both. The imputation of identity is essential, here, to fixing the reference of the ostension.

These reflections are reminiscent of Hume's account of our idea of external objects. Hume's theory was that the idea of external objects arises from an error of identification. Various similar impressions separated in time are mistakenly treated as identical; and then, as a means of resolving this contradiction of identifying momentary events which are separated in time, we invent a new nonmomentary object to serve as subject matter of our statement of identity. Hume's charge of erroneous identification here is interesting as a psychological conjecture on origins, but there is no need for us to share that conjecture. The important point to observe is merely the direct connection between identity and the positing of processes, or time-extended objects. To impute identity rather than river kinship is to talk of the river Caÿster rather than of *a* and *b*.

Pointing is of itself ambiguous as to the temporal spread of the indicated object. Even given that the indicated object is to be a process with considerable temporal spread, and hence a summation of momentary objects, still pointing does not tell us *which* summation of momentary objects is intended, beyond the fact that the momentary object at hand is to be in the desired summation. Pointing to *a*, if construed as referring to a time-extended process and not merely to the momentary object *a*, could be interpreted either as referring to the river Caÿster of which *a* and *b* are stages, or as referring to the water of which *a* and *c* are stages, or as referring to any one of an unlimited number of further less natural summations to which *a* also belongs.

Such ambiguity is commonly resolved by accompanying the pointing with such words as "this river," thus appealing to a prior concept of a river as one distinctive type of time-consuming process, one distinctive form of summation of momentary objects. Pointing to *a* and saying "this river" – or ὅδε δι ποταμός, since we are in 400 BC – leaves no ambiguity as to the object of reference if the word "river" itself is already intelligible. "This river" means "the riverish summation of momentary objects which contains this momentary object."

But here we have moved beyond pure ostension and have assumed conceptualization. Now suppose instead that the general term "river" is not yet understood, so that we cannot specify the Caÿster by pointing and saying "This river is the Caÿster." Suppose also that we are deprived of other descriptive devices. What we may do then is point to *a* and two days later to *b* and say each time, "This is the Caÿster." The word "this" so used must have referred not to *a* nor to *b*, but beyond to something more inclusive, identical in the two cases. Our specification of the Caÿster is not yet unique, however, for we might still mean any of a vast variety of other collections of momentary objects, related in other modes than that of river kinship; all we know is that *a* and *b* are among its constituents. By pointing to more and more stages additional to *a* and *b*, however, we eliminate more and more alternatives, until our listener, aided by his own tendency to favor the most natural groupings, has grasped the idea of the

Cayster. His learning of this idea is an induction: from our grouping the sample momentary objects *a*, *b*, *d*, *g*, and others under the head of Cayster, he projects a correct general hypothesis as to what further momentary objects we would also be content to include.

Actually there is in the case of the Cayster the question of its extent in space as well as in time. Our sample pointings need to be made not only on a variety of dates, but at various points up and down stream, if our listener is to have a representative basis for his inductive generalization as to the intended spatio-temporal spread of the four-dimensional object Cayster.

In ostension, spatial spread is not wholly separable from temporal spread, for the successive ostensions which provide samples over the spatial spread are bound to consume time. The inseparability of space and time characteristic of relativity theory is foreshadowed, if only superficially, in this simple situation of ostension.

The concept of identity, then, is seen to perform a central function in the specifying of spatio-temporally broad objects by ostension. Without identity, *n* acts of ostension merely specify up to *n* objects, each of indeterminate spatio-temporal spread. But when we affirm identity of object from ostension to ostension, we cause our *n* ostensions to refer to the same large object, and so afford our listener an inductive ground from which to guess the intended reach of that object. Pure ostension plus identification conveys, with the help of some induction, spatiotemporal spread. . . .

21 Identity: an Excerpt from *Quiddities**

W. V. O. Quine

The term is used loosely. We speak of identical twins. We say that you and I drive identical station wagons. But for all the looseness of common usage, the term in its strict sense is as tight as a term can be. A thing is identical with itself and with nothing else, not even its identical twin.

David Hume was puzzled. Identity seems like a relation, but it does not relate things pairwise as a relation should; things are identical only to themselves. How then does identity differ from a mere property? Moreover, it applies to everything. How then does it differ from the mere property of existence, the property enjoyed by everything?

It is hard to project oneself into the confusions of even so gifted a mind as Hume's, after those confusions have given way to the progress of science.

* Reprinted by permission of the author and publisher from *Quiddities: an Intermittently Philosophical Dictionary*, by W. V. Quine (Cambridge, Mass.: Harvard University Press). Copyright © 1987 by the Presidents and Fellows of Harvard College.

A relation is now clearly conceived as consisting of pairs of objects: the uncle relation comprises all the uncle–nephew and uncle–niece pairs. The identity relation comprises all and only the repetitious pairs, $\langle x, x \rangle$; $\langle x, x \rangle$ is still not to be confused with x .

On confusions over identity see also USE VERSUS MENTION.¹ And there are the makings of further confusion in the following reflection: evidently to say of anything that it is identical with itself is trivial, and to say that it is identical with anything else is absurd. What then is the use of identity? Wittgenstein put this question.

Genuine questions of identity can arise because we may refer to something in two ways and leave someone wondering whether we referred to the same thing. Thus I mention Simon, someone mentions Peter, and we explain that Simon is Peter; they are identical. It is neither trivial to say so nor absurd to doubt it.

There is little need to give a man two names, nor much interest in developing an identity concept solely for that contingency. What is more important is reference to something not by two names but by two descriptions, or by a name and a description. We need to be able to identify Ralph with the man who mows the lawn, and his house with the one nearest the station. Identities such as these permeate our daily discourse.

A philosophical riddle was propounded in antiquity about the identification, early and late, of a ship belonging to Theseus: was it the same ship despite successive replacement, over the years, of all its parts? The same riddle is familiar from Heraclitus in application to a river: you cannot step into the same one twice, he claimed, for its substance is continually renewed. For that matter, is Ralph as of now the same man that was mowing the grass eight years ago, if, as the saying goes, our bodily substance fully renews itself in the course of seven years? Are you indeed still you after all this time?

These three riddles – one, really – are wrongly reckoned as identity crises; they hinge not on the nature of identity, but on what we choose to count as a boat, a river, a person. Words are instruments, and their vagueness is tolerated where it does not impair their utility.

The continuing identity of a person over the years is predicated not on his retention of substance, but on the continuity of replacement of substance, and the continuity of change in his shape, mass, and habits. Continuity also of his memory is expected, but occasionally a lapse in this quarter is taken in stride. How far back to place a person's beginning – whether at birth or conception or somewhere between – is up for grabs, because the utility of the word 'person' has not hinged much on that detail until recent times.

A point that has seemed strangely in need of being driven home is that it is simply a question of the human use of the word 'person', whether the actual use or some use that is being proposed. It is not a question of discerning a hitherto undiscovered meaning of the word 'person'. Words, as Humpty Dumpty appreciated, are no more than what we make them.

I have dwelt here on persons, but the case is the same with the river of Heraclitus and the boat of Theseus. The truth of an identity statement hinges on the general term involved or implied – 'person', 'boat', 'river'. Ralph is the

same person now as eight years ago, but his stages are distinct. When on the other hand the ornithologist says ‘This is the same as that’, pointing in two directions, it would be absurd to accuse him of meaning what he says. He means that the species of this bird is identical with the species of that one.

A vital use of identity lurks unobserved in much of our use of ‘only’ and ‘else’ and ‘nothing but’. When I say that the hiding place is known to Ralph and only him, nobody else, I mean to say two things: that Ralph knows the hiding place and that whoever knows the hiding place is identical with Ralph. To say that there is no God but Allah is to affirm, of whatever Gods there be, that Each, or He, is identical with Allah.

Note

- 1 Another entry in the author’s *Quiddities*. – [Eds]
-

22 In Defense of Stages¹: Postscript B to “Survival and Identity”*

David Lewis

Some would protest that they do not know what I mean by “more or less momentary person-stages, or time-slices of continuant persons, or persons-at-times.” Others do know what I mean, but don’t believe there are any such things.

The first objection is easy to answer, especially in the case where the stages are less momentary rather than more. Let me consider that case only; though I think that instantaneous stages also are unproblematic, I do not really need them. A person-stage is a physical object, just as a person is. (If persons had a ghostly part as well, so would person-stages.) It does many of the same things that a person does: it talks and walks and thinks, it has beliefs and desires, it has a size and shape and location. It even has a temporal duration. But only a brief one, for it does not last long. (We can pass over the question how long it can last before it is a segment rather than a stage, for that question raises no objection of principle.) It begins to exist abruptly, and it abruptly ceases to exist soon after. Hence a stage cannot do everything that a person can do, for it cannot do those things that a person does over a longish interval.

That is what I mean by a person-stage. Now to argue for my claim that they exist, and that they are related to persons as part to whole. I do not suppose the

* From David Lewis, *Philosophical Papers*, vol. I (New York: Oxford University Press, 1983). Reprinted by permission of the author.

doubters will accept my premises, but it will be instructive to find out which they choose to deny.

First: it is possible that a person-stage might exist. Suppose it to appear out of thin air, then vanish again. Never mind whether it is a stage of any person (though in fact I think it is). My point is that it is the right sort of thing.

Second: it is possible that two person-stages might exist in succession, one right after the other but without overlap. Further, the qualities and location of the second at its appearance might exactly match those of the first at its disappearance. Here I rely on a *patchwork principle* for possibility: if it is possible that X happen intrinsically in a spatiotemporal region, and if it is likewise possible that Y happen in a region, then also it is possible that both X and Y happen in two distinct but adjacent regions. There are no necessary incompatibilities between distinct existences. Anything can follow anything.

Third: extending the previous point, it is possible that there might be a world of stages that is exactly like our own world in its point-by-point distribution of intrinsic local qualities over space and time.

Fourth: further, such a world of stages might also be exactly like our own in its causal relations between local matters of particular fact. For nothing but the distribution of local qualities constrains the pattern of causal relations. (It would be simpler to say that the causal relations supervene on the distribution of local qualities, but I am not as confident of that as I am of the weaker premise.)

Fifth: then such a world of stages would be exactly like our own simpliciter. There are no features of our world except those that supervene on the distribution of local qualities and their causal relations.

Sixth: then our own world is a world of stages. In particular, person-stages exist.

Seventh: but persons exist too, and persons (in most cases) are not person-stages. They last too long. Yet persons and person-stages, like tables and table-legs, do not occupy spatiotemporal regions twice over. That can only be because they are not distinct. They are part-identical; in other words, the person-stages are parts of the persons.

Let me try to forestall two misunderstandings. (1) When I say that persons are maximal R-interrelated² aggregates of person-stages, I do *not* claim to be reducing “constructs” to “more basic entities.” (Since I do not intend a reduction to the basic, I am free to say without circularity that person-stages are R-interrelated aggregates of shorter person-stages.) Similarly, I think it is an informative necessary truth that trains are maximal aggregates of cars interrelated by the ancestral of the relation of being coupled together (count the locomotive as a special kind of car). But I do not think of this as a reduction to the basic. Whatever “more basic” is supposed to mean, I don’t think it means “smaller.” (2) By a part, I just mean a subdivision. I do not mean a well-demarcated subdivision that figures as a unit in causal explanation. Those who give “part” a rich meaning along these lines³ should take me to mean less by it than they do.

Notes

- 1 On this topic I am much indebted to discussions with Saul Kripke and with Denis Robinson. Kripke's views on related matters were presented in his lectures on "Identity through Time," given at Princeton in 1978 (and elsewhere); Robinson's in "Re-Identifying Matter," *Philosophical Review*, 91 (1982), pp. 317–41.
- 2 "The R-relation" is that "relation of mental continuity and connectedness that matters in survival." – [Eds].
- 3 Such as D. H. Mellor, in his *Real Time* (Cambridge: Cambridge University Press, 1981), ch. 8.

23 Some Problems about Time*

Peter Geach

When I was invited to give this philosophical lecture and was considering which subject to talk about, I found my mind turning towards a great philosopher, a Fellow of this Academy, who died just forty years ago: John Ellis McTaggart. I consider myself very lucky to have been introduced to McTaggart's work early in my philosophical life; McTaggart sets high standards of clarity, rigour, and seriousness for a young philosopher to try to live up to. I suppose McTaggart is little read nowadays; he was a metaphysician, and metaphysics is not in fashion; even those who stridently call out for metaphysics to be done do not produce any themselves, and ignore the one British metaphysical work of genius in this century. But I make bold to put into McTaggart's mouth the words of one of his favourite poets:

But after, they will know me. If I stoop
Into a dark tremendous sea of cloud,
It is but for a time; I press God's lamp
Close to my breast; its splendour, soon or late,
Will pierce the gloom: I shall emerge one day.
(Browning's *Paracelsus*)

I shall be talking about a subject that was of central concern for McTaggart – the problems of time. I begin by examining a view of time that is now widely held in one form or another. In its crudest form, this view makes time out to be simply one of the dimensions in which bodies are extended; bodies have not three dimensions but four. An instantaneous solid is as much a mere artificially

* © The British Academy 1966. Reproduced by permission of the author and publisher from *Proceedings of the British Academy*, vol. 11.

abstracted aspect of a concrete thing as a surface without depth is; photographs of a man at different ages represent different three-dimensional cross-sections of a four-dimensional whole. Time is only subjectively and relatively distinct from the other dimensions in which things are extended. We may illustrate this by the simile of horizontal and vertical; though at any given point on the Earth's surface a unique vertical direction can be picked out, there is no cosmic distinction of horizontal and vertical, and people at different places on the Earth will take different directions to be vertical. Or again, as Quine says: 'Just as forward and backward are distinguishable only relative to an orientation, so, according to Einstein's relativity principle, space and time are distinguishable only relative to a velocity'; and he speaks of 'an hour-thick slice of the four-dimensional material world . . . perpendicular to the time axis'.¹

Since Einstein, indeed, this sort of view has been very popular with philosophers who try to understand physics and physicists who try to do philosophy. Some of the arguments used in its favour are decidedly odd. Thus, it is supposed to be supported by the fact that we can represent local motion in a graph with axes representing space and time; the line drawn on the graph-paper is taken to represent a 'world line' or 'four-dimensional worm' stretching through a 'space-time continuum'. We might as well be asked to believe that the use of temperature charts requires the physical existence of 'world lines' in a 'temperature-time continuum'. Obviously the two axes of a graph, though themselves magnitudes of the same sort, may represent quite heterogeneous magnitudes.

Another odd argument is that modern formal logic, in particular quantification theory, can be applied to propositions about physical objects only if these objects are regarded as four-dimensional. This is not at all true. In Quine's *Methods of Logic*,² for example, we learn from his precept and practice how to apply modern formal logic to propositions of ordinary language; there is no obstacle to such application, he points out, in the sort of ambiguity that is resolvable by considering 'circumstances of the argument as a whole – speaker, hearer, scene, date, and underlying problem and purpose'; all that we really need is that the sense and reference of expressions should 'stay the same throughout the space of the argument' (*Methods of Logic*, p. 56). In a later work, *Word and Object*, Quine does indeed pay lip service to the need of four-dimensional talk; but the parts of his book essentially involving such talk could easily be cut out; the great majority of the sentences given as logical examples are in a streamlined version of English, not in four-dimension-ese; and Quine's discussions almost all relate to the mode of significance of terms and the structure of propositions in this near-vernacular language. Thus it is not open to Quine to maintain that if we are to be 'serious about applying modern logic to temporal entities', in particular if we are so to apply quantification theory, then we need 'the four-dimensional view' as 'part and parcel' of what we are doing.³

Logic would not be much use for arguments about concrete realities if we had to hold that, outside pure mathematics, logic applied only to a language yet to be constructed, one that nobody talks or writes. Logic was a going concern, and was applied to inferences about concrete matters, long before anyone ever

dreamed up four-dimensional language. If all these past applications of logic had to be written off as misconceived, we could not have high hopes for future applications to an as yet non-existent language. Quine is certainly not himself prepared to write off so much of logic's past.

Nor ought any logician to try to accommodate his doctrines to demands made in the name of contemporary physics. Logic must be kept rigid, come what may in the way of physical theories; for only so can it serve as a crowbar to overthrow unsatisfactory theories. Lavoisier remarked that the phlogistonists ascribed different and indeed incompatible properties to phlogiston in order to explain different experimental results; what a good thing there were not then logicians prepared to bend logic in the interests of the phlogiston theory – to say that these were 'complementary' accounts of phlogiston, both true so long as you did not combine them!

The view that time is merely a fourth dimension in which things extend is in any event quite untenable. On this view, the variation of a poker's temperature with time would simply mean that there were different temperatures at different positions along the poker's time-axis. But this, as McTaggart remarked, would no more be a *change* in temperature than a variation of temperature along the poker's length would be.⁴ Similarly for other sorts of change. A man's growth would be regarded as the tapering of a four-dimensional body along its time-axis from later to earlier; but this again would no more be a change than is a poker's tapering along its length towards its point. We thus have a view that really abolishes change, by reducing change to a mere variation of attributes between different parts of a whole. But, as McTaggart again remarked, no change, no time; the view we are discussing countenances talk of a *time*-axis, but such talk is inappropriate on these premises.

The view really commits us to saying that time is an illusion. In Absolute Reality there is a changeless arrangement of four-dimensional solids; in Present Experience certain aspects of this arrangement appear to our perceptions as changes of three-dimensional bodies. McTaggart too thought that time was an illusion – though he had a very different account to give of the Absolute Reality that we misperceive as changeable bodies. But time cannot be an illusion; and certain arguments of McTaggart's own, ironically enough, are readily adapted to prove this.

The arguments in question show that certain features other than time in our experience cannot possibly be illusory. Thus, there really must be error in the universe; for there appears to be error, and if this appearance is false, then again there is error.⁵ Parmenides and Mrs Eddy alike are in a quandary what to say about the 'error of mortal mind'. Again (as mention of Mrs Eddy reminds me) there is plain incoherence in the optimistic doctrine that misery is only an 'error of mortal mind': if my 'mortal mind' thinks I am miserable, then I am miserable, and it is not an illusion that I am miserable.⁶ (Of course, so far as this goes, it might still be true that our misery would vanish if we all perceived things without illusions; McTaggart could consistently hold that, as he in fact did.) But now, quite similarly, even if my distinction between past, present and future aspects of physical things is a fragmentary misperception of changeless

realities, it remains true that I have various and uncombinable illusions as to which realities are present. I must therefore have these illusions not simultaneously but one after another; and then there is after all real time and real change.

One might perhaps hold that time and change are only in the mind, in the sense that only a mind lives through time and undergoes change; in this sense, misery is 'only in the mind'. But this sense of the phrase must be sharply distinguished from the sense in which a thing's being 'only in the mind' implies its unreality. A man can no more 'only think' he has changing impressions of the world than he can 'only think' he is unhappy.

McTaggart tried to show that there was a difference between error and misery, on the one hand, and time on the other. A state of error or misery cannot be just illusory, because to be under such an illusion would be a state of real error or misery; but a state of self-consciousness that presents itself as temporal need not, he argued, be on that account really temporal.⁷ This distinction is sound, so far as it goes; however, it misses the point that temporal appearance requires the existence of diverse *and uncombinable* impressions as to what is present. I am not arguing that *each single* state of self-consciousness must really be temporal because it presents itself as temporal; I am arguing that the *variety* of states each person experiences must really be, as it appears to be, a change in his experience, because these states are combinable only in succession, and not simultaneously.

However, we might try modifying the view of a four-dimensional and changeless *physical* reality by allowing that there is real change in the world of experience. There would then be a set of observing minds each of which continuously 'moved on' from one part of the four-dimensional physical world to another; though the ordered cross-sections of four-dimensional bodies would then appear to an observing mind as earlier and later, they would not really stand in temporal relations – only in the experiences of the observing minds would there be real time and change.

To make this story consistent, the observing minds must be supposed incorporeal and physically dimensionless; otherwise there would, contrary to hypothesis, be real change in the physical world. How then can mind be said to *move*? We need not make heavy weather of this; a simple analogy may help us out. The order of printed words on a page is an unchanging spatial order; but it appears as a temporal order to a reader whose attention moves on from word to word and from line to line – and surely nobody will have felt a difficulty over my use of 'moves on' in this context.

The theory I have just sketched is *one* theory of time to be found in the opening discourse of the Time Traveller in Wells; and it is a theory that lends itself to speculative developments. Why should we assume that an observing mind's attention must always travel on in one direction like that of a slow, plodding, reader? Even normal minds may sometimes slip back to a part of the physical continuum that their attention has already scanned; Wells in fact gives us this 'explanation' of vivid reminiscence. And why should not a practised observer learn a skill like that of the practised reader, of looking before and after, seeing, for example, by anticipation those parts of the physical continuum

that he would observe only later on by the normal movement of his focus of observation?

This whole theory, though, is open to the gravest objections. It incorporates an extreme form of Cartesian dualism: the human body is a changeless four-dimensional solid, the human mind a changeable dimensionless entity that reads off data for its *cogitationes* along one dimension of this solid. The theory is thus exposed to all the general arguments against Cartesian dualism; and also, to certain special objections. Though admitting an inability to understand the mind's power to move the body, Descartes did not venture to deny this power; even the Occasionalist disciples of Descartes, who did deny such a power to the mind, held that God would miraculously tamper with our normally automatic bodily machinery so that within limits it should move as we wish. On the theory we are now considering, there is no time or change except in minds; the four-dimensional physical world is an absolutely fixed order, not to be altered by any will, human or divine. The mind just cannot interfere with what will physically come to be; in fact, the very phrase I just used is only a loose manner of referring to those regions of the changeless four-dimensional world which a given mind is next going to observe.

Such a view would reduce the will of man to an impotent chimera, buzzing in a void and feeding upon second intentions (in the words of the perhaps legendary medieval conundrum). It may be beneath the dignity of philosophy to say 'We know our will is free, Sir, and there's an end on't'; but we do know that our plans and purposes radically alter our physical environment, and there's an end on't; any contrary theory, however plausibly argued, just has to be false.

The view that our decisions cannot bring about physical changes may be called *fatalism*. Fatalism has a bad name among philosophers, like solipsism; arguments in favour of either will be dismissed as ingenious sophistries, and a reduction of a thesis to either counts as checkmate in the philosophical game. Determinists are mostly anxious to repudiate fatalism: to maintain only that human designs are predictable from causes, not that they do not have effects. I think this defence is open only to some varieties of determinist; other determinists evade fatalism only by a sort of doublethink; indeed, it sometimes looks as though doublethink were being deliberately advocated as a way out of free-will puzzles. Be that as it may, fatalism naked and undisguised has a strong imaginative and emotional appeal for many people. John Buchan was such a person; in his admirable novel *The Gap in the Curtain* he worked out the consequences of that purely mental 'time-travel' into the future which, as we just saw, would be allowed as a theoretical possibility by the theory of mental observers' scanning an unchanging physical world. I will not spoil this novel, for those of you who have not read it, by giving away the plot; I will just remark that the fatalism is consistently upheld. Buchan's characters merely get a glimpse of the future, with no power to change it; as in Oriental tales of Fate, what is to be comes to pass regardless of man's designs.

We find it easy to imagine the future as a country into which we are travelling and which is there before we travel into it; a country of which we might get a Pisgah sight through a break in the clouds before we actually get there. Here it

is interesting to notice the change of meaning that has happened to the phrases ‘the next world’ or ‘the world to come’. They originally meant the *age* to come, *vitam venturi saeculi*, which is to follow the return of Messiah; nowadays, to many people, they suggest some other *place*, as when one calls Mars ‘another world’.

The fundamental difficulty about this picture is quite different from the obvious one. At the price of adopting dualistic fatalism, one can, as I have shown, make some kind of sense out of this talk about travelling; it is not the travelling that raises the real difficulty, but the destination. What *is* (say) the England of 1984? Is there really such an object *in rerum natura*, distinct from the England of 1965?

It is very natural to talk this way: very natural to think of the successive phases in an object’s history as ordered parts of the object itself – somehow like the segments of a worm’s body. I shall here borrow an example from McTaggart; he, of course, did not believe in Time, but his example suits well enough for recent statements of this view, for example, by Quine and J. J. C. Smart. The phrase ‘St Paul’s in the nineteenth-century’ would designate an individual, and so would, for example, ‘St Paul’s in 1801’; and these must be two distinct individuals, for many predication that are true of St Paul’s in (the whole of) the nineteenth-century are false of St Paul’s in 1801 and vice versa. Moreover, ‘St Paul’s in 1801’ will designate a part of the whole designated by ‘St Paul’s in the nineteenth-century’; and if we take the individuals designated by ‘St Paul’s in 1801’, ‘St Paul’s in 1802’, up to ‘St Paul’s in 1900’, they will together include all the content of the individual designated by ‘St Paul’s in the nineteenth-century’.⁸

I think this account involves an erroneous analysis of propositions into subject and predicate. Let us consider one sort of predication that might be used to discriminate the individuals designated by phrases like ‘St Paul’s in 1856’: if you were answering the question ‘How many visitors were there?’ you might have to give a different answer for each year of the nineteenth-century and of course a different answer again for the century as a whole. We can certainly consider a proposition: ‘There were *n* visitors to St Paul’s in 1856’, as a predication about St Paul’s; I have chosen this example to show that the problem I am raising does not arise from superficial grammatical considerations, for here we have in any case a logical subject of predication that is not a grammatical subject.⁹ The question is whether we can also analyse the same proposition as a predication about St Paul’s in 1856; as attaching to the subject ‘St Paul’s in 1856’ the predicate: ‘There were *n* visitors to . . .’. This analysis is not excluded because the other is possible; we may surely analyse ‘Queen Anne’s hat was red’ equally well as predication of Queen Anne’s hat that it was red and as predication of Queen Anne that she had a red hat; similarly, it could be argued, our example *both* predicates something of St Paul’s *and* predicates something of St Paul’s in 1856. But I think the second analysis can be excluded on other grounds; phrases like ‘St Paul’s in 1856’ cannot be taken as logical subjects at all.

Let us shift to another example: ‘McTaggart in 1901 was a philosopher holding Hegel’s dialectic to be valid, and McTaggart in 1921 was a philosopher not

holding Hegel's dialectic to be valid.' If we regarded 'McTaggart in 1901' and 'McTaggart in 1921' as designating two individuals, then we must also say they designate two philosophers: one philosopher believing Hegel's dialectic to be valid, and another philosopher believing Hegel's dialectic not to be valid. To be sure, on the view I am criticizing, the phrases 'McTaggart in 1901' and 'McTaggart in 1921' would not designate two philosophers, but two temporal slices of one philosopher. But just that is the trouble: for a predicate like 'philosopher believing so-and-so' can of course be true only of a philosopher, not of a temporal slice of a philosopher. So if our example, which is a plain and true¹⁰ empirical proposition, were construed as a conjunction of two predication about temporal slices of McTaggart, then it would turn out necessarily false; which is an absurd result. The absurdity does not come about just for my chosen example; it arises equally for Quine's example 'Tabby at *t* is eating mice',¹¹ for a cat can eat mice at time *t*, but a temporal slice of a cat, Tabby-at-*t*, cannot eat mice anyhow.

The friends of temporal slices will no doubt here pray leave to amend the examples so that they contain predicates fitting temporal slices, instead of predicates like 'philosopher believing so-and-so' or 'cat eating mice', which fit living beings and not temporal slices of living beings. But we ought not to grant them leave to amend. The whole ground for treating, for example, 'McTaggart in 1901' and 'McTaggart in 1921' as designating two distinct individuals was that we seemed to find predicates true of the one and false of the other. But now we find that such predicates as appear in ordinary empirical propositions are often of a kind that could not be true of temporal slices; so the ground for recognizing temporal slices as distinct individuals has been undercut; and we ought to reject temporal slices from our ontology, rather than cast around for new-fashioned predicates to distinguish them by.

I conclude that temporal slices are merely 'dreams of our language'. It is no less a mistake to treat 'McTaggart in 1901' and 'McTaggart in 1921' as designating individuals than it would be so to treat 'nobody' or 'somebody'. If we take the name 'McTaggart' as logical subject of both clauses in our example, no such troubles arise; for, on the face of it, the predicates we are attaching to this subject are a compatible pair, namely 'philosopher believing in 1901 that Hegel's dialectic is valid' and 'philosopher not believing in 1921 that Hegel's dialectic is valid'.

Predicates of this sort, in which dates are mentioned, are a long way above the most fundamental level of temporal discourse. Our ability to keep track of the date and the time of day depends on a set of enormously complicated natural phenomena; such phenomena, serving 'for signs and for seasons and for days and for years', might easily not have been available. We can easily imagine rational beings, living on a cloud-bound planet like Venus, who had no ready means of keeping dates or telling the time, and were too well endowed by Nature with the necessities and amenities of life to feel any need to contrive such means. Clearly, such creatures might still speak of one thing's happening at the same time as another, or after another, and might have past, present, and future tenses in their language. This is grass-roots temporal discourse; it is per-

verse to try to analyse it by means of the vastly more complex notions that are involved in saying ‘in 1901’ or ‘at time t ’.

In particular, it is definitely wrong to analyse an unsophisticated simultaneity proposition, like ‘Peter was writing a letter and (at the same time) Jenny was practising the piano’, in terms of what happened at some one time t – ‘For some time t , Peter was writing a letter at t and Jenny was practising the piano at t ’. Such a use of ‘at the same time’ as we have here does not involve any reference to an apparatus or technique for telling the time (and still less, a reference to Absolute Time). On the contrary, telling the time depends on knowing some of these primitive simultaneity propositions to be true. Telling the time by an ordinary clock involves observing that the long hand points (say) to the twelve and the short hand *at the same time* points to the six; clearly we do not need another clock to verify that it *is* at the same time. A physicist may protest that he simply cannot understand ‘at the same time’ except via elaborate stipulations about observing instruments; his protest may be dismissed out of hand, for he could not describe the set-up of any apparatus except by certain conditions’ having to be fulfilled *together*, i.e. simultaneously, by the parts of the apparatus.

Simultaneity is involved in empirical statements; but it is not an empirical relation like neighbourhood in space. The natural expression for simultaneity is not a relative term like ‘simultaneous with’, but a conjunction like ‘while’ joining clauses; it is an accident of English idiom that ‘at the same time’ seems to refer to a certain *time* that has to be *the same*, and the words for ‘at the same time’ in other languages – Latin *simul*, Greek *άμα*, Polish *razem* – have no such suggestion.

These conjunctions joining clauses no more stand for a proper relation than, for example, ‘or’ does. If I say I can see with my myopic eyes something over there that is *either* a hawk *or* a hand-saw, I do not claim to observe a hawk in the act of being an alternative to a hand-saw; to try to conceive a relation of alternateness between such concrete objects would soon land us in paradoxes. Like alternateness, simultaneity is not a relational concept, but is one of those concepts called transcendental by the medievals, formal in Wittgenstein’s *Tractatus*, and topic-neutral by Ryle; the last term is the most informative of the three – it shows us that these concepts are not departmental but crop up in discourse generally.

Because of this topic-neutrality, ‘at the same time’ belongs not to a special science but to logic; its laws are logical laws, like the so-called De Morgan laws for ‘or’. Physicists may have interesting things to tell us about the physical possibilities of synchronizing clocks by the transmission of electromagnetic signals; but this information is wholly irrelevant to the logic of basic simultaneity propositions. Our practical grasp of this logic is not to be called in question on account of recondite physics; for without such a practical grasp we could not understand even elementary propositions in physics, so a physicist who casts doubt upon it is sawing off the branch he sits upon. And a theoretical account of this logic must be given not by physicists but by logicians.

I remarked just now that the natural, primitive, way to speak of simultaneity

is to use a conjunction joining clauses, rather than a relational term like ‘simultaneous with’. In general, I think we need to get events expressed in a propositional style, rather than by using name-like phrases (what Kotarbiński has called ‘onomatoids’). We need, that is to say, propositions like ‘Wellington fought Napoleon at Waterloo after George III first went mad’, rather than ‘George III’s first attack of madness is earlier than the Battle of Waterloo’.

Some years ago philosophers were all the while talking of people and things as being ‘logical constructions out of events’. This was a topsy-turvy view: nobody ever has talked or is going to talk a language containing no names of people or things but only names of events, and the claim that our language could in principle be replaced by such a language is perfectly idle. On the other hand, any sentence in which an event is represented by a noun-phrase like ‘Queen Anne’s death’ appears to be easily replaceable by an equivalent one in which this onomatoid is paraphrased away; we could use instead a clause attaching some part of the verb ‘to die’ to the subject ‘Queen Anne’. Any ordinary sentence, that is, will allow of such paraphrase; philosophical sentences like ‘Queen Anne’s death is a particular’ may resist translation, but we can get on very well without them. On the other hand, ‘Queen Anne’s death is a past event’ goes over into ‘Queen Anne has died’ (or ‘is dead’), and ‘The news of Queen Anne’s death made Lord Bolingbroke swear’ goes over into ‘Lord Bolingbroke swore because he heard Queen Anne had died’. Cutting out the onomatoids in this way, we get a manner of speaking in which persons and things are mentioned but events do not even appear to be mentioned; so far from its being people and things that are logical constructions out of events, events are logical constructions out of people and things.

McTaggart’s proof that time is unreal has often been criticized on the score that it essentially depends on treating ‘past’, ‘present’, and ‘future’ as logical predicates in propositions like ‘Queen Anne’s death is past’. I think I could show that this is too easy a way of dismissing McTaggart; some at least of his arguments could be restated so as to avoid the criticism. Anyhow, the critics have oddly failed to see that if the ostensible predicate ‘past’ in ‘Queen Anne’s death is past’ is not to be parsed as a logical predicate, then equally the phrase ‘Queen Anne’s death’ is not to be regarded as being, or even going proxy for, a logical subject.

In his lectures on Logical Atomism, Bertrand Russell forcibly argued that a phrase like ‘the Kaiser’s death’ is not even a description, let alone a name, of an object nameable by a proper name, but rather goes proxy for the corresponding proposition ‘The Kaiser is dead’. For example, people might in 1918 assert or deny or doubt the Kaiser’s death; this shows that the onomatoid ‘the Kaiser’s death’ goes proxy for a clause ‘The Kaiser is dead’. (Observe that it would be nonsense to speak of asserting or denying the Kaiser’s *spiked helmet* – this phrase is a description of a nameable object.)

To be sure, later on in the same course of lectures Russell tells us that a person or thing is ‘a series of classes of particulars, and therefore a logical fiction’.¹² This often happens with a work of Russell’s: you pays your money and you takes your pick. I have no hesitation which of the two views I should pick. For the first, there

are sound logical reasons; for the second, there is only an ontological prejudice of Russell's – 'the things that are really real last a very short time'.¹³

There is more than this wrong with Russell's treatment of persons. He is trying to ride two theories of classes at once: the no-class theory (that classes are fictions) and what we may call the composition theory (that classes are composed of their members and series of their terms). Only the composition theory, *plus* the segmented-worm idea of a person's temporal parts, can make it plausible that a series of classes is what a person is; Russell then concludes that, being a series of classes, a person is a fiction, by jumping over to the no-class theory. I doubt the staying power of either horse; to try to ride both at once is really desperate.

If my own arguments are sound, time-order and space-order are radically different. We can indeed verbally use such forms as '*A* is between *B* and *C*' for either sort of order; but I think this only leads to confusion. Spatial order relates individual objects: Bill is between Tom and Joe. We can get grammatically similar sentences about time-order by using onomatoids like 'the Battle of Waterloo'; but the logically perspicuous way to represent time-order is a complex sentence whose sub-clauses *report* (not name) events, these clauses being joined by temporal conjunctions like 'and then', 'and at the same time', 'while', etc. Such conjunctions, which form narrative propositions out of simpler ones, are of course quite different in category from relative terms that form propositions out of names or name-substitutes; and time 'relations' are not to be spoken of in the same logical tone of voice as space relations.

If in '*x* adjoins *y*' we replace the schematic letters by names or descriptions of bodies, the resulting proposition will not be even a description, let alone a name, of something that can itself adjoin a body. On the other hand, if we replace the letters in '*p* and then *q*' by narrative propositions like 'Queen Anne died' or 'Wellington defeated Napoleon', the result is again a narrative proposition reporting a course of events; and this can be used to build up more complex narrative propositions, of such forms as 'while *r*, (*p* and then *q*)'. Nothing analogous to this is possible for propositions describing spatial order: '*x* is between (*y* is above *w*) and *z*' gives us mere gibberish if we replace the schematic letters by names.

Miss Anscombe has raised an interesting objection to this argument. She rightly remarked that from a grammatical point of view 'where' will serve as a conjunction forming sentences out of sentences just as well as 'when' will. To give an example: we may join 'The Dome of the Rock was built' and 'Solomon's Temple was built' either with 'when' or with 'where' so as to make sense; the 'when' proposition is of course false, but that is no objection to it as a logical example. Some medieval logicians did in fact class both conjunctions as means of forming 'hypotheticals', i.e. complex propositions, out of simpler propositions; there were temporal hypotheticals and local hypotheticals. But without going into the analysis of local hypotheticals, we can quickly see that their logic does not run at all parallel to that of temporal hypotheticals. For, as I just now remarked, a temporal hypothetical '*p* and then *q*' can be used as a clause in a more complex one such as 'while *r*, (*p* and then *q*)'. We can play no similar

tricks with local hypotheticals: ‘where *r*, (*p* to the southeast of where *q*)’ – e.g. ‘Where the Dome of the Rock was built, (the Pyramids were built to the south-east of where the Parthenon was built)’ – is just not an intelligible build-up for a proposition. The Pyramids just *were* built to the south-east of where the Parthenon was built; this just *is* so, and there’s no sense in trying to say *where* it was so. The more we try to assimilate space and time, the more we shall find ourselves logically impeded from doing so.

I am strongly inclined to maintain that the rules for our grass-roots employment of temporal conjunctions – not only ‘at the same time’, but also ‘before’ and ‘after’ – belong to the domain of formal logic. This claim is highly disputable, and I can here only sketch my reasons for it. They derive from the branch of logic called modal logic – the logic of necessity and possibility. Tie-ups between modal logic and our elementary temporal discourse might well have been suspected; for is not the future precisely the domain of unrealized possibility? Arthur Prior was a pioneer in these researches, and further work has been done by a band of younger logicians, including Hintikka, Dummett, Lemmon, and Kripke. The March 1965 number of the *Journal of Symbolic Logic* contains an important article on the adequacy of certain modal-logic calculi for dealing with temporal order.¹⁴ I feel confident that much progress will be made in these researches; I am not invoking anyone’s authority, but you can see that the idea of clearing up time problems with tools of modal logic is not just a programme vaguely sketched by me here and now. Nor would it be fair to say that calling these researches ‘logic’ is an arbitrary bit of nomenclature; modal logic is traditionally a part of logic, from Aristotle onwards; and the systems now being used in tense logic are based on modal systems originally devised by Lewis and Langford with no such application in mind.

People have long felt inclined to ascribe to some truths about time the same necessity as logical truths have: one could as easily describe a world in which *modus ponens* broke down as a world in which time was two-dimensional or the past was changeable. If I turn out to have been right in my conjecture about the possibility of reducing to modal logic the rules that govern temporal discourse, then this feeling will have been a divination of the truth. Geometrical truths, as is well known, are not necessary in this way; we can describe without contradiction a world whose geometry is non-Euclidean just as well as a Euclidean world. But if these basic truths about time are logical, then a world differing from ours in regard to them is a mere chimera.

However this may be, it is certain that there is a category-difference between space and time order, between events and individuals; and this can be brought out in quite ordinary language. But sometimes important things are too close to us to be clearly visible, or are concealed like faces in a puzzle picture; the labour of bringing them into plain view is then not wasted. And mistakes and confusions about this sort of thing are both common – witness the reams of nonsense about time you can find in bookshops – and of some practical importance. Squandering vast sums on foolish enterprises is an everyday occurrence; we may yet be witnesses of a ‘time race’ between East and West. Will the US time explorer get back and eliminate Lenin before his Russian rival gets back

even earlier and eliminates George Washington? In a few years the world may be anxiously waiting for the answer. If such spectacular folly once gets under way because governments have been convinced of some nonsensical theory, a logician will not waste effort on protests that will certainly go unheeded; he need not, after all, lose any sleep about who is going to succeed, and he could be glad that destructive efforts were directed where they would only squander human resources in a silly way.

One does what one can, though, against the Kingdom of Darkness; and perhaps less spectacular follies can be cured by exposing them to the light. Let me just instance a sophistry often used on one side of a current controversy. Some people are wont to say that it cannot make any significant moral difference whether you avoid something you wish to avoid by interposing a spatial barrier or by interposing a temporal barrier. If we do not let ourselves be fooled by the merely verbal assimilation of temporal and spatial barriers, the principle is really not a bit plausible; we need only test it on a case that rouses nobody's passions.

Let us suppose that it is my duty to organize a meeting in Cambridge. I fix a date for the meeting; then I suddenly realize that that ass Smith, whose presence would be disastrous, is coming to Cambridge for the day on that date, and will certainly attend given this opportunity. I may avoid this disaster either by changing the date of the meeting – ‘interposing a temporal barrier’ – or by locking Smith in his hotel room – ‘interposing a spatial barrier’. It really is not morally indifferent which of these methods I adopt.

When we find writers copying from one another the false moral principle I have just attacked – particularly when we find one of them supporting it with talk of ‘space-time’ – we may be pretty confident what the trouble is; here we have, to use Hobbes’s phrase, Darkness from Vain Philosophy. It is not for me here and now to enter upon a discussion ‘of the Benefit that proceedeth from such Darkness, and to whom it accrueth’.

Notes

- 1 *Word and Object* (Cambridge, Mass.: MIT Press, 1960), p. 172.
- 2 W. V. Quine, *Methods of Logic*, 4th edn (Cambridge, Mass.: Harvard University Press, 1982), p. 56.
- 3 ‘Mr Strawson on Logical Theory’, *Mind*, October 1953, p. 443. On the previous page of the same article, Quine had quoted the very passage from his own *Methods of Logic* that I quoted just now!
- 4 *The Nature of Existence*, vol. ii, ed. C. D. Broad (Cambridge: Cambridge University Press, 1927), sections 315–16.
- 5 *The Nature of Existence*, vol. ii, section 510.
- 6 *Ibid.*, section 857.
- 7 *Op. cit.*, section 511.
- 8 *The Nature of Existence*, vol. i, section 163. It is of no present concern that McTaggart chose to use the word ‘substance’ where I use ‘individual’. He was clearly assuming that the Christian era begins on 1 January AD 1, so that the nineteenth-century runs from 1 January 1801 to 31 December 1900.
- 9 Anyone disturbed by this sort of subject – predicate analysis may be reminded that

it has an Aristotelian precedent. Aristotle analyses ‘There is a single science of (a pair of) contraries’, into subject ‘(pair of) contraries’, predicate ‘there being a single science of them’; and he explains this as meaning, not that contraries *are* there being a single science of them, but that *it is true to say of them* that there is a single science of them (*Analytica Priora*, 48b 4 ff.).

- 10 Cf. *The Nature of Existence*, vol. i, sections 48–50.
 - 11 *Word and Object*, p. 173.
 - 12 *Logic and Knowledge* (London: Allen and Unwin, 1956), pp. 186–9.
 - 13 *Ibid.*, p. 274.
 - 14 R. A. Bull, ‘An Algebraic Study of Diodorean Modal Systems’.
-

24 The Problem of Temporary Intrinsics: an Excerpt from *On the Plurality of Worlds**

David Lewis

Let us say that something persists iff,¹ somehow or other, it exists at various times; this is the neutral word.² Something perdures iff it persists by having different temporal parts, or stages, at different times, though no one part of it is wholly present at more than one time; whereas it endures iff it persists by being wholly present at more than one time. Perdurance corresponds to the way a road persists through space; part of it is here and part of it is there, and no part is wholly present at two different places. Endurance corresponds to the way a universal, if there are such things, would be wholly present wherever and whenever it is instantiated. Endurance involves overlap: the content of two different times has the enduring thing as a common part. Perdurance does not.

(There might be mixed cases: entities that persist by having an enduring part and a perduring part. An example might be a person who consisted of an enduring entelechy ruling a perduring body; or an electron that had a universal of unit negative charge as a permanent part, but did not consist entirely of universals. But here I ignore the mixed cases. And when I speak of ordinary things as perduring, I shall ignore their enduring universals, if such there be.)

Discussions of endurance versus perdurance tend to be endarkened by people who say such things as this: ‘Of course you are wholly present at every moment of your life, except in case of amputation. For at every moment all your parts are there: your legs, your lips, your liver. . . .’ These endarkeners may think themselves partisans of endurance, but they are not. They are perforce neutral because they lack the conceptual resources to understand what is at issue. Their

* From David Lewis, *On the Plurality of Worlds* (Oxford: Blackwell, 1986). Reprinted by permission of the author.

speech betrays – and they may acknowledge it willingly – that they have no concept of a temporal part. (Or at any rate none that applies to a person, say, as opposed to a process or a stretch of time.) Therefore they are on neither side of a dispute about whether or not persisting things are divisible into temporal parts. They understand neither the affirmation nor the denial. They are like the people – fictional, I hope – who say that the whole of the long road is in their little village, for not one single lane of it is missing. Meaning less than others do by ‘part’, since they omit parts cut crosswise, they also mean less than others do by ‘whole’. They say the ‘whole’ road is in the village; by which they mean that every ‘part’ is; but by that, they only mean that every part cut lengthwise is. Divide the road into its least lengthwise parts; they cannot even raise the question whether those are in the village wholly or only partly. For that is a question about crosswise parts, and the concept of a crosswise part is what they lack. Perhaps ‘crosswise part’ really does sound to them like a blatant contradiction. Or perhaps it seems to them that they understand it, but the village philosophers have persuaded them that really they couldn’t, so their impression to the contrary must be an illusion. At any rate, I have the concept of a temporal part; and for some while I shall be addressing only those of you who share it.³

. . . The principal and decisive objection against endurance, as an account of the persistence of ordinary things such as people or puddles, is the problem of temporary intrinsics. Persisting things change their intrinsic properties. For instance shape: when I sit, I have a bent shape; when I stand, I have a straightened shape. Both shapes are temporary intrinsic properties; I have them only some of the time. How is such change possible? I know of only three solutions.

(It is not a solution just to say how very commonplace and indubitable it is that we have different shapes at different times. To say that is only to insist – rightly – that it must be possible somehow. Still less is it a solution to say it in jargon – as it might be, that bent-on-Monday and straight-on-Tuesday are compatible because they are ‘time-indexed properties’ – if that just means that, somehow, you can be bent on Monday and straight on Tuesday.)

First solution: contrary to what we might think, shapes are not genuine intrinsic properties. They are disguised relations, which an enduring thing may bear to times. One and the same enduring thing may bear the bent-shape relation to some times, and the straight-shape relation to others. In itself, considered apart from its relations to other things, it has no shape at all. And likewise for all other seeming temporary intrinsics; all of them must be reinterpreted as relations that something with an absolutely unchanging intrinsic nature bears to different times. The solution to the problem of temporary intrinsics is that there aren’t any temporary intrinsics. This is simply incredible, if we are speaking of the persistence of ordinary things. (It might do for the endurance of entelechies or universals.) If we know what shape is, we know that it is a property, not a relation.

Second solution: the only intrinsic properties of a thing are those it has at the present moment. Other times are like false stories; they are abstract representations, composed out of the materials of the present, which represent or misrep-

resent the way things are. When something has different intrinsic properties according to one of these ersatz other times, that does not mean that it, or any part of it, or anything else, just has them – no more so than when a man is crooked according to the Times, or honest according to the News. This is a solution that rejects endurance; because it rejects persistence altogether. And it is even less credible than the first solution. In saying that there are no other times, as opposed to false representations thereof, it goes against what we all believe. No man, unless it be at the moment of his execution, believes that he has no future; still less does anyone believe that he has no past.

Third solution: the different shapes, and the different temporary intrinsics generally, belong to different things. Endurance is to be rejected in favour of perdurance. We perdure; we are made up of temporal parts, and our temporary intrinsics are properties of these parts, wherein they differ one from another. There is no problem at all about how different things can differ in their intrinsic properties.

Notes

- 1 ‘Iff’ is short for ‘if and only if’. – [Eds]
- 2 My discussion of this problem is much indebted to David M. Armstrong, ‘Identity Through Time’, in *Time and Cause: Essays Presented to Richard Taylor*, ed. by Peter van Inwagen (Dordrecht: D. Reidel, 1980); and to Mark Johnston. I follow Johnston in terminology.
- 3 I attempt to explain it to others in *Philosophical Papers*, vol.1 (Oxford: Oxford University Press, 1983), pp. 76–7; reprinted in this volume as ‘In Defense of Stages: Postscript B to “Survival and Identity”’. But I have no great hopes, since any competent philosopher who does not understand something will take care not to understand anything else whereby it might be explained.

25 Temporary Intrinsics and Presentism*

Dean W. Zimmerman

David Lewis develops something like an antinomy concerning change which he calls “the problem of temporary intrinsics.” The resolution of this puzzle provides his primary motivation for the acceptance of a metaphysics of temporal parts.¹ Lewis’s own discussion is extremely compressed, showing up as a digression in a book about modality. So I shall set forth in some detail what I take to be his line of reasoning before suggesting that, at least for those philosophers who take seriously the distinction between past, present, and future, the argument poses no special threat.

The Structure of Lewis's Argument

Lewis's argument for temporal parts has the following structure. He offers reasons to deny that “the only intrinsic properties of a thing are those it has at the present moment”² – reasons, that is, for rejecting the “second solution” he considers. But if, in addition to the intrinsic properties I have now, I also have the intrinsic properties I have at other times, then I will end up having pairs like *being bent* and *being straight* – pairs that are, in some sense, incompatible. The challenge is then to answer the question: How can I have a pair of incompatible properties?³ Lewis thinks there are only two possible ways to answer this question. The first is unacceptable, and the second leads to the doctrine of temporal parts:

(1) My being both bent and straight is like my son’s being both tall and short – tall for a two-year-old, say, but short by comparison with most people. This strategy for dealing with apparent contradiction construes the seemingly incompatible properties as really relations to other things (in the case of tall and short, relations to different comparison classes). The version of this strategy that Lewis considers for temporary intrinsics is his “first solution”: that shapes and other seemingly intrinsic properties “are disguised relations, which an enduring thing may bear to times.”⁴ There is no more difficulty in standing in the *bent-at* relation to one time and the *straight-at* relation to another than there is in bearing the *tall-for-a* relation to two-year-olds and the *short-for-a* relation to the citizens of the United States as a whole. But Lewis doesn’t like this solution; he thinks it is tantamount to the rejection of intrinsic properties altogether.

(2) There’s only one way left, says Lewis, to make the apparent contradiction go away while retaining the incompatibility of *being bent* and *being straight*; and that is to treat it as we do the case of the road that is both bumpy and smooth. How can a road be both? Easily: by having one part that is bumpy and another that is smooth. So, analogously, the only way for me to be both bent and straight is for me to have a part that is bent and a part that is straight. But these cannot be ordinary spatial parts of me, like an arm or a hand. The bent “part” of me is exactly my size and shape, with arms, legs, torso, and head; and likewise for the straight “part” of me. And, like the different spatial parts of the road, these different parts of me must be distinct one from another. So I emerge as a whole spread out along the temporal dimension with different (temporal) parts for the different times I occupy, much as the road is a whole spread out along the spatial dimension with different (spatial) parts for the different places it occupies.

I am willing to grant Lewis’s assertion that, once someone admits that I have more properties than just those I have now, she must choose between alternatives (1) and (2). And perhaps it is true that (1) eliminates temporary intrinsics altogether. At the very least, it eliminates temporary *monadic* properties (“one-

place” properties, properties that are not relations); and it’s easy to see why someone might think that *really* intrinsic properties should be monadic.⁵ What I want to question instead is the very first move: Why suppose that I must have more than just the properties I have now?

Serious Tensors and Presentists

Before looking at Lewis’s answer, I want to make clear what view Lewis is targeting: namely, “presentism.” A closely related position is that of one who “takes tense seriously.” As shall appear, one can’t very well be a presentist without taking tense seriously, although it is possible to do the reverse.

When a philosopher says, “The only properties I have are those I have now,” it is tempting to respond by saying: This thesis is either an uninteresting, tautologous truth; or an obvious falsehood. If the first occurrence of “have” is in the present tense, then the assertion is equivalent to “The only properties I have now are those I have now.” Who could disagree? But how dull! On the other hand, suppose this “have” is an instance of what philosophers sometimes call a “tenseless” verb. To say that I (tenselessly) have some property, for instance that I (tenselessly) am straight, is to say something more or less equivalent to this: “Either I was straight, or I am straight, or I will be straight.” But “The only properties I (tenselessly) have are those I have now” is true only if either I never change or I exist for but an instant. Taken, then, in the only way in which it can be true (i.e., with the first “have” in the present tense), the claim seems too trivial to be the focus of a substantive philosophical debate.

I am convinced that there *is* an important disagreement between those who take tense seriously and those who don’t. Precisely what the disagreement boils down to will depend to some extent upon metaphysical theses about what kinds of things are, in the first instance, true and false. Here is one example; but I believe that nothing much hinges on accepting just this view about the most fundamental bearers of truth. Suppose you think that the sentences we write down and utter are true or false in virtue of their expressing *propositions* that are true or false in some more basic sense. A proposition is something that can be expressed in many different ways; it can be believed by one person and disbelieved by another; and, at least in the case of a proposition that isn’t about a particular sentence or thought, it would have existed and been either true or false even in the absence of all sentences or thoughts. This familiar conception of the ultimate bearers of truth and falsehood⁶ can be conjoined with a tensed or a tenseless theory about the nature of the proposition. On a tensed construal, a proposition’s being true is not typically a once-and-for-all thing. The sentence “I am bent” could now be used by me to express a true proposition; but the proposition in question hasn’t always been true, and it won’t continue to be true for very long. A tenseless account of propositions, on the other hand, takes them to be like statements made using tenseless verbs: each is either always true, or never true.

The competition between the tensed and tenseless approaches to the funda-

mental bearers of truth gives rise to a familiar dispute over the importance of “tense logic.” Logic is all about describing the most general patterns of truth-preserving inference. If the things that are true and false can be true but *have been false*, or *be about to become false*, then some of the patterns of inference logicians should be interested in will involve temporal notions. On the tensed conception of truths, it is a question of logic whether, for example, the proposition: It will be the case that I am bent, implies the proposition: It was the case that it will be the case that I am bent. Thus relations like being true simultaneously, and being true earlier or later than, will turn out to be, at least in part, logical notions.⁷ On the other hand, those who take truth-bearers to correspond to tenseless statements will regard this as a blunder: temporal relations are for science and (perhaps) metaphysics to explore; but they are not part of the subject matter of logic.⁸

The philosopher who takes a tensed approach to the bearers of truth regards each of them as making a claim about what is the case *now*. Of course some propositions are eternally true: in other words, there are propositions which, either necessarily, or as a matter of contingent fact, have always been true and will always be true. That two and two make four is an example of the first sort. And historical propositions expressed by tenseless statements, such as my utterance in a lecture of “Plato believes in universals,” are examples of the latter sort. But the proponent of tensed truth-bearers will insist that the true proposition expressed is composed of tensed propositions; it’s a disjunction of three propositions: Either Plato (now) believes in universals, or he did, or he will.⁹ This is a truth, but it is made out of three other propositions, only one of which is true, and each of which concerns what is now the case.

I shall call a philosopher who takes this sort of position a “serious tenser.” Those who do not take tense seriously include D. H. Mellor, J. J. C. Smart, and others who defend what is sometimes called the “new tenseless theory of time”: the thesis that the meaning of every tensed utterance can be captured by stating, in a tenseless language of eternal truths, the conditions in which the utterance would be true.¹⁰ It should be evident that there can be a real dispute between the serious tenser and someone who rejects this view.¹¹

Many serious tensers are also *presentists*. The presentist says: “The only things that exist are those that exist at present.” The “once was” no longer exists and the “will be” doesn’t exist yet. But the proponents of presentism are also confronted with a skeptical challenge to the significance of their thesis. Is the first occurrence of “exist” in the presentist’s assertion a tensed one? Then the presentist is simply making a fuss over a pointless tautology: “The only things that exist now (i.e. at present) are those that exist at present.” Who denies this? Or is “exist” here a tenseless verb, equivalent to “existed or exists now or will exist”? But then it’s an implausible metaphysical thesis: the claim that everything exists at all times, that nothing can have a less than eternal history. So presentism is either a boring truth or an interesting falsehood.

Presentism is neither; it is a substantive thesis, and one that is not equivalent to the claim that everything exists eternally. Just as the serious tenser thinks there is, at bottom, only one kind of truth, and that is “truth-now”; so the

presentist thinks there is only one largest class of all real things, and this class contains nothing that lies wholly in the past or future.¹² Presentism is, in fact, a thesis about the range of things to which one should be “ontologically committed.”

Philosophers are always looking out for the ontological commitments of their views – where someone is ontologically committed to a certain kind of thing just in case something she believes implies that something of that kind exists. There are many perfectly sensible truths which, on the surface, seem to require the existence of highly problematic entities – *entia non grata*, as it were. Consider, for example, the following:

- (1) Jeeves was nonplussed by the dearth of champagne in the ice box.
- (2) Moriarty is the most well-known criminal in detective fiction.
- (3) Courage is a virtue displayed by many people.
- (4) There could have been a person who is not one of those who actually exist.

On the face of it, these are statements about such things as dearths, fictional characters, characteristics that may be possessed by many people, and merely possible persons. One might think that it could be inferred from them that: there is at least one dearth, there are some fictional criminals, there is something displayed by every courageous person, there are merely possible people. But each of these statements can seem hard to swallow for one reason or another:

A dearth of champagne isn’t a kind of *thing*, a sort of invisible anti-champagne located where the champagne should be. To say that Jeeves was nonplussed by the dearth of champagne is simply to say that there was no champagne in the ice box, and that he was taken aback by the situation.

Nor are there some criminals (among the least dangerous of criminals) who are fictional. Fictional characters are not an odd group of people who, for some reason, we cannot meet in the way we meet other people, but can only get to know through stories. Statements about, say, Moriarty must really be elliptical for claims about the stories Arthur Conan Doyle wrote that had the name “Moriarty” in them.¹³

It might seem less problematic to suppose that there are some things called “virtues,” of which courage is one. But if courage is something that can be displayed (or possessed or exemplified) by many different people at once, then some puzzling questions immediately arise. For how could anything be displayed in many different places at once, except by having a part displayed in each of those places and only there? Those philosophers particularly perplexed by this question (called “nominalists”) claim that (3) doesn’t imply that there is one thing possessed by all the courageous people. Some nominalists would say that each of the courageous people has his or her own particular instance of courageousness (in D. C. Williams’s terminology, a courageousness “trope”¹⁴), and that statements about courage are really about the big group or heap of all these instances.

For present purposes, the final case is the most illuminating. Do we really want to say that there are some merely possible people? That some people are tall, some are short, and some are nonexistent – the limiting case of diminutive stature, as it were? Philosophers who answer, No, are called “modal actualists”: they hold that there are no nonactual things. But then how to make sense of (4)? One strategy is to posit individual essences for nonexistent individuals, and then construe talk about nonactuals as really talk about these essences. Then (4) becomes the claim that there is an unexemplified individual essence that would be the essence of a person if it were exemplified.¹⁵ Another is to say that what (4) really comes to is the claim that it is possible that there be something that is a person and is not identical with Jones, Robinson, . . . or any of the other actual people. This is an assertion about the possible truth of a certain proposition (that there be something that is a person and is not identical with . . .); the proposition itself isn’t about any particular nonactual thing; and it is not equivalent to the claim that there is something that is a possible person and is not identical with. . . .¹⁶

These are some typical attempts to avoid ontological commitment to undesirable entities. Statements which, on the surface, seem to imply that there are certain problematic entities, are given philosophical glosses or paraphrases which seem to capture the truth in question while avoiding the implication that the troublesome things exist. The presentist is engaged in precisely the same sort of enterprise. But the truths that bother her are of this sort:

- (5) There was a person who is not one of the people who presently exist.
- (6) There will be a person who is not one of the people who presently exist.

The presentist is a “temporal actualist” – she is troubled by the fact that (5) and (6) seem to imply that there are some people who do not now exist, just as the modal actualist is bothered by the fact that (4) seems to imply that there are some people who do not actually exist. How can there *be* something that no longer exists, or that hasn’t existed yet, she wonders? And so the presentist tries to show that the truth of (5) and (6) doesn’t really conflict with her thesis that no nonpresent things exist.

One way of trying to show this would be to make use of individual essences again: (5) becomes the proposition that there is an individual essence not now exemplified that was once exemplified, and was then the essence of a person; and analogously for (6). Another is to insist that the truth of (5) implies only that it was the case that there is someone not identical with Jones, Robinson, . . . or any other presently existing person; but not that there is someone who used to exist and is not identical with Jones, Robinson, And likewise for (6).¹⁷

How is presentism related to taking tense seriously? The presentist must, I think, be a serious tenser. At the very least, tenseless statements that ostensibly require ontological commitment to past and future things must be treated as equivalent to tensed truths that do not. And the presentist could not very well regard all the fundamental truth-bearers as eternally true, corresponding to

tenseless statements. For, she says, one of the truths is that wholly future things, like my first grandchild, do not exist – and such truths had better be susceptible to change. On the other hand, the serious tenser need not be a presentist. Quentin Smith, for example, is a non-presentist serious tenser.¹⁸ According to Smith, fundamental truths are all tensed; but past and future individuals and events, although no longer present, nonetheless exist. Ostensible ontological commitment to such things cannot, on Smith's view, be paraphrased away.

But the combination of rejecting presentism while taking tense seriously is an unstable one. For the primary motivation for treating the fundamental truth-bearers as mutable and true *now* is the desire to do justice to the feeling that what's in the past is over and done with, and that what's in the future only matters because it will eventually be present. This is the source of the importance Prior attaches to the exclamation “Thank goodness that's over!”¹⁹ If yesterday's headache still exists, and remains as painful as ever, then why should I be relieved now? Would the mere fact that it's no longer present justify this attitude? Most serious tensers, including Smith himself, will agree that it would not. And so, to render reasonable our special concern for the present, Smith strips past and future events of all their interesting intrinsic properties. For instance, yesterday's headache, although it exists, is no longer painful. It has a past-oriented property, *having been painful* – a sort of backwards-looking relation to the property *being painful*. But it is not painful now, and that's why it no longer concerns us.²⁰

Although this view makes sense of our relief when pain is past, I find it unappealing in the extreme. The past and future events and objects it posits are too ghostly to be real. A painful headache cannot exist without being painful; a tanker explosion cannot exist without being violent and loud; Plato cannot exist while having neither body nor soul. What's left of these past and future things and events is too thin: yesterday's headache is still an event, but it isn't painful or throbbing or much of anything else; Plato is still a substance, I suppose, but he doesn't talk or think or walk or sleep or have any spatial location. Neither Plato nor headache has any of the ordinary intrinsic properties it displayed while present. Smith's efforts to preserve the intuition behind “Thank goodness that's over!” while rejecting presentism are, I judge, unsuccessful. Past and future things become nearly-bare particulars, unreal echoes of their once or future selves. The serious tenser is much better off without them.

Why does Lewis Reject Presentism?

The serious tenser says that it is simply not true that I have the property *being straight* if I am bent now. I was straight, and will be again; but I am not now, and so there is no problem of my having incompatible intrinsic properties. Of course philosophers are free to invent a tenseless language in which “I am straight” is true just in case I either am now or was or will be straight. Who can stop philosophers from inventing peculiar ways of speaking? But the bare fact

that one can talk this way doesn't create any problem about my having incompatible properties.

What is Lewis's response to this serious-tenser solution of the problem of temporary intrinsics? He seems to suppose (reasonably, I think) that someone who takes this line must be a presentist. But, by Lewis's lights, presentism is too incredible to be believed. Presentism "rejects endurance; because it rejects persistence altogether"; and it "goes against what we all believe" by implying that "there are no other times." "No man, unless it be at the moment of his execution, believes that he has no future; still less does anyone believe that he has no past." And yet, says Lewis, the presentist denies these obvious facts.²¹

This string of claims represents what might be called the "no persistence objection" to presentism. Lewis takes it that the following thesis of "Persistence through Change" is obviously true:

- (PC) There are (at least) two different times; one at which I am bent, another at which I am straight.²²

Lewis thinks that (PC) is a simple expression of my belief that I persist through changes in my posture: there are times when I'm bent and times when I'm straight. The presentist is committed to the nonexistence of all times but one, the present. (PC) says there is more than one time; so presentism and (PC) are incompatible.

The serious-tenser dissolution of the problem of temporary intrinsics given at the beginning of this section does not require the truth of presentism; a non-presentist serious tenser like Smith has little to fear from Lewis's argument. But Smith's combination of views has turned out to be unacceptable; and so the tensed response to the problem of temporary intrinsics stands or falls with presentism.

In order for Lewis's argument to have any teeth, (PC) must have two features: (i) it must be something we all, on reflection, believe; and (ii) it must require ontological commitment to the existence of more than one time. To be something commonly believed, (PC) must correspond to the humdrum assertion that I am bent at some times and straight at others. The question is whether this belief in my persistence through change – and the similar belief had by anyone who can remember changing posture – implies that there exist more times than the present.

If the statements used to express ordinary beliefs could be counted on to wear their ontological commitments on their sleeves, then an affirmative answer would be justified. But virtually everyone must allow that many statements expressing commonsensical beliefs do *not* wear their ontological commitments on their sleeves. It would be just like Bertie Wooster to respond to Jeeves's report about the dearth of champagne in the ice box by saying: "Well, at least there's *something* in the ice box." The source of the joke here would be that, generally speaking, from the fact that there's a such-and-such in the ice box, it follows that there is something in the ice box. But when the "such-and-such" is a *dearth* of something, it doesn't follow. Why? Because the assertion that there's

a dearth of something is just a fancy (and old-fashioned) way of saying that there isn't any of that something – and that's compatible with there being *nothing at all* in the ice box.

Compare (PC) with a precisely parallel case involving ontological commitment to nonactuals. I suppose that most of us believe that we could have been put in situations that would have resulted in our lives going differently than they have in fact gone. There are certain possible experiences and events which, had they happened, would have prevented me from becoming a philosopher. But does this statement commit me to the existence of nonactual experiences and events? I should think not.

A few people have believed in the existence of alternative universes, just as real and concrete as this one, but with things going differently in them – worlds in which, for instance, the Axis powers win World War II, and the US is partitioned between Germany and Japan.²³ David Lewis, in fact, believes in the literal existence of alternative universes, just as concrete as our actual world, in which every possible way things could go actually plays itself out.²⁴ But Lewis is one of the exceptions that prove the rule. The rest of us cannot bring ourselves to believe that there is such an event as the Axis powers' winning the war, an event with which, fortunately, we are not space-time neighbors. It's not that we ordinarily ignore these nonactual events because they are "far" from us, unreachable from our world. Rather, we think they simply are not.

How do we know that we aren't, implicitly, committed to the existence of such merely possible events? Well, we just ask ourselves whether we think they exist – whether we think that there are such things, whether we think we stand in real relationships with them. The answer comes back a resounding No. And then, if we are philosophers, we go about the business of finding plausible paraphrases for our beliefs ostensibly about nonactual possibilities – paraphrases that seem to us to capture more or less what we believed all along, but which do not even appear to imply that there are situations involving me that don't occur, or whole worlds full of people and events that are not actual. If it were to become clear that there is no way to do this, then perhaps we would feel forced to reconsider our judgment that our beliefs about alternate possibilities do not implicitly commit us to the existence of such things. But that's not usually the way things go in philosophy: there's usually more than one way to skin a philosophical cat; usually several competing approaches to a given philosophical problem emerge as favorites, with much to be said for and against each of them. And so it is here: there are ever so many fairly plausible projects under way for paraphrasing away ostensible commitment to nonactual things and situations, each with its own advantages and disadvantages, and few confront such grave obstacles as to suggest that they are absolute dead-ends.

The presentist believes that the situation is precisely parallel when it comes to my belief that there are times at which I'm bent and times at which I'm straight. Does this commit me to the existence of times other than the present? Well, when I ask myself whether I think that my childhood exists, or the time of my death, the snows of yesteryear, or the light of other days, the answer comes back a resounding No. Is it just that I feel that past and future things and events can

be regarded as nonexistent because they are “temporally far” from me? I think not – the past is no more, and the future is not yet, in the strictest sense. And so those who share this judgment begin the work of philosophical paraphrase, trying to find plausible construals of statements like (5), (6), and (PC) that capture what is meant but do not involve direct reference to nonpresent times, individuals, and events. So, for instance, (PC) can be taken as a tenseless statement expressing a disjunction of tensed propositions: Either I was bent and would become or had previously been straight, or I was straight and would become or had previously been bent, or I will be bent and will have been or be about to become straight, or I will be straight and will have been or be about to become bent. Surely this tensed disjunction is true if (PC) is true; furthermore, it contains no mention of anything like a nonpresent time. So, given the presentist’s desire to avoid ontological commitment to nonpresent times, this tensed statement provides a perfectly sensible paraphrase of my conviction that I can persist through change of shape.

Furthermore, it is not as if Lewis himself allows (PC) to stand as it is, with no paraphrastic gloss. After all, he thinks that I am bent at one time and straight at another only in virtue of the fact that I have temporal parts located at these times, one of which is bent, the other straight. So “there is a time at which I am bent,” as it occurs in (PC), receives the paraphrase “there is a time at which I have a temporal part that is bent.” Lewis salvages our common conviction that we persist through change by introducing the uncommon notion of a temporal part. But if his temporal-parts reading of (PC) captures enough of our pretheoretical convictions to be acceptable, then surely he must allow the presentist similar leeway in her attempt to affirm persistence through change while avoiding talk of nonpresent times.²⁵

The large-scale project of paraphrasing truths ostensibly about nonpresent times and things is as complex and difficult as the counterpart project concerning nonactUALS. Ways must be found to capture all truths about past and future things without the appearance of ontological commitment to such things.²⁶ Presentists must, for example, find a way to understand statements ostensibly about relations that hold between presently existing things and things in the past and future. Causation is one instance of this problem: the causal relation holds between events; but no relation can hold between a present event and some future or past event, since such events do not exist. Must the presentist then conclude that no event in the present can be caused by anything earlier, or cause anything later?²⁷ Such difficulties must be overcome for presentism to remain plausible.

And there are familiar chestnuts bedeviling *anyone* (presentist or not) who takes tense seriously, such as McTaggart’s paradox²⁸ and the puzzle about the rate at which the present “moves.” Is this rate one minute per minute? It couldn’t very well move any faster! And yet this doesn’t sound like a proper rate at all.²⁹ Perhaps most worrisome is that positing facts about what is present *absolutely* (and not merely about what is “present relative to me” or “present relative to my inertial frame”) seems inconsistent with a well-confirmed scientific theory: special relativity.³⁰ But, as indicated in the notes to this and the previous

paragraph, these are problems which presentists and others who take tense seriously have tried to address. Have the solutions been satisfactory? Perhaps not in every case. But rejecting presentism on the basis of such problems would require careful exploration of these debates – debates which have nothing to do with the problem of temporary intrinsics *per se*. Furthermore, there is reason to be hopeful that they will be resolved in the presentist's favor – or at least that they will not be resolved decisively in favor of her opponents. After all, as John Bigelow points out, presentism was accepted everywhere by nearly everyone until a mere hundred or so years ago.³¹ A thesis with a track record like that shouldn't be expected to go down without a fight.

So far as I know, all presentists (and almost all who take tense seriously) reject the doctrine of temporal parts; indeed, Prior, Geach, and Chisholm have been among its most vocal opponents.³² What I have tried to show is that the part of Lewis's argument aimed at these philosophers requires considerable buttressing before it will convince. In particular, we need a reason to think that some truths ostensibly about nonpresent things cannot be given plausible paraphrases that eschew commitment to such things. So far as I can see, there isn't any reason to think this is so. At any rate, Lewis hasn't (yet) given us one.

Notes

* A very distant ancestor of this paper was presented at meetings of the Central States Philosophical Association in 1990. I am grateful to my commentator on that occasion, Mark Heller, for his excellent criticisms and suggestions (which formed the basis of his paper, "Things Change," *Philosophy and Phenomenological Research*, 52 (1992), pp. 695–704); and to Roderick Chisholm, who was also present. Later on, Trenton Merricks (who, for no good reason, was *not* present at the talk) provided useful comments as well. Recent work that should be consulted includes: Merricks, "Endurance and Indiscernibility," *Journal of Philosophy*, 91 (1994), pp. 165–84; Sally Haslanger, "Endurance and Temporary Intrinsics," *Analysis*, 49 (1989), pp. 119–25; Peter M. Simons, "On Being Spread Out in Time: Temporal Parts and the Problem of Change," in *Existence and Explanation*, ed. by Wolfgang Spohn, Bas C. van Fraassen, and Brian Skyrms (Dordrecht: Kluwer, 1991), pp. 131–47; and Mark Hinchliff, "The Puzzle of Change," *Philosophical Perspectives*, vol. 10: *Metaphysics*, ed. by James E. Tomberlin (Oxford: Blackwell, 1996), pp. 119–36. I am extremely grateful to David Lewis, who provided extensive criticism of a late draft, and saved me from a number of serious mistakes. But it should not be assumed that he agrees with anything that I now say about his argument from temporary intrinsics.

- 1 Cf. David Lewis, *On the Plurality of Worlds* (Oxford: Blackwell, 1986), pp. 202–3; the relevant passage is reprinted in this volume as "The Problem of Temporary Intrinsics: an Excerpt from *On the Plurality of Worlds*." Page references are to the excerpt in this volume.
- 2 Cf. Lewis, "The Problem of Temporary Intrinsics," p. 205.
- 3 It might be replied that there is no problem with having both if the verb "having" is taken *tenselessly* (that is, in such a way that "I have both" is equivalent to something like: "I had, now have, or will have the one; and I had, now have, or will have the other"). But then we should want to know why these properties deserve the

- label “incompatible.” How do they differ from a pair of compatible intrinsics, like *being red* and *being round*?⁴
- 4 Lewis, “The Problem of Temporary Intrinsics,” p. 205.
- 5 One might, however, allow intrinsic properties to be monadic but treat the *having* of them as a relation between a thing, a property, and a time. See, for example, Peter van Inwagen, “Four-dimensional Objects,” *Noûs*, 24, and Sally Haslanger, “Endurance and Temporary Intrinsics.”⁵
- 6 It can be found in Bolzano, Frege, Church, Chisholm, and Plantinga, to name but a few.
- 7 This point of view is exemplified by A. N. Prior, “The Notion of the Present” and “Some Free Thinking about Time,” and Peter Geach, “Some Problems about Time,” all reprinted in this volume.
- 8 See Gerald J. Massey, “Tense Logic! Why Bother?,” *Noûs*, 3 (1969), pp. 17–32.
- 9 If the tenseless verb used in my lecture were the ordinary historical present tense, the proposition in question would lack the final conjunct; only the more arcane tenseless verb introduced by philosophers is used to express disjunctive propositions with disjuncts concerning the future.
- 10 Mellor, *Real Time* (Cambridge: Cambridge University Press, 1981), ch. 5; J. J. C. Smart, “Time and Becoming,” reprinted in his *Essays Metaphysical and Moral* (Oxford: Blackwell, 1987). For further defenders of the new tenseless theory, see the contributions of L. Nathan Oaklander, Michelle Beer, and Clifford Williams to Part I of *The New Theory of Time*, ed. by Oaklander and Quentin Smith (New Haven, Conn.: Yale University Press, 1994); for criticisms, see Quentin Smith’s contributions to Part I of the same volume, and his *Language and Time* (Oxford: Oxford University Press, 1993).
- 11 Some have held that there are both tenselessly true propositions and “tensedly true” propositions, and that the former are not equivalent to disjunctions of the latter. But this is a minority opinion; most who take at least *some* fundamental truth-bearers to be mutable regard all truth as tensed truth. For some reasons to do so, see Roderick M. Chisholm and Dean W. Zimmerman, “Tense and Theology,” *Noûs*, 31 (1997), pp. 262–5.
- 12 For a paradigmatic statement of this position, see Prior, “The Notion of the Present,” reprinted in this volume.
- 13 On some problems for carrying out this project, see Peter van Inwagen, “Creatures of Fiction,” *American Philosophical Quarterly*, 14 (1977), pp. 299–308; “Fiction and Metaphysics,” *Philosophy and Literature*, 7 (1983), pp. 67–77; and “Pretense and Paraphrase,” in Peter J. McCormick, ed., *The Reasons of Art* (Ottawa: University of Ottawa Press, 1985), pp. 414–22.
- 14 See Williams, “The Elements of Being,” reprinted in this volume.
- 15 Compare Alvin Plantinga, “Actualism and Possible Worlds,” *Theoria*, 42 (1976), pp. 139–60; reprinted in *The Possible and the Actual*, ed. by Michael J. Loux (Ithaca, N.Y.: Cornell University Press, 1979).
- 16 See Prior, *Papers on Time and Tense* (Oxford: Clarendon Press, 1968), pp. 142–3; and Kit Fine’s Postscript to Prior and Fine, *Worlds, Times and Selves* (London: Duckworth, 1977).
- 17 For a general discussion of the treatment of past and future individuals in tense logic, see Prior, *Past, Present and Future* (Oxford: Clarendon Press, 1967), ch. 8. See also Chisholm, “Referring to Things That No Longer Exist,” *Philosophical Perspectives*, vol. 4: *Action Theory and Philosophy of Mind* (1990), pp. 545–56.

- 18 See Quentin Smith, *Language and Time* (New York: Oxford University Press, 1993), see esp. ch. 5.
- 19 See Prior, “Some Free Thinking about Time,” reprinted in this volume; and his “Thank Goodness That’s Over,” reprinted in Prior, *Papers in Logic and Ethics* (London: Duckworth, 1976).
- 20 Incidentally, Smith’s approach to past and future events and things provides him with the means to define “being present” – something he claims cannot be done. Just take all the kinds of intrinsic properties which a contingent thing cannot have when it is wholly past or future; and then say a thing is present just in case either it is a necessary thing (and so must always be present), or it is a contingent thing that has properties belonging to this special class.
- 21 Lewis, “The Problem of Temporary Intrinsics,” p. 206.
- 22 This thesis, and its name, are taken from personal correspondence with Lewis, and used with his permission.
- 23 Such is the world of the “alternate history” novel, *The Man in the High Castle*, by Philip K. Dick. Dick became convinced that such alternate streams of history are not mere fictions, but that they are real; he claimed to have been able to “recall” events from lives lived in other worlds.
- 24 For Lewis’s reasons for believing in concrete worlds besides this one, see his *On the Plurality of Worlds* (for the senses in which his worlds are *concrete*, see section 1.7 of the book). I should point out that, unlike Dick, Lewis’s reasons are purely theoretical and *a priori*, not empirical.
- 25 I owe this point to Trenton Merricks.
- 26 For some presentist responses to the problem, cf. Prior, *Past, Present and Future*, ch. 8 (“Time and Existence”); Prior, *Papers on Time and Tense*, ch. 8 (“Time, Existence, and Identity”); R. M. Chisholm, “Referring to Things That No Longer Exist”; and John Bigelow, “Presentism and Properties,” in *Philosophical Perspectives*, vol. 10: *Metaphysics*, pp. 35–52.
- 27 John Bigelow and I have offered, independently, very similar solutions to this problem. Cf. the final section of my “Chisholm and the Essences of Events,” in *The Philosophy of Roderick M. Chisholm* (The Library of Living Philosophers), ed. by Lewis Hahn (La Salle, Ill.: Open Court, 1997); and Bigelow’s “Presentism and Properties,” p. 47.
- 28 Cf. McTaggart, “Time: An Excerpt from *The Nature of Existence*,” this volume; and, for a response which, by my lights, completely dissolves this “paradox,” cf. Reading 6: C. D. Broad, “McTaggart’s Arguments against the Reality of Time,” also in this volume.
- 29 J. J. C. Smart raises the puzzle about the rate of passage in “The Space–Time World,” this volume. Presentists can hope that Ned Markosian has settled the problem for good and all in his “How Fast Does Time Pass?,” *Philosophy and Phenomenological Research*, 53 (1993), pp. 829–44.
- 30 Prior’s description of the problem and his response may be found in “Some Free Thinking about Time,” reprinted in this volume. Cf. also Geach, “Some Problems about Time,” this volume. More recent treatments may be found in Quentin Smith, *Language and Time*, ch. 7. For one scientist who thinks that Prior may have been right about the prematurity of giving up on the notion of absolute simultaneity, cf. J. S. Bell, *Speakable and Unspeakable in Quantum Mechanics* (Cambridge: Cambridge University Press, 1987), p. 77; and Bell’s remarks in *The Ghost in the Atom*, ed. by P. C. W. Davies and J. R. Brown (Cambridge: Cambridge University Press, 1986), esp. pp. 48–51.

- 31 Bigelow, "Presentism and Properties," pp. 35–6.
- 32 For characteristic rejections of temporal parts, cf. Prior, "Some Free Thinking about Time"; Geach, "Some Problems about Time"; and Chisholm, *Person and Object: A Metaphysical Study* (La Salle, Ill.: Open Court, 1976), Appendix A, pp. 138–44. For a serious tenser who accepts temporal parts, see Quentin Smith, "Personal Identity and Time," *Philosophia*, 22 (1993), pp. 155–67.

How do Causes Bring about their Effects?

26 Constant Conjunction: an Excerpt from *A Treatise of Human Nature**

David Hume

To begin regularly, we must consider the idea of *causation*, and see from what origin it is deriv'd. 'Tis impossible to reason justly, without understanding perfectly the idea concerning which we reason; and 'tis impossible perfectly to understand any idea, without tracing it up to its origin, and examining that primary impression, from which it arises. The examination of the impression bestows a clearness on the idea; and the examination of the idea bestows a like clearness on all our reasoning.

Let us therefore cast our eye on any two objects, which we call cause and effect, and turn them on all sides, in order to find that impression, which produces an idea of such prodigious consequence. At first sight I perceive, that I must not search for it in any of the particular *qualities* of the objects; since, which-ever of these qualities I pitch on, I find some object, that is not possest of it, and yet falls under the denomination of cause or effect. And indeed there is nothing existent, either externally or internally, which is not to be consider'd either as a cause or an effect; tho' 'tis plain there is no one quality, which universally belongs to all beings, and gives them a title to that denomination.

The idea, then, of causation must be deriv'd from some *relation* among objects; and that relation we must now endeavour to discover. I find in the first place, that what-ever objects are consider'd as causes or effects, are *contiguous*; and that nothing can operate in a time or place, which is ever so little remov'd from those of its existence. Tho' distant objects may sometimes seem productive of each other, they are commonly found upon examination to be link'd by a chain of causes, which are contiguous among themselves, and to the distant objects; and when in any particular instance we cannot discover this connexion, we still presume it to exist. We may therefore consider the relation of *CONTIGUITY* as essential to that of causation; at least may suppose it such, according to the general opinion, till we can find a more¹ proper occasion to clear up this matter, by examining what objects are or are not susceptible of juxtaposition and conjunction.

* From David Hume, *A Treatise of Human Nature*, Book I, Part III, Sections ii and xiv; first published in 1739.

The second relation I shall observe as essential to causes and effects, is not so universally acknowledg'd, but is liable to some controversy. That of PRIORITY of time in the cause before the effect. Some pretend that 'tis not absolutely necessary a cause shou'd precede its effect; but that any object or action, in the very first moment of its existence, may exert its productive quality, and give rise to another object or action, perfectly co-temporary with itself. But beside that experience in most instances seems to contradict this opinion, we may establish the relation of priority by a kind of inference or reasoning. 'Tis an establish'd maxim both in natural and moral philosophy, that an object, which exists for any time in its full perfection without producing another, is not its sole cause; but is assisted by some other principle, which pushes it from its state of inactivity, and makes it exert that energy, of which it was secretly possest. Now if any cause may be perfectly co-temporary with its effect, 'tis certain, according to this maxim, that they must all of them be so; since any one of them, which retards its operation for a single moment, exerts not itself at that very individual time, in which it might have operated; and therefore is no proper cause. The consequence of this wou'd be no less than the destruction of that succession of causes, which we observe, in the world; and indeed, the utter annihilation of time. For if one cause were co-temporary with its effect, and this effect with its effect, and so on, 'tis plain there wou'd be no such thing as succession, and all objects must be co-existent.

If this argument appear satisfactory, 'tis well. If not, I beg the reader to allow me the same liberty, which I have us'd in the preceding case, of supposing it such. For he shall find, that the affair is of no great importance.

Having thus discover'd or suppos'd the two relations of *contiguity* and *succession* to be essential to causes and effects, I find I am stopt short, and can proceed no farther in considering any single instance of cause and effect. Motion in one body is regarded upon impulse as the cause of motion in another. When we consider these objects with the utmost attention, we find only that the one body approaches the other; and that the motion of it precedes that of the other, but without any sensible interval. 'Tis in vain to rack ourselves with *farther* thought and reflexion upon this subject. We can go no *farther* in considering this particular instance.

Shou'd any one leave this instance, and pretend to define a cause, by saying it is something productive of another, 'tis evident he wou'd say nothing. For what does he mean by *production*? Can he give any definition of it, that will not be the same with that of causation? If he can; I desire it may be produc'd. If he cannot; he here runs in a circle, and gives a synonymous term instead of a definition.

Shall we then rest contented with these two relations of contiguity and succession, as affording a compleat idea of causation? By no means. An object may be contiguous and prior to another, without being consider'd as its cause. There is a NECESSARY CONNEXION to be taken into consideration; and that relation is of much greater importance, than any of the other two above-mention'd.

Here again I turn the object on all sides, in order to discover the nature of this necessary connexion, and find the impression, or impressions, from which its idea may be deriv'd. When I cast my eye on the *known qualities* of objects, I

immediately discover that the relation of cause and effect depends not in the least on *them*. When I consider their *relations*, I can find none but those of contiguity and succession; which I have already regarded as imperfect and unsatisfactory. Shall the despair of success make me assert, that I am here possesst of an idea, which is not preceded by any similar impression? This wou'd be too strong a proof of levity and inconstancy; since the contrary principle has been already so firmly establish'd, as to admit of no farther doubt; at least, till we have more fully examin'd the present difficulty.

We must, therefore, proceed like those, who being in search of any thing, that lies conceal'd from them, and not finding it in the place they expected, beat about all the neighbouring fields, without any certain view or design, in hopes their good fortune will at last guide them to what they search for. . . .

Suppose two objects to be presented to us, of which the one is the cause and the other the effect; 'tis plain, that from the simple consideration of one, or both these objects we never shall perceive the tie, by which they are united, or be able certainly to pronounce, that there is a connexion betwixt them. 'Tis not, therefore, from any one instance, that we arrive at the idea of cause and effect, of a necessary connexion of power, of force, of energy, and of efficacy. Did we never see any but particular conjunctions of objects, entirely different from each other, we shou'd never be able to form any such ideas.

But again; suppose we observe several instances, in which the same objects are always conjoin'd together, we immediately conceive a connexion betwixt them, and begin to draw an inference from one to another. This multiplicity of resembling instances, therefore, constitutes the very essence of power or connexion, and is the source, from which the idea of it arises. In order, then, to understand the idea of power, we must consider that multiplicity; nor do I ask more to give a solution of that difficulty, which has so long perplex'd us. For thus I reason. The repetition of perfectly similar instances can never *alone* give rise to an original idea, different from what is to be found in any particular instance, as has been observ'd, and as evidently follows from our fundamental principle, *that all ideas are copy'd from impressions*. Since therefore the idea of power is a new original idea, not to be found in any one instance; and which yet arises from the repetition of several instances, it follows, that the repetition *alone* has not that effect, but must either *discover* or *produce* something new, which is the source of that idea. Did the repetition neither discover nor produce any thing new, our ideas might be multiply'd by it, but wou'd not be enlarg'd above what they are upon the observation of one single instance. Every enlargement, therefore, (such as the idea of power or connexion) which arises from the multiplicity of similar instances, is copy'd from some effects of the multiplicity, and will be perfectly understood by understanding these effects. Wherever we find any thing new to be discover'd or produc'd by the repetition, there we must place the power, and must never look for it in any other object.

But 'tis evident, in the first place, that the repetition of like objects in like relations of succession and contiguity *discovers* nothing new in any one of them; since we can draw no inference from it, nor make it a subject either of our demonstrative or probable reasonings,² as has been already prov'd. Nay sup-

pose we cou'd draw an inference, 'twou'd be of no consequence in the present case; since no kind of reasoning can give rise to a new idea, such as this of power is; but wherever we reason, we must antecedently be possest of clear ideas, which may be the objects of our reasoning. The conception always precedes the understanding; and where the one is obscure, the other is uncertain; where the one fails, the other must fail also.

Secondly, 'Tis certain that this repetition of similar objects in similar situations *produces* nothing new either in these objects, or in any external body. For 'twill readily be allow'd, that the several instances we have of the conjunction of resembling causes and effects are in themselves entirely independent, and that the communication of motion, which I see result at present from the shock of two billiard-balls, is totally distinct from that which I saw result from such an impulse a twelve-month ago. These impulses have no influence on each other. They are entirely divided by time and place; and the one might have existed and communicated motion, tho' the other never had been in being.

There is, then, nothing new either discover'd or produc'd in any objects by their constant conjunction, and by the uninterrupted resemblance of their relations of succession and contiguity. But 'tis from this resemblance, that the ideas of necessity, of power, and of efficacy, are deriv'd. These ideas, therefore, represent not any thing, that does or can belong to the objects, which are constantly conjoin'd. This is an argument, which, in every view we can examine it, will be found perfectly unanswerable. Similar instances are still the first source of our idea of power or necessity; at the same time that they have no influence by their similarity either on each other, or on any external object. We must therefore, turn ourselves to some other quarter to seek the origin of that idea.

Tho' the several resembling instances, which give rise to the idea of power, have no influence on each other, and can never produce any new quality *in the object*, which can be the model of that idea, yet the *observation* of this resemblance produces a new impression *in the mind*, which is its real model. For after we have observ'd the resemblance in a sufficient number of instances, we immediately feel a determination of the mind to pass from one object to its usual attendant, and to conceive it in a stronger light upon account of that relation. This determination is the only effect of the resemblance; and therefore must be the same with power or efficacy, whose idea is deriv'd from the resemblance. The several instances of resembling conjunctions lead us into the notion of power and necessity. These instances are in themselves totally distinct from each other, and have no union but in the mind, which observes them, and collects their ideas. Necessity, then, is the effect of this observation, and is nothing but an internal impression of the mind, or a determination to carry our thoughts from one object to another. Without considering it in this view, we can never arrive at the most distant notion of it, or be able to attribute it either to external or internal objects, to spirit or body, to causes or effects.

The necessary connexion betwixt causes and effects is the foundation of our inference from one to the other. The foundation of our inference is the transition arising from the accustom'd union. These are, therefore, the same.

The idea of necessity arises from some impression. There is no impression

convey'd by our senses, which can give rise to that idea. It must, therefore, be deriv'd from some internal impression, or impression of reflexion. There is no internal impression, which has any relation to the present business, but that propensity, which custom produces, to pass from an object to the idea of its usual attendant. This therefore is the essence of necessity. Upon the whole, necessity is something, that exists in the mind, not in objects; nor is it possible for us ever to form the most distant idea of it, consider'd as a quality in bodies. Either we have no idea of necessity, or necessity is nothing but that determination of the thought to pass from causes to effects and from effects to causes, according to their experienc'd union. . . .

'Tis now time to collect all the different parts of this reasoning, and by joining them together form an exact definition of the relation of cause and effect, which makes the subject of the present enquiry. . . .

There may two definitions be given of this relation, which are only different, by their presenting a different view of the same object, and making us consider it either as a *philosophical* or as a *natural* relation; either as a comparison of two ideas, or as an association betwixt them. We may define a CAUSE to be 'An object precedent and contiguous to another, and where all the objects resembling the former are plac'd in like relations of precedence and contiguity to those objects, that resemble the latter.' If this definition be esteem'd defective, because drawn from objects foreign to the cause, we may substitute this other definition in its place, *viz.* 'A CAUSE is an object precedent and contiguous to another, and so united with it, that the idea of the one determines the mind to form the idea of the other, and the impression of the one to form a more lively idea of the other.' Shou'd this definition also be rejected for the same reason, I know no other remedy, than that the persons, who express this delicacy, should substitute a juster definition in its place. But for my part I must own my incapacity for such an undertaking. When I examine with the utmost accuracy those objects, which are commonly denominated causes and effects, I find, in considering a single instance, that the one object is precedent and contiguous to the other; and in inlarging my view to consider several instances, I find only, that like objects are constantly plac'd in like relations of succession and contiguity. Again, when I consider the influence of this constant conjunction, I perceive, that such a relation can never be an object of reasoning, and can never operate upon the mind, but by means of custom, which determines the imagination to make a transition from the idea of one object to that of its usual attendant, and from the impression of one to a more lively idea of the other. However extraordinary these sentiments may appear, I think it fruitless to trouble myself with any farther enquiry or reasoning upon the subject, but shall repose myself on them as on establish'd maxims.

Notes

1 [Book I,] part IV, sect. 5.

2 [Book I,] part III, sect. 6.

27 Efficient Cause and Active Power: an Excerpt from *Essays on the Active Powers of the Human Mind**

Thomas Reid

That active power is an attribute, which cannot exist but in some being possessed of that power, and the subject of that attribute, I take for granted as a self-evident truth. Whether there can be active power in a subject which has no thought, nor understanding, no will, is not so evident. . . .

When I observe a plant growing from its seed to maturity, I know that there must be a cause that has power to produce this effect. But I see neither the cause nor the manner of its operation.

But in certain motions of my body, and directions of my thought, I know, not only that there must be a cause that has power to produce these effects, but that I am that cause; and I am conscious of what I do in order to the production of them.

From the consciousness of our own activity, seems to be derived, not only the clearest, but the only conception we can form of activity, or the exertion of active power. . . .

If it be so that the conception of an efficient cause enters into the mind, only from the early conviction we have that we are the efficient causes of our own voluntary actions, which I think is most probable, the notion of efficiency will be reduced to this, that it is a relation between the cause and the effect, similar to that which is between us and our voluntary actions. This is surely the most distinct notion, and, I think, the only notion we can form of real efficiency.

Now it is evident, that, to constitute the relation between me and my action, my conception of the action, and will to do it, are essential. For what I never conceived, nor willed, I never did.

If any man, therefore, affirms, that a being may be the efficient cause of an action, and have power to produce it, which that being can neither conceive nor will, he speaks a language which I do not understand. If he has a meaning, his notion of power and efficiency must be essentially different from mine; and, until he conveys his notion of efficiency to my understanding, I can no more assent to his opinion, than if he should affirm, that a being without life may feel pain.

It seems therefore to me most probable, that such beings only as have some degree of understanding and will, can possess active power; and that inanimate beings must be merely passive, and have no real activity. Nothing we perceive without us affords any good ground for ascribing active power to any inanimate

* From Thomas Reid, *Essays on the Active Powers of the Human Mind*, Essay I, ch. v; first published in 1788.

being; and every thing we can discover in our own constitution, leads us to think, that active power cannot be exerted without will and intelligence.

28 Psychological and Physical Causal Laws: an Excerpt from *The Analysis of Mind**

Bertrand Russell

The traditional conception of cause and effect is one which modern science shows to be fundamentally erroneous, and requiring to be replaced by a quite different notion, that of *laws of change*. In the traditional conception, a particular event A caused a particular event B, and by this it was implied that, given any event B, some earlier event A could be discovered which had a relation to it, such that—

- (1) Whenever A occurred, it was followed by B;
- (2) In this sequence, there was something “necessary,” not a mere *de facto* occurrence of A first and then B.

The second point is illustrated by the old discussion as to whether it can be said that day causes night, on the ground that day is always followed by night. The orthodox answer was that day could not be called the cause of night, because it would not be followed by night if the earth’s rotation were to cease, or rather to grow so slow that one complete rotation would take a year. A cause, it was held, must be such that under no conceivable circumstances could it fail to be followed by its effect.

As a matter of fact, such sequences as were sought by believers in the traditional form of causation have not so far been found in nature. Everything in nature is apparently in a state of continuous change,¹ so that what we call one “event” turns out to be really a process. If this event is to cause another event, the two will have to be contiguous in time; for if there is any interval between them, something may happen during that interval to prevent the expected effect. Cause and effect, therefore, will have to be temporally contiguous processes. It is difficult to believe, at any rate where physical laws are concerned, that the earlier part of the process which is the cause can make any difference to the effect, so long as the later part of the process which is the cause remains unchanged. Suppose, for example, that a man dies of arsenic poisoning, we say that his taking arsenic was the cause of death. But clearly the process by which he acquired the arsenic is irrelevant: everything that happened before he swallowed it may be ignored, since

* From Bertrand Russell, *The Analysis of Mind* (London: Routledge and Kegan Paul, 1921). Reprinted by permission of Routledge and the Bertrand Russell Peace Foundation.

it cannot alter the effect except in so far as it alters his condition at the moment of taking the dose. But we may go further: swallowing arsenic is not really the proximate cause of death, since a man might be shot through the head immediately after taking the dose, and then it would not be of arsenic that he would die. The arsenic produces certain physiological changes, which take a finite time before they end in death. The earlier parts of these changes can be ruled out in the same way as we can rule out the process by which the arsenic was acquired. Proceeding in this way, we can shorten the process which we are calling the cause more and more. Similarly we shall have to shorten the effect. It may happen that immediately after the man's death his body is blown to pieces by a bomb. We cannot say what will happen after the man's death, through merely knowing that he has died as the result of arsenic poisoning. Thus, if we are to take the cause as one event and the effect as another, both must be shortened indefinitely. The result is that we merely have, as the embodiment of our causal law, a certain direction of change at each moment. Hence we are brought to differential equations as embodying causal laws. A physical law does not say "A will be followed by B," but tells us what acceleration a particle will have under given circumstances, i.e. it tells us how the particle's motion is changing at each moment, not where the particle will be at some future moment.

Laws embodied in differential equations may possibly be exact, but cannot be known to be so. All that we can know empirically is approximate and liable to exceptions; the exact laws that are assumed in physics are known to be somewhere near the truth, but are not known to be true just as they stand. The laws that we actually know empirically have the form of the traditional causal laws, except that they are not to be regarded as universal or necessary. "Taking arsenic is followed by death" is a good empirical generalization; it may have exceptions, but they will be rare. As against the professedly exact laws of physics, such empirical generalizations have the advantage that they deal with observable phenomena. We cannot observe infinitesimals, whether in time or space; we do not even know whether time and space are infinitely divisible. Therefore rough empirical generalizations have a definite place in science, in spite of not being exact or universal. They are the data for more exact laws, and the grounds for believing that they are *usually* true are stronger than the grounds for believing that the more exact laws are *always* true.

Science starts, therefore, from generalizations of the form, "A is usually followed by B." This is the nearest approach that can be made to a causal law of the traditional sort. It may happen in any particular instance that A is *always* followed by B, but we cannot know this, since we cannot foresee all the perfectly possible circumstances that might make the sequence fail, or know that none of them will actually occur. If, however, we know of a very large number of cases in which A is followed by B, and few or none in which the sequence fails, we shall in *practice* be justified in saying "A causes B," provided we do not attach to the notion of cause any of the metaphysical superstitions that have gathered about the word.

There is another point, besides lack of universality and necessity, which it is important to realize as regards causes in the above sense, and that is the lack of

uniqueness. It is generally assumed that, given any event, there is some one phenomenon which is *the cause* of the event in question. This seems to be a mere mistake. Cause, in the only sense in which it can be practically applied, means "nearly invariable antecedent." We cannot in practice obtain an antecedent which is *quite* invariable, for this would require us to take account of the whole universe, since something not taken account of may prevent the expected effect. We cannot distinguish, among nearly invariable antecedents, one as *the cause*, and the others as merely its concomitants: the attempt to do this depends upon a notion of cause which is derived from will, and will (as we shall see later) is not at all the sort of thing that it is generally supposed to be, nor is there any reason to think that in the physical world there is anything even remotely analogous to what will is supposed to be. If we could find one antecedent, and only one, that was *quite* invariable, we could call that one *the cause* without introducing any notion derived from mistaken ideas about will. But in fact we cannot find any antecedent that we know to be quite invariable, and we can find many that are nearly so. For example, men leave a factory for dinner when the hooter sounds at twelve o'clock. You may say the hooter is *the cause* of their leaving. But innumerable other hooters in other factories, which also always sound at twelve o'clock, have just as good a right to be called the cause. Thus every event has many nearly invariable antecedents, and therefore many antecedents which may be called its cause. . . .

Note

- 1 The theory of quanta suggests that the continuity is only apparent. If so, we shall be able theoretically to reach events which are not processes. But in what is directly observable there is still apparent continuity, which justifies the above remarks for the present.

29 Causality: an Excerpt from *A Modern Introduction to Logic**¹

L. Susan Stebbing

1 Uniformities and Multiformities

Few people would be seriously perplexed by the discovery that swans may be black; the appearance of a white peacock is interesting but not alarming. We are accustomed to seeing brown hens and white hens, black horses and chestnut

* From L. Susan Stebbing, *A Modern Introduction to Logic*, 2nd edition (London: Methuen, 1933), ch. 15.

horses, red tulips and yellow tulips, stormy seas and calm seas. But most men acquainted with snow would be startled if a lump of snow placed in front of a fire were not to melt; or if the pavements remained dry throughout a downpour of rain; or if a man shot at short range through the heart were to remain standing upright and apparently unaffected. Again, a man might be surprised if on arriving one morning at a station he found that there were no porters, nor trains, nor any signs of the bustling activity usually to be seen at railway stations. His surprise would be of a very different kind from that with which he would observe pavements remaining dry in spite of the heavy rain that fell upon them, or with which he would observe the man shot through the heart yet remaining undisturbed in position. . . . A world in which rain did not wet pavements would appear chaotic. It would appear chaotic because, since in our experience we have always found rain to wet the surface upon which it falls, we have come to expect that it will always do so. Rain that lacked this property would not *be* rain. But that a man should be caught in a shower without his umbrella surprises no one who lives in England. In other words, we believe that there are dependable regularities in the external world, although some things 'just happen so'. The life of civilized man is conditioned by the belief that if he acts in such and such a manner such and such a result will follow. Every one believes that if he is hungry and eats food, his hunger will be satisfied; that water will quench his thirst; that fire will warm him; that the ground upon which he stands will support his weight; that heat will melt snow, and that day will alternate with night. Such beliefs as these are held with varying degrees of strength. They may be mistaken. The thirst of fever is not quenched by water; a dying man is not warmed by the fire; the solid ground may quake. Nevertheless, without belief in some dependable regularities we should not act as we all in fact do. That our expectations are sometimes unfulfilled presupposes that we have formed expectations. . . .

Such experiences show that we are accustomed to distinguish between occurrences that we regard as being regularly connected and occurrences that we consider to be accidentally, or casually, conjoined. Occurrences of the first type we shall call *uniformities*; occurrences of the second type we shall call *multiformities*. Science begins with what may be described as the discovery of the minor uniformities of nature, regular connexions between facts taken in relative isolation from other facts. Simple enumeration is a method by which we may discover such minor uniformities as the connexion between *crows and black colour*, or between *flames and warmth*, or between *drinking water and quenching thirst*, or between *heavy rain and wet pavements*. These examples include such empirical generalizations as *All crows are black*, which we should not be unprepared to discover were not uniform, as well as such uniformities as *Rain wets pavements*, which we should not expect to fail. Often we find multiformities where some element of analogy might have suggested regular connexion. Sometimes in England the month of June is hot; more often it is cold. . . . Sometimes a person who has looked at the new moon through glass suffers misfortune before the moon wanes, whilst others who looked with him suffer no disaster. The first stage in science is to distinguish such multiformities from uniformities. It may then be possible by analysis of relevant factors in the complex occurrence

that constitutes a multiformity to resolve it into a uniformity of higher generality and greater abstractness. In this stage we pass insensibly from common-sense knowledge through organized common sense to knowledge that would be called ‘scientific’ in the strict sense. . . .

Daily experiences lead us, then, to distinguish between what *always happens* and what *sometimes happens but not always*. If we are successfully to order our experience and to know what to expect, we must be able to replace *sometimes* by *always*. We have now to consider whether there are any characteristics belonging to uniformities such that they are intrinsically different from multiformities. There would be such an intrinsic difference if all uniformities were *causal* connexions. We have to inquire whether this is the case, and what precisely we mean by saying that a uniformity is a *causal* uniformity. Before discussing this problem we need a more exact definition of the terms we have used. These definitions may be given as follows:

- (1) A *multiformity* is a set of occurrences, or of properties, such that some one, or more than one, member of the set sometimes recurs without the rest.
- (2) A *uniformity* is a set of occurrences, or of properties, such that if any one of them recurs, the others recur.

2 The Common-sense Notion of Cause

The plain man quite well understands how to use the word ‘cause’. Most transitive verbs, except those that express emotional attitudes, express causation, e.g. *make, produce, influence, cure, fell, cook, raise, build, destroy*. If the plain man is asked, ‘What do you mean by a cause?’ he will probably reply ‘What makes a thing happen.’ He knows that he is using the notion of cause when he says, ‘The child died from pneumonia’, ‘It was a fused wire that set the house on fire’, ‘The heat has expanded the railway line’, ‘She moved the clock so roughly that it has stopped.’ He means something definite when he says, ‘You won’t find a cure for cancer until you know its cause.’ This correct *use* of the notion of causation is, however, compatible with an extremely confused conception of what exactly causation is. The discussions of philosophers have done little, if anything, to clear up these confusions. There is some justification for Mr Russell’s remark that ‘the word “cause” is so inextricably bound up with misleading associations as to make its complete extrusion from the philosophical vocabulary desirable’.¹ But, whatever may be the case with philosophy, it is not possible to expel the word or the conception from science. ‘Cause’ expresses a concept indispensable to the earlier stages of the attempt to order the facts of experience. It is by reference to this concept that the conception of uniformities may be made determinate. It is from this point of view that we have now to consider what is meant by causal connexion. . . .

Consider the proposition *The rain wets the pavement*. This might be expressed by ‘The rain causes the pavement to be wet.’ In asserting such a proposition the plain man would mean that on this occasion the pavement would not have been

wet unless the rain had fallen on it. He would be ready to admit that on other occasions the wetness might be due to the spraying from a water-cart, or from a burst water-main. Again, ‘Charles I died because his head was cut off’ would be taken to mean that the beheading of Charles I was that occurrence that ended his life; that had the axe not struck his neck with some degree of force he would not have died as and when he did die. Thus common sense seems to regard the cause as an occurrence *relevant* to the happening of the effect. Given that the cause occurs, then the effect occurs. It seems clear that the conception of causation is confined by common sense to what happens in space and time (or in time only in the case of mental events²) and to this only in so far as what happens is regarded as changing, that is, as altering in character. In the example, ‘The rain wets the pavement’, clearly it is *the falling of the rain* that is the cause, and what it causes is a change *in the character* of the pavement. It does not cause *the pavement* but the *wetness* of the pavement which was previously dry. If the pavement had not been there, there could not be a wet *pavement*; but the pavement may be there without being *wet*. Or, to take an example given in an earlier chapter, ‘The air pulse chills the hot wire.’ Here the effect is *a change in the temperature* of the wire. Thus the notion of causation seems to be applied to a change in the character of something. We have used ‘occurrence’ to denote a spatio-temporal happening having determinate characters, or properties. Thus *the cause* is an occurrence related to some other occurrence, *the effect*. ‘Occurrence’ suggests something changing. But the effect is considered to be a change in something which relatively to it continues unchanged.

The notion of cause, then, seems to arise when we observe a change occurring in something. It is obvious that common sense will pay most attention to striking changes. A change is striking when it is sensationally impressive or emotionally affecting. It is for such changes that common sense seeks causes. Further, in determining which of the various occurrences that are present is to be taken as *the cause* common sense again selects what is striking. This selection is due to the practical attitude of the plain man who wants to know not only *what has caused* a given effect but *how to produce* such an effect on another occasion. This practical attitude is reflected in the traditional problems of causation. On the one hand, the occurrence selected as *the cause* has been isolated from other occurrences which are in fact joint-factors with it; on the other hand, the occurrence regarded as *the effect* has been left unanalysed so that different sets of factors have been regarded as the *same* effect. Hence has arisen what is known as the problem of the plurality of causes. The hackneyed illustration ‘Many causes may produce death’ affords the best example. There are more ways of killing a cat than drowning it in butter. Each of these ways would effectually kill the cat, although its state of mind and body might be very different according to what mode of killing it was actually adopted. The procedure of a coroner’s court is based upon the assumption that if the total characters of the effect-occurrence, *viz.* *the death of the person*, be made determinate, then the precise character of the cause-occurrence can be ascertained. This assumption may be mistaken but it is at least plausible. It suggests a refinement of the common-sense notion of cause, and one, moreover, that would be quite useless to common sense. For

practical purposes it is a positive advantage to know various different ways of obtaining a certain result, and it is often irrelevant for the given purpose what *other* results are also brought about. Thus, if a man desires to kill his enemy he can achieve his object by stabbing him through the heart, or by poisoning him, or by drowning him, and so on. One can obtain roast pig by burning down the house that contains the pig. The method may be wasteful, but it does not *therefore* fail of its effect. A desire to roast pigs with less expensive apparatus would suggest the elimination of certain factors from the causal occurrence, and this would involve the elimination of certain factors from the effect-occurrence.

Again, since its standpoint is practical, common sense can afford to ignore those conditions that are usually present and can therefore be taken for granted. For example, the plain man wants to light a match. He rubs it on the side of the match-box and obtains the desired result. He would say that the friction caused the flame. If, however, the operation were performed inside a jar from which the air had been exhausted, it would be found that the match did not light. He would thus find that the presence of oxygen is also necessary for the production of the effect. Since, however, air is always present when the plain man strikes a match, he takes its presence for granted and pays attention to those factors only in the total situation which he is aware of as changing. . . .

3 Development of the Common-sense Notion of Cause

We have seen that as practical agents we start from a complex situation within which we desire to bring about certain changes. Provided that the desired result is achieved what *else* is achieved can be neglected. Similarly with what is *not* desired. It is what is *always* present when death is present that matters from the practical point of view. Hence, ‘death’ stands for a set of properties abstracted from a complex set of conditions. Whenever a man is shot through the heart, *he dies*. Whenever *a man is dead*, he ceases to respond to our entreaties. The italicized words stand for complex situations which, in each case, is from the practical point of view a *single* occurrence. Such occurrences, taken as *single*, are of varying degrees of abstractness. Thus *death* is an abstraction requiring analysis. Such analysis takes us away from the standpoint of common sense. It involves looking at the whole situation retrospectively, not prospectively. The former attitude is that of the coroner’s court and the scientific investigator, the latter is that of the practical agent; the one is concerned with *knowing*, the other with *doing*. Both are concerned with uniformities, i.e. regular connexions. The practical agent, however, is content with a relation that is determinate only in the direction *from cause to effect*: *wherever X occurs, E occurs*. Such a relation may be many-one: given the cause, then the effect is determined, but not conversely. But the scientific investigator wants to find a relation that is equally determinate in either direction, that is, he seeks a one-one relation: *wherever X occurs, E occurs, and E does not occur unless X has occurred*. He has accordingly to analyse the conditions into their constituent factors so that he may ascertain whether any are irrelevant, and whether any, though necessary, are not sufficient to the occurrence of the result. The appearance of a plurality of

causes, for example, that death may sometimes be caused by pneumonia, sometimes by drowning, etc., or that thirst may be quenched by water or by cider, arises from the neglect of certain factors in the total situation that constitutes the effect-occurrence. This should be clear from what has preceded. . . .

We pinch a piece of india-rubber and its shape changes. We drop a lump of sugar into hot coffee and it dissolves. Here we have two examples of common-sense things whose characteristics change. The india-rubber left lying on the table does not change in shape. The sugar in the bowl does not dissolve. If the table is pinched, it does not change in shape. This last example suggests that the occurrence of an effect depends upon the nature of both the things that are brought into relation. The same movement of pinching will change the shape of the india-rubber but will not change the shape of the table. Thus the common-sense notion of cause seems to involve three assumptions: (1) that it is things that enter into the causal relation; (2) that the characteristics which belong to the thing, or, as common sense would say, 'the nature of the thing', is relevant to the causal situation; (3) that things left to themselves do not undergo changes. The attempt to see what precisely is involved in these assumptions may enable us to understand more clearly what causation is.

(1) The conception of what constitutes a *thing* is more or less vague. . . . But common sense distinguishes between a *thing* and *its states*. For example, the paper covering a wall would be regarded as a thing; the changes in colour as the wall-paper fades would be regarded as *states* of the wall-paper. These states also have characteristics. For example, the state of the wall-paper has the characteristic of being a pale grey-blue. The wall-paper is a thing; it has characteristics of a different kind; for example, it has the characteristic of altering in colour under the action of sunlight. Or consider *this piece of india-rubber*. It is a thing; it has the characteristic of altering in shape when pinched. *This lump of sugar* is another thing; it has the characteristic of *dissolving in water*. These characteristics of *fading*, of *elasticity*, of *solubility* belong to the thing not to its states. We shall call such characteristics *causal characteristics*.³ Each state of the thing has determinate characteristics, a definite shade of colour, a definite shape, and so on. Such characteristics we shall call *primary characteristics*. When a thing changes from one state to another these primary characteristics may be different. Since these states are *states of the thing* we say that *the thing* changes. But we want also to regard the thing as *persisting through* its changes. It is for this reason that we seek for a cause of change but not of persistence. What changes are the states; what does not change is the thing of which the states are states. The state of a thing is an occurrence. We easily recognize this in the case of water that has become frozen. We recognize the water *in a frozen state* and we see that this is an occurrence that has happened to the water. We do not so easily recognize *persisting in a state* as an occurrence. For example, if the table is in the state of continuing to be a definite shade of brown, we do not commonly think of this persisting in a definite shade as an occurrence. But if the table is knocked over we think of it as in the state of falling. There is no logical justification for thus distinguishing between these two cases. In both cases the table *has* a certain state, or is *in* a certain state, and each state has determinate,

primary characteristics. Common sense usually calls such primary characteristics ‘simple qualities’. Although it is *things* that common sense regards as entering into the causal relation, it is not *a thing* that is taken to be the cause but a certain *state* of the thing. For example, a table is not a cause, but a table *in the state of falling* may cause some one’s leg to be hurt. It is, however, a state *of the table*, so that the causal relation involves reference to things.

(2) What common sense calls ‘the nature of the thing’ is the set of characteristics that belong to it. But common sense does not clearly distinguish between the causal characteristics which belong to the thing and the primary characteristics which belong to its states. Nor is common sense at all clear with regard to the distinction between a *state* of a thing and a *characteristic*. . . . The causal characteristics of a thing are what the chemist calls the ‘properties’ of a chemical substance, such as a fat or a metal. We can now define this notion. *A causal characteristic of a thing is a characteristic mode of behaviour in relation to other things.* Thus ‘*the nature of a thing*’ includes those characteristics that it exhibits in relation to other things. . . .

(3) The assumption that ‘things left to themselves’ do not undergo change, also fails of precision, since it involves the notion of ‘one thing’. The conception of what constitutes *one thing* is vague. Whether X is to be called one thing or an aggregate of things depends, so far as common sense is concerned, mainly upon practical considerations. A lampstand and the electric-light bulb and the shade all constitute *one thing* if the lamp be used to light a room. From the point of view of purchasing the lamp, there are at least three things. . . . Apart from purely practical purposes common sense would probably regard *one thing* as definable by reference to the occupation of a sensibly continuous spatial boundary, either neglecting the time dimension or including it under the notion of persistence of sensibly similar characteristics through a period of time. We have seen that common sense distinguishes between a thing and its states. When there is considerable alteration in the primary characteristics, then common sense would refuse to admit the persistence of the thing. Thus common sense requires sensible continuity of characteristics, and assumes that there is such sensible continuity even when it has not been continuously perceived. Hence, it is argued, if there is a change in the sensible characteristics manifested by a thing in a certain state, there must be something to *make* it change. In this way arises the assumption that things left to themselves do not change. For example, given that a candle is one thing, then common sense does not expect it to change while the candle is standing unlighted on the table. If the candle which was standing upright in the candlestick is, after a few hours, bending over the candlestick, common sense assumes that something other than the candle has caused the change, e.g. the heat of the room. It is a causal characteristic of the candle to become bent under certain conditions of temperature. These conditions are dependent upon other things, for example, the fire, the relative positions of the fire and the candle, etc. From the causal point of view these conditions constitute one situation, or set of related things, which may be regarded as a

system. If the candle in its stick were regarded as one system, we should have to distinguish at least three different things, viz. candlestick, tallow, wick. These things are in spatio-temporal relations. If no change were occurring in this system, then it would be assumed that no change would occur *unless* something outside the system, e.g. a fire, or a lighted match, came into spatio-temporal relation with it. If, however, change were occurring in the candle-system independently of anything outside it, then it would be assumed that something was going on all the time *in* the system, whether it were at first perceptible or not.

We see, then, that the attempt to analyse a total causal situation involves the distinction of different factors standing in spatio-temporal relations. *What* occurs will be dependent upon the causal characteristics of the things in that situation. Thus the fire which melts the candle merely warms the brass candlestick. A roaring fire in the kitchen does not melt the candle in the bedroom. The factors in a causal situation must be in spatio-temporal proximity.⁴ But not everything in the given situation is relevant to the given causal occurrence. If it were there would be no causal uniformities since some factors in the situation do not recur. No two causal situations are exactly alike. A causal uniformity is a connexion between factors *recognizable as the same* on different occasions of their occurrence, i.e. under varying conditions and at different places and times.

The development of the common-sense notion of cause brings out several points of importance. The consideration of these will enable us to make clear certain distinctions with regard to which common sense is confused.

(1) A causal uniformity is an abstraction since it connects sets of recurrent characteristics belonging to events which do not recur.

(2) Neither the distinction between a thing and its states, nor the distinction between the qualities that a thing has and the way in which it behaves in relation to other things is clearly drawn at the level of common sense. . . .

. . . These distinctions throw light upon the distinction, so vaguely conceived by common sense, between *cause* and *condition*. Since the causal characteristics of the thing are its characteristic modes of behaviour in relation to other things, it follows that how a thing behaves depends upon what other things are in relation to it. This may be made clearer by means of an example. Let us consider a simple experiment. A bell so arranged that it can be continuously rung by clockwork is hung by silk threads inside a glass jar. The air in the jar is exhausted by an air-pump. As the air is withdrawn, the sound decreases, and very soon ceases to be heard, although the tongue of the bell still strikes against its sides. Given this arrangement, then, the air is a necessary condition for the propagation of sound. Now, it would commonly be said that the striking of the bell was the sufficient cause of the sound. This experiment shows that a material medium, such as air or water, is also required for the production of the sound. This material medium is then said to be a *condition*. Both the air and the striking of the bell are necessary for the production of the sound; together they are sufficient. Reflection upon this distinction emphasizes the importance of the causal

characteristics of things. In the bell experiment the medium has the causal characteristic of being able to propagate sound-waves; the bell has the causal characteristic of vibrating in such a way when struck as to set up sound-waves in a suitable medium in spatio-temporal proximity to it. A condition is, then, whatever must be present in a given situation in order that a causal characteristic of a thing may be manifested in a state of the thing, which state will have certain determinate characteristics. This state is the effect. The cause is that state of some other thing upon which the effect is consequent. In the example of the bell, the cause may be said to be the impact of the tongue upon the sides; the effect is the vibration of the sides which has for *its* effect the communication of sound-waves to the surrounding air. This distinction between cause and condition cannot be made perfectly precise and is misleading if pressed too far. What is important is to distinguish between a *sufficient* condition and a *necessary* condition. A condition X is a *sufficient condition* of an occurrence A provided that whenever X is present A occurs. But if A may occur when X is absent, then X, though a sufficient, is not a *necessary condition* of A. Thus a condition N is a *necessary condition* of A provided that A never occurs in the absence of N. A condition NS is a *necessary and sufficient condition* of an occurrence A provided that (i) whenever NS is present A occurs, and (ii) whenever NS is absent A does not occur. Owing to the failure of common sense to recognize these distinctions X is sometimes said to be 'the cause' of A when it is a necessary but not a sufficient condition, and also when it is a sufficient but not a necessary condition. This ambiguity in the use of the word 'cause' is due to the *practical* interests of common sense, which, as we saw, leads to the selection of a striking, or impressive, factor out of the set of factors that are jointly sufficient and independently necessary to the production of the effect. Hence, common sense fails to recognize that what we have to take into account is a system the parts of which are in mutual dependence. This dependence is causal dependence.

(3) It follows from what has just been said that the distinction between cause and effect cannot be made as sharply as common sense makes it. The emphasis must be placed upon the relation, *cause* and *effect* being merely the terms in the relation, selected because they are striking, or practically important, or are easily discriminated. This practical emphasis leads, as we saw, to the neglect of other factors that are relevant, and hence to the conception of the causal relation as being not only asymmetrical but also many-one. But it is usually assumed that if the cause and the effect are determined with equal precision, the relation will be one-one, so that given the effect, the cause is thereby determined, given the cause, the effect is thereby determined. . . .

(4) Common sense assumes that if in a system in which no change has been occurring, a change begins to occur, then that system must be in causal relation to something outside it which causes the change. Such causation is called *transeunt*. Thus we are led to the distinction between a thing left to itself and a thing not so left. We saw that this distinction is vague. It must be replaced by the distinction between an *isolated system* and a *system in causal relations to something outside the*

system. Changes occurring in an isolated system are determined by the mutual relations of the parts. Such determination is called *immanent causation*. For example, the works of a watch constitute an isolated system. Once the watch is wound up the changes occurring in it are causally determined by the mutual relations of the parts of the works. Thus the movement of the hands over the dial is immanently caused. If, however, the watch is put in very cold or very hot temperatures, the temperature of the surrounding medium will cause a change in the metal case which will cause a change in the working of the watch. This would be an example of transeunt causation. The business of a good watch-maker is to construct a watch as little subject as possible to changes occurring outside the watch-case. His ideal would be the construction of a completely isolated system, save for the fact that the watch must be periodically wound up by external agency. This ideal is unattainable. The distinction between systems that are causally isolated and systems that are not cannot be made absolute. The latter may always be regarded as sub-systems in a wider system. But unless there were systems that are practically isolated with regard to many changes occurring in these systems, the discovery of causal uniformities would be impossible. . . .

Our discussion of causation has shown that there is a close interrelation between causal uniformities, or, as we may call them, causal laws, and things. The attempt to determine more precisely the nature of this interrelation takes us beyond the standpoint of common sense.

4 Causal Laws and the Behaviour of Things

We have seen that the way in which a thing, for example, a lump of sugar, a candle, a poker, a living being, will behave in a given situation depends both upon the nature of the thing and upon the nature of the situation in which it is placed. This lump of sugar dissolves in water; this piece of gold does not. The poker put into a fire will become red-hot; when it is taken out and put in the fender it will become cold again, and will revert (approximately) to its former condition. The thing has characteristics which distinguish it from other things. Some of its characteristics are causal characteristics, i.e. modes of behaviour in relation to other things, e.g. *solubility in water* which belongs to this lump of sugar. The thing has also non-causal characteristics relating to the kind of primary characteristics exhibited by its states. The states of the thing have determinate characteristics. These determinate characteristics of the states are *caused* by the causal characteristics and the situation in which the thing is placed. For example, the determinate characteristics of the state of the poker when it is red-hot are *being red* and *being hot*. These characteristics are caused by the causal characteristics *altering in colour* and *altering in temperature* (which belong to the poker, not to its states) and by the situation, viz. the fire.

So far we have considered definite examples of things, *this poker, this lump of sugar*. But each of these is recognized as belonging to a class consisting of *things of a certain sort*, or as we have called them *natural kinds*. Every instance

of a *kind of thing* has certain characteristics of a certain sort which makes it the *sort of thing* it is, and is what we mean by a *kind*. The way in which a thing behaves depends upon its kind. These modes of behaviour are causal laws. Wherever there are things of a certain kind in certain situations there will be certain modes of behaviour, that is certain variations in the primary characteristics of the states of the thing. These changes *recur* under suitable conditions at different times and places. Hence, the characteristic modes of behaviour of things are recurrent modes of change. Causal laws are the laws of these recurrent modes of change.

There can be no doubt that we do distinguish kinds of things by observing their modes of behaviour in the presence of other things, that is in different situations. We observe the primary characteristics of the states of a thing, and we know that that kind of thing has states having those characteristics. If the thing fails to exhibit that mode of behaviour which is characteristic of that kind of thing, we know that we were mistaken as to the kind of thing it was. For example, we may see a dish of apples which *look like* Blenheim pippins. We may take up one and bite it, only to find that it tastes of soap. We conclude (rightly) that *this thing* is a piece of soap made to look like a Blenheim pippin. It *isn't* a Blenheim pippin because it doesn't behave like one. Thus we see that the distinguishing characteristics of a kind involve modes of behaviour, i.e. causal laws. The notion of kinds of things, then, leads us to the consideration of causation and conditions. . . .

To say that X is the necessary and sufficient condition of the occurrence of E is, then, to say that X *alone* is relevant to the occurrence of E. It might be objected that, if we can discover what is relevant only by eliminating what is irrelevant, then we could never tell whether permanent factors in the universe, for example, the presence of the fixed stars, are relevant to a given causal situation, for example, *sugar dissolving in water*. This is true, but it is also unimportant since the statement of the given causal law does not require us to take account of the fixed stars, nor should we ever be concerned with causal situations in the absence of the fixed stars. Moreover, although the annihilation of the fixed stars might affect the result, we have not the slightest reason for supposing that this would be the case. On the other hand, we do find that the substitution of gold for sugar in the water does not yield the result *dissolving in water*. It seems, then, that experience does provide us with examples of multiformities, and with examples of causal occurrences that are independent of other causal occurrences happening contemporaneously and in the same neighbourhood. That is to say that there are relatively independent causal series. The difference between the causal set A₁, A₂, A₃ . . . and the causal set B₁, B₂, B₃ . . . depends upon the different natures of A and B. We have seen what is meant by the phrase 'the nature of' in this context. It is the fact that A has a certain nature, or is a thing of a certain kind, which determines in what situations A is a causal factor.

It is important to distinguish causal laws from the particular causal propositions which exemplify them. It is the causal law that is fundamental. A particular causal proposition asserts a definite causal occurrence happening once, and once

only, for example, *This shot through his heart caused this man's death*. In asserting that this man's death was *caused* by his being shot through the heart we are asserting more than the historical fact that two particular occurrences were conjoined. This is clear, since there are many occurrences happening together (simultaneously or successively) which we should not regard as being causally connected. It may be that what *more* we are asserting is simply that this is an instance of a conjunction of two occurrences of a certain sort such that the one is *always* conjoined with the other, or it may be that we are asserting that the two occurrences are related by a unique relation of causation. We shall consider in the next section what can be said in favour of either of these two views. Whatever view we adopt we must admit that there would be no significance in the assertion of causation unless we at least meant to assert that *whenever* a given occurrence happens, then some other given occurrence happens. The causal law which the example given above exemplifies can be stated precisely. Whenever there is an occurrence which is the passage of a bullet through a man's heart, there follows an occurrence which is the cessation of the beating of the heart. Thus the form of such a causal law is: Whenever an occurrence having the property Φ happens at a time t_1 to a thing of the kind K_1 , then an occurrence having the property Ψ happens at a time t_2 to a thing of the kind K_2 . It may be the case that (i) Φ and Ψ are properties of the same sort; (ii) K_1 and K_2 are the same thing; (iii) t_1 and t_2 are the same time. When (iii) is the case, there is an instance of simultaneous causality. In the examples we have given the *things* have been of different degrees of complexity, e.g. *gold*, an element, *water*, an inorganic compound, *heart*, *man*, organic compounds. The behaviour of each of these kinds is expressed by causal laws. These causal laws will differ in the degree of their abstractness, and from some points of view and with regard to certain problems the differences between the *kind* of these kinds, or sorts, of things will be very important. But it is sufficient here to notice that the *simplest* causal law is abstract. . . .

6 Causation and Regular Sequence⁵

The problem we have now to discuss is whether causal laws express *nothing but* regularities of sequence. If so, it would follow that all regular sequences are causal, e.g. the sequence of day and night. If not, the difficulty arises of finding some characteristic distinguishing regular sequences that are causal from those that are not.

There is undoubtedly something to be said in favour of the view that causal regularities are nothing but observed regularities of sequence. Its best-known recent exponent is Mr Bertrand Russell.⁶ Unfortunately his argument is so carelessly expressed that it is difficult to extract the main points. Perhaps they may be said to lie in the two following considerations. Starting with the admission that causal laws are of the form 'A causes B', e.g. 'Arsenic causes death', Mr Russell argues that such laws are liable to exception, and that, consequently, they cannot be universal and necessary. Now a 'law' that has exceptions would not generally be regarded as a law. Mr Russell, however, does not take this

view, for he seems to wish to maintain that *A causes B*, expresses a law and that it means 'A is the nearly invariable antecedent of B'. By 'nearly invariable' Mr Russell seems to mean 'almost *unvarying*'. He argues that we cannot say that arsenic always causes death since a man who has swallowed arsenic 'might be shot through the head immediately after taking the dose, and then it would not be of arsenic that he would die.' Again, it 'may happen that immediately after the man's death his body is blown to pieces by a bomb. We cannot say what will happen after the man's death, through merely knowing that he has died as the result of arsenic poisoning.' Accordingly, he argues that 'if we are to take the cause as one event and the effect as another, both must be shortened indefinitely.' We are thus left with laws expressing the direction of change from moment to moment. The upshot of this argument appears to be that since a change occupies a finite time, and since a change A that is usually followed by a change B may be interrupted before the completion of the process, we cannot assert that 'A is always followed by B', so long as we are concerned with perceptible changes. Thus causal laws are not universal. The second point concerns the difficulty of finding any *one* event which can be regarded as *the* cause of a given event. This difficulty leads Mr Russell to deny the uniqueness of the causal relation. He argues: 'Cause, in the only sense in which it can be practically applied, means "nearly invariable antecedent". We cannot in practice obtain an antecedent which is *quite* invariable, for this would require us to take account of the whole universe, since something not taken account of may prevent the expected effect.' The man who first swallowed arsenic, who was immediately afterwards shot through the head, and whose body was, immediately after death, blown to pieces by a bomb is said to provide an illustration of such 'prevention'. Hence Mr Russell concludes that 'in fact we cannot find any antecedent that we know to be quite invariable', but 'we can find many that are nearly so. For example, men leave a factory for dinner when the hooter sounds at twelve o'clock. You may say the hooter is *the* cause of their leaving. But innumerable other hooters in other factories, which also always sound at twelve o'clock, have just as good a right to be called the cause. Thus every event has many nearly invariable antecedents, and therefore many antecedents which may be called its cause.'

If Mr Russell's view be correct, then *every* regular sequence is causal, since there is nothing more in the notion of cause than regularity of sequence. Thus night will be the cause of day and day will be the cause of night. On this view we should have to admit that the blowing of the hooters was the cause of the position of the hands of the clock when the men began to leave the factory for dinner, and that the blowing of the hooter in a Manchester factory caused both the departure of the men from that factory and also the departure of the men from factories in Liverpool and in London, and conversely. The most surprising point with regard to Mr Russell's argument is his belief that such an account of causation gives the only sense in which the notion can be practically applied. Presumably we apply the notion of cause when we use it for purposes of inference. It is not obvious that the hooter illustration could be used satisfactorily in practice, so that it may be doubted whether this definition of causal connexion

could be recommended on the ground of its practical utility. It is unlikely, however, that the reasons Mr Russell gives for his view are in fact the reasons that led him to it. Possibly his main reason for adopting such a paradoxical view is to be found in the extreme difficulty of pointing out *any* characteristic which suffices to distinguish regular sequences that are causal from those that are not. This difficulty may have led to the conclusion that there is no such characteristic. Such an argument is by no means conclusive. Dr Broad has put this point very clearly.⁸ He argues that if causation did involve a unique and not further analysable relation it would be ‘impossible to define it in any but tautologous terms’. In that case it ‘might be that regular sequence was not even *part* of what we mean by causation, but was merely a sign (though by no means an infallible one) by which the presence of this other relation is indicated’. Dr Broad admits that we do not seem to be directly acquainted with any ‘extra factor’ in causation in the way in which, for instance, we are directly acquainted with the unique and unanalysable relation of inside and outside in space. Thus it remains *possible* that the main reason for rejecting Mr Russell’s view may simply be its paradoxical consequences. But, as Dr Broad further argues, ‘there are many cases where we should admit regular sequence and *unhesitatingly deny* causation’, although, he adds, ‘there are perhaps no cases where we can *unhesitatingly assert* causation in addition to regular sequence’. It may certainly be admitted that the working scientist would unhesitatingly deny that the blowing of a hooter in Manchester was the cause of the departure of the London workmen.

If we ask why the hooter illustration is paradoxical we may be able to discover the ‘extra factor’ that is missing from Mr Russell’s account. Dr Broad says, ‘the missing factor seems to be a certain spatio-temporal continuity between the sequent events’, and he adds, ‘I am inclined to think that it is the absence of such continuity between the blowing of the Manchester hooter and the movement of the London workmen which makes me so certain that the former is not a cause of the latter.’ This suggestion will meet the difficulty only if ‘a *certain* spatio-temporal continuity’ be interpreted as involving reference to continuity of change of character of the events happening in Manchester, or in London. It is the absence of continuous change of character that leads to the paradox. If what we have already said about causal laws is correct, then it is a mistake to suppose that one *event* causes another *event*. We have insisted that it is an event’s having a certain character that causes another event having a certain character to have some other character. The missing factor must then be found in the character of the event. Causal laws connect changes in the characters of events, and there must be continuity in this change of character. That this reference to character is essential is shown by the fact that we speak of ‘*regular sequences*’. Events do not recur. As we have seen, we can speak of the *same* cause on different occasions only because the causal connexion is primarily between the characters, and is derivatively between the events to which these characters belong. Mr Russell’s view is, then, to be rejected because it takes no account of the continuity of change of *character* that is essential to causation. We conclude that causation cannot be regarded as *equivalent to* regular sequence.

... Those who hold the view that science makes no use of the notion of cause

have generally been more interested in the physical than in the biological and social sciences. Thus Mr Russell says 'in advanced sciences such as gravitational astronomy, the word "cause" never occurs', and he adds, 'the reason why physics has ceased to look for causes is that, in fact, there are no such things'.⁹ It may certainly be admitted that 'in advanced sciences' the notion of cause is replaced by the notion of functional dependence. But it is a mistake to suppose that apart from the 'advanced sciences' there is no scientific method. On the contrary, the development of science from its earliest stages to its most advanced stages has been continuous. It is the merest dogmatism to confine the 'sciences' to physics and to argue that because the physicist does not employ the notion of cause, therefore 'science' has no use for it. The most superficial acquaintance with the earlier stages of a science is enough to reveal that the notion of cause is indispensable. There is no doubt, for instance, that the word 'cause' frequently occurs in the works of biologists.¹⁰ Those sciences that are concerned with the recurrent modes of behaviour of different *kinds* of things undoubtedly use the notion of cause in the form of causal laws. The bio-chemist carries out careful experiments with regard to the action of chemicals upon living organisms in order to discover their modes of behaviour, that is, their causal laws. Thus, for instance, he uses such expressions as 'Nitrites cause a fall in blood-pressure', and he employs the notion of cause when he infers that an injection of amylnitrite will cause such a fall in blood-pressure. He is content to leave to philosophers doubts as to the validity of the concept of cause so long as he is able to continue to use it. Hence, it seems rash to conclude from an examination of the words used by physicists that 'there are no such things' as causes.

Notes

- 1 *Mysticism and Logic* (New York: Longmans, Green, 1918), p. 180. Mr Russell himself has recently based his philosophy of science upon the conception of 'causal lines', which, however, he does not attempt to analyse, and which perhaps cannot bear the weight of the construction he rests upon it.
- 2 Common sense certainly regards mental events as non-spatial, and somewhat waveringly applies the conception of causation to such events. It is not necessary for our purpose to discuss this application. Hence, our discussion is limited to non-mental events.
- 3 The expression 'causal characteristic' is taken from Dr C. D. Broad (see *The Mind and its Place in Nature* (London: Routledge & Kegan Paul), p. 432). In the discussion of this problem I am, as always, much indebted to Dr Broad's writings.
- 4 In saying that there must be 'spatio-temporal proximity', I mean that the various factors cannot be separated by a spatio-temporal gap.
- 5 This section should be omitted on a first reading.
- 6 See *The Analysis of Mind*, chap. V [reprinted, in part, in this volume].
- 7 See Reading 28, p. 228, in this volume.
- 8 *The Mind and its Place in Nature*, pp. 453–6.
- 9 *Mysticism and Logic*, p. 180.
- 10 We have seen that the physicist also employs the notion of cause, although no doubt its analysis would be different in the case of the physical sciences from its analysis in the biological sciences.

30 Causality and Determination*

G. E. M. Anscombe

I

It is often declared or evidently assumed that causality is some kind of necessary connection, or alternatively, that being caused is – non-trivially – instancing some exceptionless generalization saying that such an event always follows such antecedents. Or the two conceptions are combined.

Obviously there can be, and are, a lot of divergent views covered by this account. any view that it covers nevertheless manifests one particular doctrine or assumption. Namely:

If an effect occurs in one case and a similar effect does not occur in an apparently similar case, there must be a relevant further difference.

Any radically different account of causation, then, by contrast with which all those diverse views will be as one, will deny this assumption. Such a radically opposing view can grant that often – though it is difficult to say generally when – the assumption of relevant difference is a sound principle of investigation. It may grant that there are necessitating causes, but will refuse to identify causation as such with necessitation. It can grant that there are situations in which, given the initial conditions and no interference, only one result will accord with the laws of nature; but it will not see general reason, in advance of discovery, to suppose that any given course of things has been so determined. So it may grant that in many cases difference of issue can rightly convince us of a relevant difference of circumstances; but it will deny that, quite generally, this *must* be so.

The first view is common to many philosophers of the past. It is also, usually but not always in a neo-Humeian form, the prevailing received opinion throughout the currently busy and productive philosophical schools of the English-speaking world, and also in some of the European and Latin American schools where philosophy is pursued in at all the same sort of way; nor is it confined to these schools. So firmly rooted is it that for many even outside pure philosophy, it routinely determines the meaning of ‘cause’, when consciously used as a theoretical term: witness the terminology of the contrast between ‘causal’ and ‘statistical’ laws, which is drawn by writers on physics – writers, note, who would not conceive themselves to be addicts of any philosophic school when they use this language to express that contrast.

* An Inaugural Lecture delivered at Cambridge, and published by the Cambridge University Press (1971). Reprinted by permission of the author and Cambridge University Press.

The truth of this conception is hardly debated. It is, indeed, a bit of *Weltanschauung*: it helps to form a cast of mind which is characteristic of our whole culture.

The association between causation and necessity is old; it occurs for example in Aristotle's *Metaphysics*: 'When the agent and patient meet suitably to their powers, the one acts and the other is acted on OF NECESSITY.' Only, with 'rational powers', an extra feature is needed to determine the result: 'What has a rational power [e.g. medical knowledge, which can kill *or* cure] OF NECESSITY does what it has the power to do and as it has the power, when it has the desire' (Book IX, Chapter V).

Overleaping the centuries, we find it an axiom in Spinoza, 'Given a determinate cause, the effect follows OF NECESSITY, and without its cause, no effect follows' (*Ethics*, Book I, Axiom III). And in the English philosopher Hobbes:

A cause simply, or an entire cause, is the aggregate of all the accidents both of the agents how many soever they be, and of the patients, put together; which when they are supposed to be present, IT CANNOT BE UNDERSTOOD BUT THAT THE EFFECT IS PRODUCED at the same instant; and if any of them be wanting, IT CANNOT BE UNDERSTOOD BUT THAT THE EFFECT IS NOT PRODUCED. (*Elements of Philosophy Concerning Body*, Chapter IX)

It was this last view, where the connection between cause and effect is evidently seen as *logical* connection of some sort, that was overthrown by Hume, the most influential of all philosophers on this subject in the English-speaking and allied schools. For he made us see that, given any particular cause – or 'total causal situation' for that matter – and its effect, there is not in general any contradiction in supposing the one to occur and the other not to occur. That is to say, we'd know what was being described – what it would be like for it to be true – if it were reported for example that a kettle of water was put, and kept, directly on a hot fire, but the water did not heat up.

Were it not for the preceding philosophers who had made causality out as some species of logical connection, one would wonder at this being called a discovery on Hume's part: for vulgar humanity has always been over-willing to believe in miracles and marvels and *lusus naturae*. Mankind at large saw no contradiction, where Hume worked so hard to show the philosophic world – the Republic of Letters – that there was none.

The discovery was thought to be great. But as touching the equation of causality with necessitation, Hume's thinking did nothing against this but curiously reinforced it. For he himself assumed that NECESSARY CONNECTION is an essential part of the idea of the relation of cause and effect (*A Treatise of Human Nature*, Book I, Part III, Sections II and VI), and he sought for its nature. He thought this could not be found in the situations, objects or events called 'causes' and 'effects', but was to be found in the human mind's being determined, by experience of CONSTANT CONJUNCTION, to pass from the sensible impression or memory of one term of the relation to the convinced idea of the other. Thus to say that an event was caused was to say that its occurrence was an instance of some

exceptionless generalization connecting such an event with such antecedents as it occurred in. The twist that Hume gave to the topic thus suggested a connection of the notion of causality with that of deterministic laws – i.e. laws such that always, given initial conditions and the laws, a unique result is determined.

The well-known philosophers who have lived after Hume may have aimed at following him and developing at least some of his ideas, or they may have put up a resistance; but in no case, so far as I know,¹ has the resistance called in question the equation of causality with necessitation.

Kant, roused by learning of Hume's discovery, laboured to establish causality as an *a priori* conception and argued that the objective time order consists 'in that order of the manifold of appearance according to which, IN CONFORMITY WITH A RULE, the apprehension of that which happens follows upon the apprehension of that which precedes. . . . In conformity with such a rule there must be in that which precedes an event the condition of a rule according to which this event INVARIABLY AND NECESSARILY follows' (*Critique of Pure Reason*, Book II, Chapter II, Section III, Second Analogy). Thus Kant tried to give back to causality the character of a *justified* concept which Hume's considerations had taken away from it. Once again the connection between causation and necessity was reinforced. And this has been the general characteristic of those who have sought to oppose Hume's conception of causality. They have always tried to establish the necessitation that they saw in causality: either *a priori*, or somehow out of experience.

Since Mill it has been fairly common to explain causation one way or another in terms of 'necessary' and 'sufficient' conditions. Now 'sufficient condition' is a term of art whose users may therefore lay down its meaning as they please. So they are in their rights to rule out the query: 'May not the sufficient conditions of an event be present, and the event yet not take place?' For 'sufficient condition' is so used that if the sufficient conditions for *X* are there, *X* occurs. But at the same time, the phrase cozens the understanding into not noticing an assumption. For 'sufficient condition' sounds like: 'enough'. And one certainly can ask: 'May there not be *enough* to have made something happen – and yet it not have happened?'

Russell wrote of the notion of cause, or at any rate of the 'law of causation' (and he seemed to feel the same way about 'cause' itself), that, like the British monarchy, it had been allowed to survive because it had been erroneously thought to do no harm. In a destructive essay of great brilliance he cast doubt on the notion of necessity involved, unless it is explained in terms of universality, and he argued that upon examination the concepts of determination and of invariable succession of like objects upon like turn out to be empty: they do not differentiate between any conceivable course of things and any other. Thus Russell too assumes that necessity or universality is what is in question, and it never occurs to him that there may be any other conception of causality ('The Notion of Cause', in *Mysticism and Logic*).²

Now it's not difficult to show it *prima facie* wrong to associate the notion of cause with necessity or universality in this way. For, it being much easier to trace effects back to causes with certainty than to predict effects from causes, we often know a cause without knowing whether there is an exceptionless generalization of the kind envisaged, or whether there is a necessity.

For example, we have found certain diseases to be contagious. If, then, I have had one and only one contact with someone suffering from such a disease, and I get it myself, we suppose I got it from him. But what if, having had the contact, I ask a doctor whether I will get the disease? He will usually only be able to say, 'I don't know – maybe you will, maybe not.'

But, it is said, knowledge of causes here is partial; doctors seldom even know any of the conditions under which one invariably gets a disease, let alone all the sets of conditions. This comment betrays the assumption that there is such a thing to know. Suppose there is: still, the question whether there is does not have to be settled before we can know what we mean by speaking of the contact as cause of my getting the disease.

All the same, might it not be like this: knowledge of causes is possible without any satisfactory grasp of what is involved in causation? Compare the possibility of wanting clarification of 'valency' or 'long-run frequency', which yet have been handled by chemists and statisticians without such clarification; and valencies and long-run frequencies, whatever the right way of explaining them, have been known. Thus one of the familiar philosophic analyses of causality, or a new one in the same line, may be correct, though knowledge of it is not necessary for knowledge of causes.

There is something to observe here, that lies under our noses. It is little attended to, and yet still so obvious as to seem trite. It is this: causality consists in the derivativeness of an effect from its causes. This is the core, the common feature, of causality in its various kinds. Effects derive from, arise out of, come of, their causes. For example, everyone will grant that physical parenthood is a causal relation. Here the derivation is material, by fission. Now analysis in terms of necessity or universality does not tell us of this derivedness of the effect; rather it forgets about that. For the necessity will be that of laws of nature; through it we shall be able to derive knowledge of the effect from knowledge of the cause, or vice versa, but that does not show us the cause as source of the effect. Causation, then, is not to be identified with necessitation.

If *A* comes from *B*, this does not imply that every *A*-like thing comes from some *B*-like thing or set-up or that every *B*-like thing or set-up has an *A*-like thing coming from it; or that given *B*, *A* had to come from it, or that given *A*, there had to be *B* for it to come from. Any of these may be true, but if any is, that will be an additional fact, not comprised in *A*'s coming from *B*. If we take 'coming from' in the sense of travel, this is perfectly evident.

'But that's because we can observe travel!' The influential Humeian argument at this point is that we can't similarly observe causality in the individual case (*A Treatise of Human Nature*, Book I, Part III, Section II). So the reason why we connect what we call the cause and what we call the effect as we do must lie elsewhere. It must lie in the fact that the succession of the latter upon the former is of a kind regularly observed.

There are two things for me to say about this. First, as to the statement that we can never observe causality in the individual case. Someone who says this is just not going to count anything as 'observation of causality'. This often happens in philosophy; it is argued that 'all we find' is such-and-such, and it turns out that

the arguer has excluded from his idea of ‘finding’ the sort of thing he says we don’t ‘find’. And when we consider what we are allowed to say we do ‘find’, we have the right to turn the tables on Hume, and say that neither do we perceive bodies, such as billiard balls, approaching one another. When we ‘consider the matter with the utmost attention’, we find only an impression of travel made by the successive positions of a round white patch in our visual fields . . . etc. Now a ‘Humeian’ account of causality has to be given in terms of constant conjunction of physical things, events, etc., not of experiences of them. If, then, it must be allowed that we ‘find’ bodies in motion, for example, then what theory of perception can justly disallow the perception of a lot of causality? The truthful – though unhelpful – answer to the question: ‘How did we come by our primary knowledge of causality?’ is that in learning to speak we learned the linguistic representation and application of a host of causal concepts. Very many of them were represented by transitive and other verbs of action used in reporting what is observed. Others – a good example is ‘infect’ – form, not observation statements, but rather expressions of causal hypotheses. The word ‘cause’ itself is highly general. How does someone show that he has the concept *cause*? We may wish to say: only by having such a word in his vocabulary. If so, then the manifest possession of the concept presupposes the mastery of much else in language. I mean: the word ‘cause’ can be *added* to a language in which are already represented many causal concepts. A small selection: *scrape, push, wet, carry, eat, burn, knock over, keep off, squash, make* (e.g. noises, paper boats), *hurt*. But if we care to imagine languages in which no special causal concepts are represented, then no description of the use of a word in such languages will be able to present it as meaning *cause*. Nor will it even contain words for natural kinds of stuff, nor yet words equivalent to ‘body’, ‘wind’, or ‘fire’. For learning to use special causal verbs is part and parcel of learning to apply the concepts answering to these and many other substantives. As surely as we learned to call people by name or to report from seeing it that the cat was on the table, we also learned to report from having observed it that someone drank up the milk or that the dog made a funny noise or that things were cut or broken by whatever we saw cut or break them.

(I will mention, only to set on one side, one of the roots of Hume’s argument, the implicit appeal to Cartesian scepticism. He confidently challenges us to ‘produce some instance, wherein the efficacy is plainly discoverable to the mind, and its operations obvious to our consciousness or sensation’ (*A Treatise of Human Nature*, Book I, Part III, Section XIV). Nothing easier: is cutting, is drinking, is purring not ‘efficacy’? But it is true that the apparent perception of such things may be only apparent: we may be deceived by false appearances. Hume presumably wants us to ‘produce an instance’ in which *efficacy* is related to sensation as *red* is. It is true that we can’t do that; it is not *so* related to sensation. He is also helped, in making his argument that we don’t perceive ‘efficacy’, by his curious belief that ‘efficacy’ means much the same thing as ‘necessary connection’! But as to the Cartesian-sceptical root of the argument, I will not delay upon it, as my present topic is not the philosophy of perception.)

Secondly, as to that instancing of a universal generalization, which was supposed to supply what could not be observed in the individual case, the causal

relation, the needed examples are none too common. ‘Motion in one body in all past instances that have fallen under our observation, is follow’d upon impulse by motion in another’: so Hume (*A Treatise of Human Nature*, Book II, Part III, Section I). But, as is always a danger in making large generalizations, he was thinking only of the cases where we do observe this – billiard balls against free-standing billiard balls in an ordinary situation; not billiard balls against stone walls. Neo-Humeians are more cautious. They realize that if you take a case of cause and effect, and relevantly describe the cause *A* and the effect *B*, and then construct a universal proposition, ‘Always, given an *A*, a *B* follows’, you usually won’t get anything true. You have got to describe the absence of circumstances in which an *A* would not cause a *B*. But the task of excluding all such circumstances can’t be carried out. There is, I suppose, a vague association in people’s minds between the universal propositions which would be examples of the required type of generalizations, and scientific laws. But there is no similarity.

Suppose we were to call propositions giving the properties of substances ‘laws of nature’. Then there will be a law of nature running ‘The flash-point of such a substance is . . .’, and this will be important in explaining why striking matches usually causes them to light. This law of nature has not the form of a generalization running ‘Always, if a sample of such a substance is raised to such a temperature, it ignites’; nor is it equivalent to such a generalization, but rather to: ‘If a sample of such a substance is raised to such a temperature and doesn’t ignite, there must be a cause of its not doing so.’ Leaving aside questions connected with the idea of a pure sample, the point here is that ‘normal conditions’ is quite properly a vague notion. That fact makes generalizations running ‘Always . . .’ merely fraudulent in such cases; it will always be necessary for them to be hedged about with clauses referring to normal conditions; and we may not know in advance whether conditions are normal or not, or what to count as an abnormal condition. In exemplar analytical practice, I suspect, it will simply be a relevant condition in which the generalization, ‘Always if such and such, such and such happens . . .’, supplemented with a few obvious conditions that have occurred to the author, turns out to be untrue. Thus the conditional ‘If it doesn’t ignite then there must be some cause’ is the better gloss upon the original proposition, for it does not pretend to say specifically, or even disjunctively specifically, what *always* happens. It is probably these facts which make one hesitate to call propositions about the action of substances ‘laws of nature’. The law of inertia, for example, would hardly be glossed: ‘If a body accelerates without any force acting on it, there must be some cause of its doing so.’ (Though I wonder what the author of *Principia* himself would have thought of that.) On the other hand just such ‘laws’ as that about a substance’s flash-point are connected with the match’s igniting because struck.

Returning to the medical example, medicine is of course not interested in the hopeless task of constructing lists of all the sets of conditions under each of which people always get a certain disease. It is interested in finding what that is special, if anything, is always the case when people get a particular disease; and, given such a cause or condition (or in any case), in finding circumstances in which people don’t get the disease, or tend not to. This is connected with medicine’s concern first, and last, with things as they happen in the messy and

mixed up conditions of life: only between its first and its last concern can it look for what happens unaffected by uncontrolled and inconstant conditions.

II

Yet my argument lies always open to the charge of appealing to ignorance. I must therefore take a different sort of example.

Here is a ball lying on top of some others in a transparent vertical pipe. I know how it got there: it was forcibly ejected with many others out of a certain aperture into the enclosed space above a row of adjacent pipes. The point of the whole construction is to show how a totality of balls so ejected always build up in rough conformity to the same curve. But I am interested in this one ball. Between its ejection and its getting into this pipe, it kept hitting sides, edges, other balls. If I made a film of it I could run it off in slow motion and tell the impact which produced each stage of the journey. Now was the result necessary? We would probably all have said it was in the time when Newton's mechanics was undisputed for truth. It was the impression made on Hume and later philosophers by that mechanics, that gave them so strong a conviction of the iron necessity with which everything happens, the 'absolute fate' by which 'Every object is determin'd to a certain degree and direction of its motion' (*A Treatise of Human Nature*, Book II, Part III, Section I).

Yet no one could have deduced the resting place of the ball – because of the indeterminateness that you get even in the Newtonian mechanics, arising from the finite accuracy of measurements. From exact figures for positions, velocities, directions, spins and masses you might be able to calculate the result as accurately as you chose. But the minutest inexactitudes will multiply up factor by factor, so that in a short time your information is gone. Assuming a given margin of error in your initial figure, you could assign an associated probability to that ball's falling into each of the pipes. If you want the highest probability you assign to be really high, so that you can take it as practical certainty, it will be a problem to reckon how tiny the permitted margins of inaccuracy must be – analogous to the problem: how small a fraction of a grain of millet must I demand is put on the first square of the chess board, if after doubling up at every square I end up having to pay out only a pound of millet? It would be a figure of such smallness as to have no meaning as a figure for a margin of error.

However, so long as you believed the classical mechanics you might also think there could be no such thing as a figure for a difference that had no meaning. Then you would think that though it was not feasible for us to find the necessary path of the ball because our margins of error are too great, yet there *was* a necessary path, which could be assigned a sufficient probability for firm acceptance of it, by anyone (not one of us) capable of reducing his limits of accuracy in measurement to a sufficiently small compass. Admittedly, so small a compass that he'd be down among the submicroscopic particles and no longer concerned with the measurements, say, of the ball. And now we can say: with certain degrees of smallness we get to a region for which Newton's mechanics is no longer believed.

If the classical mechanics can be used to calculate a certain real result, we may give a sense to, and grant, the ‘necessity’ of the result, given the antecedents. Here, however, you can’t use the mechanics to calculate the result, but at most to give yourself a belief in its necessity. For this to be reasonable the system has got to be acknowledged as true. Not, indeed, that that would be enough; but if so much were secured, then it would be worthwhile to discuss the metaphysics of absolute measures of continuous quantities.

The point needs some labouring precisely because ‘the system does apply to such bodies’ – that is, to moderately massive balls. After all, it’s Newton we use to calculate Sputniks! ‘The system applies to these bodies’ is true only in the sense and to the extent that it yields sufficient results of calculations about these bodies. It does not mean: in respect of these bodies the system is the truth, so that it just doesn’t matter that we can’t use it to calculate such a result in such a case. I am not saying that a deterministic system involves individual predictability: it evidently does not. But in default of predictability the determinedness declared by the deterministic system has got to be believed because the system itself is believed.

I conclude that we have no ground for calling the path of the ball determined – at least, until it has taken its path – but, it may be objected, is not each stage of its path determined, even though we cannot determine it? My argument has partly relied on loss of information through multiplicity of impacts. But from one impact to the next the path is surely determined, and so the whole path is so after all.

It sounds plausible to say: each stage is determined and so the whole is. But what does ‘determined’ mean? The word is a curious one (with a curious history); in this sort of context it is often used as if it *meant* ‘caused’. Or perhaps ‘caused’ is used as if it meant ‘determined’. But there is at any rate one important difference – a thing hasn’t been caused until it has happened; but it may be determined before it happens.

(It is important here to distinguish between being *determined* and being *determinate*. In indeterministic physics there is an apparent failure of both. I am concerned only with the former.)

When we call a result determined we are implicitly relating it to an antecedent range of possibilities and saying that all but one of these is disallowed. What disallows them is not the result itself but something antecedent to the result. The antecedences may be logical or temporal or in the order of knowledge. Of the many – antecedent – possibilities, *now* only one is – antecedently – possible.

Mathematical formulae and human decisions are limiting cases; the former because of the obscurity of the notion of antecedent possibilities, and the latter because decisions can be retrieved.

In a chess-game, the antecedent possibilities are, say, the powers of the pieces. By the rules, a certain position excludes all but one of the various moves that were in that sense antecedently possible. This is logical antecedence. The next move is determined.

In the zygote, sex and eye-colour are already determined. Here the antecedent possibilities are the possibilities for sex and eye-colour for a child; or more narrowly: for a child of these parents. *Now*, given the combination of this ovum and this spermatozoon, all but one of these antecedent possibilities is excluded.

It might be said that anything was determined once it had happened. There is now no possibility open: it *has* taken place! It was in this sense that Aristotle said that past and present were necessary. But this does not concern us: what interests us is *pre*-determination.

Then ‘each stage of the ball’s path is determined’ must mean ‘Upon any impact, there is only one path possible for the ball up to the next impact (and assuming no air currents, etc.).’ But what ground could one have for believing this, if one does not believe in some system of which it is a consequence? Consider a steel ball dropping between two pins on a Galton board to hit the pin centred under the gap between them. That it should balance on this pin is not to be expected. It has two possibilities; to go to the right or to the left. If you have a system which forces this on you, you can say: ‘There has to be a determining factor; otherwise, like Buridan’s ass, the ball must balance.’ But if you have not, then you should say that the ball may be undetermined until it does move to the right or the left. Here the ball had only two significant possibilities and was perhaps unpredetermined between them. This was because it cannot be called determined – no reasonable account can be given of insisting that it is so – within a small range of possibility, actualization within which will lead on to its falling either to the right or to the left. With our flying ball there will also be such a small range of possibility. The further consequences of the path it may take are not tied down to just two significant possibilities, as with one step down the Galton board: the range of further possibility gets wider as we consider the paths it may take. Otherwise, the two cases are similar.

We see that to give content to the idea of something’s being determined, we have to have a set of possibilities, which something narrows down to one – before the event.

This accords well with our understanding of part of the dissatisfaction of some physicists with the quantum theory. They did not like the undeterminedness of individual quantum phenomena. Such a physicist might express himself by saying ‘I believe in causality!’ He means: ‘I believe that the real physical laws and the initial conditions must entail uniqueness of result.’ Of course, within a range of co-ordinate and mutually exclusive identifiable possible results, only one happens: he means that the result that happens ought to be understood as the only one that was possible before it happened.

Must such a physicist be a ‘determinist’? That is, must he believe that the whole universe is a system such that, if its total states at t and t' are thus and so, the laws of nature are such as then to allow only one possibility for its total state at any other time? No. He may not think that the idea of a total state of the universe at a time is one he can do anything with. He may even have no views on the uniqueness of possible results for whatever may be going on in any arbitrary volume of space. For ‘Our theory should be such that only the actual result was possible for that experiment’ doesn’t mean ‘Our theory should be such that only this result was possible as *the result of the experiment*.’ He hates a theory, even if he has to put up with it for the time being, that essentially assigns only probability to a result, essentially allows of a range of possible results, never narrowed down to one until the event itself.

It must be admitted that such dissatisfied physicists very often have been determinists. Witness Schrödinger's account of the 'principle of causality': 'The exact physical situation at *any* point P at a given moment t is unambiguously determined by the exact physical situation within a certain surrounding of P at any previous time, say $t - \Delta t$. If Δt is large, that is, if that previous time lies far back, it may be necessary to know the previous situation for a wide domain around P '.³ Or Einstein's more modest version of a notorious earlier claim: if you knew all about the contents of a sphere of radius 186,000 miles, and knew the laws, you would be able to know for sure what would happen at the centre for the next second. Schrödinger says: *any* point P ; and *a* means *any* sphere of that radius. So their view of causality was not that of my hypothetical physicist, who I said may not have views on the uniqueness of possible results for whatever may be going on in any arbitrary volume of space. My physicist restricts his demand for uniqueness of result to situations in which he has got certain processes going in isolation from inconstant external influences, or where they do not matter, as the weather on a planet does not matter for predicting its course round the sun.

The high success of Newton's astronomy was in one way an intellectual disaster: it produced an illusion from which we tend still to suffer. This illusion was created by the circumstance that Newton's mechanics *had a good model in the solar system*. For this gave the impression that we had here an ideal of scientific explanation; whereas the truth was, it was mere obligingness on the part of the solar system, by having had so peaceful a history in recorded time, to provide such a model. For suppose that some planet had at some time erupted with such violence that its shell was propelled rocket-like out of the solar system. Such an event would not have violated Newton's laws; on the contrary, it would have illustrated them. But also it would not have been calculable as the past and future motions of the planets are presently calculated on the assumption that they can be treated as the simple 'bodies' of his mechanics, with no relevant properties but mass, position and velocity and no forces mattering except gravity.

Let us pretend that Newton's laws were still to be accepted without qualification: no reserve in applying them in electrodynamics; no restriction to bodies travelling a good deal slower than light; and no quantum phenomena. Newton's mechanics is a deterministic system; but this does not mean that believing them commits us to determinism. We could say: of course nothing violates those axioms or the laws of the force of gravity. But animals, for example, run about the world in all sorts of paths and no path is dictated for them by those laws, as it is for planets. Thus in relation to the solar system (apart from questions like whether in the past some planet has blown up), the laws are like the rules of an infantile card game: once the cards are dealt we turn them up in turn, and make two piles each, one red, one black; the winner has the biggest pile of red ones. So once the cards are dealt the game is determined, and from any position in it you can derive all others back to the deal and forward to win or draw. But in relation to what happens on and inside a planet the laws are, rather, like the rules of chess; the play is seldom determined, though nobody breaks the rules.⁴

Why this difference? A natural answer is: the mechanics does not give the special laws of all the forces. Not, for example, for thermal, nuclear, electrical,

chemical, muscular forces. And now the Newtonian model suggests the picture: given the laws of all the forces, then there is total coverage of what happens and then the whole game of motion is determined; for, by the first law, any acceleration implies a force of some kind, and must not forces have laws? My hypothetical physicist at least would think so; and would demand that they be deterministic. Nevertheless he still does not have to be a ‘determinist’; for many forces, unlike gravity, can be switched on and off, are generated, and also shields can be put up against them. It is one thing to hold that in a clear-cut situation – an astronomical or a well-contrived experimental one designed to discover laws – ‘the result’ should be determined: and quite another to say that in the hurly-burly of many crossing contingencies whatever happens next must be determined; or to say that the generation of forces (by human experimental procedures, among other things) is always determined in advance of the generating procedure; or to say that there is always a law of composition, of such a kind that the combined effect of a set of forces is determined in every situation.

Someone who is inclined to say those things, or implicitly to assume them, has almost certainly been affected by the impressive relation between Newton’s mechanics and the solar system.

We remember how it was in mechanics. By knowing the position and velocity of a particle at one single instant, by knowing the acting forces, the whole future path of the particle could be foreseen. In Maxwell’s theory, if we know the field at one instant only, we can deduce from the equations of the theory how the whole field will change in space and time. Maxwell’s equations enable us to follow the history of the field, just as the mechanical equations enabled us to follow the history of material particles. . . . With the help of Newton’s laws we can deduce the motion of the earth from the force acting between the sun and the earth.⁵

‘By knowing the acting forces’ – that must of course include the *future* acting forces, not merely the present ones. And similarly for the equations which enable us to follow the history of the field; a change may be produced by an external influence. In reading both Newton and later writers one is often led to ponder that word ‘external’. Of course, to be given ‘the acting forces’ is to be given the external forces too and any new forces that may later be introduced into the situation. Thus those first sentences are true, if true, without the special favour of fate, being general truths of mechanics and physics, but the last one is true by favour, by the brute fact that only the force acting between earth and sun matters for the desired deductions.

The concept of necessity, as it is connected with causation, can be explained as follows: a cause *C* is a necessitating cause of an effect *E* *when* (I mean: on the occasions when) if *C* occurs it is certain to cause *E* unless something prevents it. *C* and *E* are to be understood as general expressions, not singular terms. If ‘certainty’ should seem too epistemological a notion: a necessitating cause *C* of a given kind of effect *E* is such that it *is not* possible (on the occasion) that *C* should occur and should not cause an *E*, given that there is nothing that prevents an *E* from occurring. A non-necessitating cause is then one that can fail of its effect without the intervention of anything to frustrate it. We may discover

types of necessitating and non-necessitating cause; e.g. rabies is a necessitating cause of death, because it is not possible for one who has rabies to survive without treatment. We don't have to tie it to the occasion. An example of a non-necessitating cause is mentioned by Feynman: a bomb is connected with a Geiger counter, so that it will go off if the Geiger counter registers a certain reading; whether it will or not is not determined, for it is so placed near some radioactive material that it may or may not register that reading.

There would be no doubt of the cause of the reading or of the explosion if the bomb did go off. Max Born is one of the people who has been willing to dissociate causality from determinism: he explicates cause and effect in terms of dependence of the effect on the cause. It is not quite clear what 'dependence' is supposed to be, but at least it seems to imply that you would not get the effect without the cause. The trouble about this is that you might – from some other cause. That this effect was produced by this cause does not at all show that it could not, or would not, have been produced by something else in the absence of this cause.

Indeterminism is not a possibility unconsidered by philosophers. C. D. Broad, in his inaugural lecture, given in 1934, described it as a possibility; but added that whatever happened without being determined was accidental. He did not explain what he meant by being accidental; he must have meant more than not being necessary. He may have meant being uncaused; but, if I am right, not being determined does not imply not being caused. Indeed, I should explain indeterminism as the thesis that not all physical effects are necessitated by their causes. But if we think of Feynman's bomb, we get some idea of what is meant by 'accidental'. It was random: it 'merely happened' that the radioactive material emitted particles in such a way as to activate the Geiger counter enough to set off the bomb. Certainly the motion of the Geiger counter's needle is caused; and the actual emission is caused too; it occurs because there is this mass of radioactive material here. (I have already indicated that, contrary to the opinion of Hume, there are many different sorts of causality.) But all the same the *causation* itself is, one could say, *mere hap*. It is difficult to explain this idea any further.

Broad used the idea to argue that indeterminism, if applied to human action, meant that human actions are 'accidental'. Now he had a picture of choices as being determining causes, analogous to determining physical causes, and of choices in their turn being either determined or accidental. To regard a choice as such – i.e. any case of choice – as a predetermining causal event, now appears as a naive mistake in the philosophy of mind, though that is a story I cannot tell here.

It was natural that when physics went indeterministic, some thinkers should have seized on this indeterminism as being just what was wanted for defending the freedom of the will. They received severe criticism on two counts: one, that this 'mere hap' is the very last thing to be invoked as the physical correlate of 'man's ethical behaviour'; the other, that quantum laws predict statistics of events when situations are repeated; interference with these, by the *will's* determining individual events which the laws of nature leave undetermined, would be as much a violation of natural law as would have been interference which falsified a deterministic mechanical law.

Ever since Kant it has been a familiar claim among philosophers, that one can

believe in both physical determinism and 'ethical' freedom. The reconciliations have always seemed to me either to be so much gobbledegook, or to make the alleged freedom of action quite unreal. My actions are mostly physical movements; if these physical movements are physically predetermined by processes which I do not control, then my freedom is perfectly illusory. The truth of physical indeterminism is thus indispensable if we are to make anything of the claim to freedom. But certainly it is insufficient. The physically undetermined is not thereby 'free'. For freedom at least involves the power of acting according to an idea, and no such thing is ascribed to whatever is the subject (what would be the relevant subject?) of unpredictability in indeterministic physics. Nevertheless, there is nothing unacceptable about the idea that that 'physical haphazard' should be the only physical correlate of human freedom of action; and perhaps also of the voluntariness and intentionality in the conduct of other animals which we do not call 'free'. The freedom, intentionality and voluntariness are not to be analysed as the same thing as, or as produced by, the physical haphazard. Different sorts of pattern altogether are being spoken of when we mention them, from those involved in describing elementary processes of physical causality.

The other objection is, I think, more to the point. Certainly if we have a statistical law, but undetermined individual events, and then enough of these are supposed to be pushed by will in one direction to falsify the statistical law, we have again a supposition that puts will into conflict with natural laws. But it is not at all clear that the same train of minute physical events should have to be the regular correlate of the same action; in fact, that suggestion looks immensely implausible. It is, however, required by the objection.

Let me construct an analogy to illustrate this point. Suppose that we have a large glass box full of millions of extremely minute coloured particles, and the box is constantly shaken. Study of the box and particles leads to statistical laws, including laws for the random generation of small unit patches of uniform colour. Now the box is remarkable for also presenting the following phenomenon: the word 'Coca-Cola' formed like a mosaic, can always be read when one looks at one of the sides. It is not always the same shape in the formation of its letters, not always the same size or in the same position, it varies in its colours; but there it always is. It is not at all clear that those statistical laws concerning the random motion of the particles and their formation of small unit patches of colour would have to be supposed violated by the operation of a cause for this phenomenon which did not derive it from the statistical laws.

It has taken the inventions of indeterministic physics to shake the rather common dogmatic conviction that determinism is a presupposition, or perhaps a conclusion, of scientific knowledge. Not that that conviction has been very much shaken even so. Of course, the belief that the laws of nature are deterministic has been shaken. But I believe it has often been supposed that this makes little difference to the assumption of macroscopic determinism: as if undeterminedness were always encapsulated in systems whose internal workings could be described only by statistical laws, but where the total upshot, and in particular the outward effect, was as near as makes no difference always the same. What difference does it make, after all, that the scintillations, whereby my watch dial is luminous, follow

only a statistical law – so long as the gross manifest effect is sufficiently guaranteed by the statistical law? Feynman's example of the bomb and Geiger counter smashes this conception; but as far as I can judge it takes time for the lesson to be learned. I find deterministic assumptions more common now among people at large, and among philosophers, than when I was an undergraduate.

The lesson is welcome, but indeterministic physics (if it succeeds in giving the lesson) is only culturally, not logically, required to make the deterministic picture doubtful. For it was always a mere extravagant fancy, encouraged in the 'age of science' by the happy relation of Newtonian mechanics to the solar system. It ought not to have mattered whether the laws of nature were or were not deterministic. For them to be deterministic is for them, together with the description of the situation, to entail unique results in situations defined by certain relevant objects and measures, and where no part is played by inconstant factors external to such definition. If that is right, the laws' being deterministic does not tell us whether 'determinism' is true. It is the total coverage of every motion that happens, that is a fanciful claim. But I do not mean that any motions lie outside the scope of physical laws, or that one cannot say, in any given context, that certain motions would be violations of physical law. Remember the contrast between chess and the infantile card game.

Meanwhile in non-experimental philosophy it is clear enough what are the dogmatic slumbers of the day. It is over and over again assumed that any singular causal proposition implies a universal statement running 'Always when this, then that'; often assumed that true singular causal statements are derived from such 'inductively believed' universalities. Examples indeed are recalcitrant, but that does not seem to disturb. Even a philosopher acute enough to be conscious of this, such as Davidson, will say, without offering any reason at all for saying it, that a singular causal statement implies *that there is* such a true universal proposition⁶ – though perhaps we can never have knowledge of it. Such a thesis needs some reason for believing it! 'Regularities in nature': that is not a reason. The most neglected of the key topics in this subject are: interference and prevention.

Notes

- 1 My colleague Ian Hacking has pointed out C. S. Peirce to me as an exception to this generalization.
- 2 Bertrand Russell, *Mysticism and Logic* (New York: Longmans, Green, 1918).
- 3 Erwin Schrödinger, 'Nature and the Greeks' and 'Science and Humanism' (Cambridge: Cambridge University Press, 1996), pp. 132–3.
- 4 I should have made acknowledgements to Gilbert Ryle (*Concept of Mind* (London: Hutchinson, 1949) p. 77) for this comparison. But his use of the openness of chess is somewhat ambiguous and is not the same as mine. For the contrast with a closed card game I was indebted to A. J. P. Kenny.
- 5 Albert Einstein and Leopold Infeld, *The Evolution of Physics* (New York, 1938; paperback edn 1967), p. 146.
- 6 'Causal Relations', *Journal of Philosophy*, 64 (November 1967).

PART TWO

WHAT IS OUR PLACE IN THE WORLD?

Introduction

How is the Appearance of a Thing Related to the Thing that Appears?

- 31 The Theory of Sensa: an Excerpt from *Scientific Thought*
C. D. BROAD
- 32 Qualities: an Excerpt from *Consciousness and Causality*
D. M. ARMSTRONG
- 33 The Status of Appearances: an Excerpt from *Theory of Knowledge* (1st edition)
RODERICK M. CHISHOLM

What is the Relation between Mind and Body?

- 34 Which Physical Thing Am I? An Excerpt from “Is There a Mind–Body Problem?”
RODERICK M. CHISHOLM
- 35 Personal Identity: a Materialist Account
SYDNEY SHOEMAKER
- 36 Divided Minds and the Nature of Persons
DEREK PARFIT
- 37 Personal Identity: the Dualist Theory
RICHARD SWINBURNE
- 38 The Puzzle of Conscious Experience
DAVID J. CHALMERS

Is it Possible for Us to Act Freely?

- 39 Free Will as Involving Determination and
Inconceivable without It
R. E. HOBART
- 40 Human Freedom and the Self
RODERICK M. CHISHOLM
- 41 The Mystery of Metaphysical Freedom
PETER VAN INWAGEN
- 42 The Agent as Cause
TIMOTHY O'CONNOR

Introduction

Some readers may have found the air a bit thin while exploring the metaphysics of universals, time and space, change, and causation. The subject matter of this section brings metaphysical reflection to bear on questions of more immediate importance to *us*. Where do we fit into reality?

Reflections on one's place in the world are most compellingly presented in the first-person-singular. For the moment, then, we shall follow the example of Descartes in his *Meditations* and use the first-person-singular to ask: What is *my* place in the world?

A How is the Appearance of a Thing Related to the Thing that Appears?

I can divide the world up into two parts: there is myself, and there is the rest of the world – all the individuals that are not myself nor any part of myself. There is, moreover, a distinction to be drawn between those properties or states that are wholly or partly dependent upon myself (such as my size and shape and weight, my beliefs, etc.), and those properties and states that are dependent neither upon me, nor upon what I am like intrinsically (such as the size and shape and weight of the moon). But where do the *appearances* of things fall with respect to these divisions?

C. D. Broad's theory of "sensa" (or "sense data") takes the appearances I experience (such as the elliptical appearance presented by a penny) to be parts of the world outside me – they are individuals in their own right, and not parts of me. But, as Broad admits, they are peculiar individuals, probably lasting no longer than I experience them; their properties and states would seem to be almost entirely dependent upon me. D. M. Armstrong agrees that the appearances I experience are not parts of me; but, in the case of veridical (non-illusory) perception, he goes further than Broad, denying that they are dependent upon me at all. The real colors, shapes, sounds, and smells I sense are "out there" in the world, and would be there even if I were not here to sense them. When things appear other than they are, the appearances are, he says, nowhere – they are literally nonexistent. Roderick Chisholm advocates an "adverbial" theory of appearances (a view with affinities to what Broad calls "the Multiple Relation Theory"). According to Chisholm, the substantive term "appearance" is misleading, suggesting that appearances are *things* to which we are somehow related in sensation. There are no such things as appearances; there are only individuals undergoing various states of "being appeared to." Although appearances are neither parts of me nor parts of the rest of the world, the ways I am being appeared to are dependent upon me, as are all my intrinsic states (my shape, size, etc.).

Armstrong invokes a "causal theory of the mind." Section B of this part includes "The Puzzle of Conscious Experience," in which David Chalmers argues that this sort of theory will inevitably leave something out – namely, the

way it feels to be in this or that sensory state. Chisholm's transformation of statements about appearances into statements about how one is "appeared to" provides another example of the use of paraphrase to avoid unwanted ontological commitments (see section 3 of the "Introduction: What is Metaphysics?", above).

Suggestions for Further Reading

- Aune, Bruce, *Metaphysics: the Elements* (Minneapolis: University of Minnesota, 1985), ch. 8: "Appearance and Reality."
- Carter, William R., *The Elements of Metaphysics* (Philadelphia, Penn.: Temple University Press, 1990), ch. 2: "Idealism."
- Hales, Steven D., *Metaphysics: Contemporary Readings* (Belmont, Cal.: Wadsworth, 1998), Section 6: "Secondary Qualities."
- Hamlyn, D. W., *Metaphysics* (Cambridge, UK: Cambridge University Press, 1984), ch. 2: "Appearance and Reality."
- Hasker, William, *Metaphysics: Constructing a World View* (Downers Grove, Ill., and Leicester, UK: InterVarsity Press, 1983), ch. 4: "The World."
- Jubien, Michael, *Contemporary Metaphysics* (Oxford: Blackwell, 1997), ch. 6: "Color."
- Price, H. H., *Perception* (London: Methuen, 1932), chs 1–5.
- Russell, Bertrand, *The Problems of Philosophy* (New York: H. Holt, 1912), chs 1–4: "Appearance and Reality," "The Existence of Matter," "The Nature of Matter," "Idealism."
- van Inwagen, Peter, *Metaphysics* (Boulder, Col.: Westview Press, 1993), ch. 3: "Externality."

B What is the Relation between Mind and Body?

There are really two questions here: (i) What is the relationship between *me* (the thing that is thinking these thoughts, asking itself these questions, and experiencing certain sensations right now) and this body? (ii) What is the relationship between my mental states (the thoughts I'm now thinking, the sensations I'm now experiencing) and the physical states of this body (its shape, its weight, the pattern of neural firings going off in its brain, etc.). The first question admits only two plausible answers: Either I'm identical with this body or some part of it (say, my brain), or I'm not a physical body of any sort. Call the first answer "substance materialism," and the second "substance dualism." The second question also admits only two plausible answers: With respect to each type of mental state (say, a certain sort of intense pain), either the state is identical with some physical state or it is not. Call the latter "mental state materialism" and the former "mental state dualism." Clearly, mental state materialism is incompatible with substance dualism. But the other three combinations of these four views seem, on the face of it, to be coherent.

These views raise several difficult questions, however. For instance, what does it mean to call a state "physical"? Is it just to say that it can be possessed by a physical object? But then the combination of substance materialism and mental

state dualism would be impossible. On the other hand, we might restrict “physical” to mean “something mentioned in the complete, true science of the ultimate constituents of matter” – call this use of the term “physics-physical.” Nothing but spatiotemporal relations and simple theoretical properties like mass and charge would count as physics-physical. But then, if we interpret mental state dualism and materialism using this notion of “physical,” *being in pain* would almost certainly not qualify as a physical state of a dog, cat, or human being. After all, not even a table’s *having four legs* and *being made of wood* will be physics-physical states. So the restriction of physical states to just those that show up in fundamental physics threatens to make mental state materialism trivially false.

A more interesting sense of “physical” would be this: If a property or state is physical, then whether a thing has it is completely determined by the physics-physical states of the thing’s parts. The table’s having four legs, for instance, is a physical property because anything made out of similar particles similarly arranged would also have to have four legs. (Contemporary philosophers would make this point by saying that physical states are those that “supervene upon” physics-physical states.)

Interpreting mental state dualism and materialism in this way, we should probably conclude that the essays by Roderick Chisholm and David Chalmers represent the combination of substance materialism and mental state dualism. Sydney Shoemaker defends substance materialism and (elsewhere) mental state materialism. And Richard Swinburne defends substance dualism and mental state dualism. (In “Qualities: an Excerpt from *Consciousness and Causality*” in this volume, D. M. Armstrong defends a version of mental state materialism.)

Much of the action here revolves around the question: What sorts of changes could I survive, and what changes would bring about my destruction? Shoemaker defends the view that, although persons are large-scale material objects (living human bodies), the conditions that determine whether or not someone survives are largely *psychological*. Both Chisholm (in the selection below, and also in “Identity Through Time,” above) and Swinburne argue that, given the sorts of changes a person can survive, a person cannot be identical with any “gross physical body.” Derek Parfit alleges that we are radically mistaken about the kinds of changes we can survive – we are not at all as we take ourselves to be.

Connections with readings by Chisholm and Armstrong have already been mentioned. Both David Lewis’s “In Defense of Stages: Postscript B to ‘Survival and Identity’” and Peter Geach’s “Some Problems about Time” consider relations between the metaphysics of temporal parts and personal identity.

Suggestions for Further Reading

- Aune, Bruce, *Metaphysics: the Elements* (Minneapolis: University of Minnesota, 1985), ch. 5: “Changing Things.”
- Campbell, Keith, *Body and Mind*, 2nd edn (Notre Dame: University of Notre Dame, 1984).
- Carter, William R., *The Elements of Metaphysics* (Philadelphia, Penn.: Temple University

METAPHYSICS: THE BIG QUESTIONS

- Press, 1990), chs 3, 7, and 8: "Material Minds," "Personal Identity," and "Responsibility."
- Disch, Thomas M., *Echo Round his Bones* (New York: Pocket Books, 1979; first publication, 1969). A science-fiction novel supporting a Parfit-like view of personal identity.
- Gardner, Martin, *The Whys of a Philosophical Scrivener* (New York: Quill, 1983), chs 17, 18, and 19: "Immortality: Why I am Not Resigned," "Immortality: Why I Do Not Think It Strange," and "Immortality: Why I Do Not Think It Impossible."
- Hamlyn, D. W., *Metaphysics* (Cambridge, UK: Cambridge University Press, 1984), chs 8 and 9: "Minds" and "Persons and Personal Identity."
- Hasker, William, *Metaphysics: Constructing a World View* (Downers Grove, Ill., and Leicester, UK: InterVarsity Press, 1983), ch. 3: "Minds and Bodies."
- Jubien, Michael, *Contemporary Metaphysics* (Oxford: Blackwell, 1997), ch. 6: "Color."
- Miedaner, Terry, *The Soul of Anna Klane* (New York: Ballantine, 1978). A science-fiction novel within which substance dualism is empirically verified.
- Perry, John, *A Dialogue on Personal Identity and Immortality* (Indianapolis: Hackett, 1978).
- (ed.), *Personal Identity* (Berkeley, Cal.: University of California Press, 1975).
- Post, John F., *Metaphysics: a Contemporary Introduction* (New York: Paragon House, 1991), ch. 6: "Metaphysics and Human Being."
- Smith, Quentin, and L. Nathan Oaklander, *Time, Change and Freedom: an Introduction to Metaphysics* (London: Routledge, 1995), Dialogue 7: "Personal Identity."
- Taylor, Richard, *Metaphysics* (Englewood Cliffs, N.J.: Prentice-Hall, 1992), chs 2, 3, and 4: "Persons and Bodies," "Interactionism," and "The Mind as a Function of the Body."
- van Inwagen, Peter, *Metaphysics* (Boulder, Col.: Westview Press, 1993), chs 9 and 10: "The Nature of Rational Beings: Dualism and Physicalism" and "The Nature of Rational Beings: Dualism and Personal Identity."

C Is it Possible for Us to Act Freely?

Questions about freedom and determinism are not difficult to motivate, and little need be said by way of introduction to these readings. Hobart (a pseudonym of Dickinson S. Miller) defends "compatibilism" – the thesis that free will is compatible with determinism. Chisholm denies this, and claims that, when we act freely, we must be able to cause certain things to happen – and not just by virtue of our having mental states (beliefs and desires) that cause our behavior. Chisholm gives the name "immanent causation" to the causal relation supposed to hold between an agent and the events she causes but that aren't caused by any other events. (Note that Chisholm's distinction between "immanent" and "transeunt" causation is quite different from the distinction Susan Stebbing makes using these terms in "Causality: an Excerpt from *A Modern Introduction to Logic*," above. Stebbing's use corresponds much more closely to that of the philosophers who introduced the terms back in the Middle Ages.) Van Inwagen, in an essay new in this volume, agrees with Chisholm that freedom and determinism are incompatible. But he also emphasizes the difficulties in supposing that freedom is compatible with *indeterminism*. He denies that Chisholm's appeal to a special sort of causal relation between agent and event can overcome

this difficulty. O'Connor's paper, also published here for the first time, includes a response to van Inwagen's skepticism about the utility of "agent causation."

There is an obvious similarity between the "immanent" or "agent" causation posited by Chisholm and O'Connor, and Reid's notion of "Active Power" (see "Efficient Cause and Active Power: an Excerpt from *Essays on the Active Powers of the Human Mind*," above). A. N. Prior defends a connection between the freedom to do otherwise and theories of time in "Some Free Thinking about Time."

Suggestions for Further Reading

- Aune, Bruce, *Metaphysics: the Elements* (Minneapolis: University of Minnesota, 1985), ch. 9: "Metaphysical Freedom."
- Borges, Jorge Luis, "The Garden of Forking Paths," included in *Labyrinths: Selected Stories and Other Writings* (New York: New Directions, 1964).
- Carter, William R., *The Elements of Metaphysics* (Philadelphia, Penn.: Temple University Press, 1990), chs 9 and 10: "Causal Determinism" and "Fate."
- Gardner, Martin, *The Whys of a Philosophical Scrivener* (New York: Quill, 1983) ch. 6: "Free Will: Why I Am Not a Determinist or Haphazardist."
- Hasker, William, *Metaphysics: Constructing a World View* (Downers Grove, Ill., and Leicester, UK: InterVarsity Press, 1983), ch. 2: "Freedom and Necessity."
- Jubien, Michael, *Contemporary Metaphysics* (Oxford: Blackwell, 1997), ch. 7: "Determinism, Freedom, and Fatalism."
- Post, John F., *Metaphysics: A Contemporary Introduction* (New York: Paragon House, 1991), ch. 1: "Is Metaphysics Possible?"
- Smith, Quentin, and L. Nathan Oaklander, *Time, Change and Freedom: An Introduction to Metaphysics* (London: Routledge, 1995), Dialogues 9, 10, and 11: "Fatalism and Tenseless Time," "God, Time and Freedom," and "Freedom, Determinism and Responsibility."
- Taylor, Richard, *Metaphysics* (Englewood Cliffs, N.J.: Prentice-Hall, 1992), chs 5 and 6: "Freedom and Determinism" and "Fate."
- van Inwagen, Peter, *Metaphysics* (Boulder, Col.: Westview Press, 1993), ch. 11: "The Powers of Rational Beings: Freedom of the Will."
- Gary Watson (ed.), *Free Will* (Oxford: Oxford University Press, 1982).

How is the Appearance of a Thing Related to the Thing that Appears?

31 The Theory of Sensa: an Excerpt from *Scientific Thought**

C. D. Broad

. . . Difficulties always arise when two sets of properties apparently belong to the same object, and yet are apparently incompatible with each other. Now the difficulty here is to reconcile the supposed neutrality, persistence, and independence of a physical object with the obvious differences between its various sensible appearances to different observers at the same moment, and to the same observer at different moments between which it is held not to have undergone any physical change. We know, e.g., that when we lay a penny down on a table and view it from different positions it generally looks more or less elliptical in shape. The eccentricity of these various appearances varies as we move about, and so does the direction of their major axes. Now we hold that the penny, at which we say that we were looking all the time, has not changed; and that it is round, and not elliptical, in shape. This is, of course, only one example out of millions. It would be easy to offer much wilder ones; but it is simple and obvious, and involves no complications about a transmitting medium; so we will start with it as a typical case to discuss.

Now there is nothing in the mere ellipticity or the mere variation, taken by itself, to worry us. The difficulty arises because of the incompatibility between the apparent shapes and the supposed real shape, and between the change in the appearances and the supposed constancy of the physical object. We need not at present ask *why* we believe that there is a single physical object with these characteristics, which appears to us in all these different ways. It is a fact that we do believe it. It is an equally certain fact that the penny does look different as we move about. The difficulty is to reconcile the different appearances with the supposed constancy of the penny, and the ellipticity of most of the appearances with the supposed roundness of the penny. It is probable that at first sight the reader will not see much difficulty in this. He will be inclined to say that we can explain these various visual appearances by the laws of perspective, and so on. This is not a relevant answer. It is quite true that we can *predict what particular*

* From C. D. Broad, *Scientific Thought* (London: Routledge and Kegan Paul, 1923). Reprinted by permission of Routledge.

appearance an object will present to an observer, when we know the shape of the object and its position with respect to the observer. But this is not the question that is troubling us at present. Our question is as to the compatibility of these changing elliptical appearances, however they may be correlated with other facts in the world, with the supposed constancy and roundness of the physical object.

Now what I call *Sensible Appearance* is just a general name for such facts as I have been describing. It is important, here as always, to state the *facts* in a form to which everyone will agree, before attempting any particular *analysis* of them, with which it is certain that many people will violently disagree. The fundamental fact is that we constantly make such judgements as: 'This *seems to me* elliptical, or red, or hot,' as the case may be, and that about the truth of these judgements we do not feel the least doubt. We may, however, at the same time doubt or positively disbelieve that this *is* elliptical, or red, or hot. I may be perfectly certain at one and the same time that I have the peculiar experience expressed by the judgement: 'This looks elliptical to me,' and that in fact the object is not elliptical but is round.

I do not suppose that anyone, on reflection, will quarrel with this statement of fact. The next question is as to the right way to analyse such facts; and it is most important not to confuse the facts themselves with any particular theory as to how they ought to be analysed. We may start with a negative remark, which seems to me to be true, and is certainly of the utmost importance if it be true. Appearance is *not* merely mistaken *judgement* about physical objects. When I judge that a penny looks elliptical I am not mistakenly ascribing elliptical shape to what is in fact round. Sensible appearances *may* lead me to make a mistaken judgement about physical objects, but they *need* not, and, so far as we know, commonly do not. My certainty that the penny looks elliptical exists comfortably alongside of my conviction that it is round. But a mistaken judgement that the penny *is* elliptical would not continue to exist after I knew that the penny was really round. The plain fact is then that 'looking elliptical to me' stands for a peculiar experience, which, whatever the right analysis of it may be, is not just a mistaken judgement about the shape of the penny.

Appearance then cannot be described as mistaken judgement about the properties of some physical object. How are we to describe it, and can we analyse it? Two different types of theory seem to be possible, which I will call respectively the *Multiple Relation Theory*, and the *Object Theory* of sensible appearance. The Multiple Relation Theory takes the view that 'appearing to be so and so' is a unique kind of relation between an object, a mind, and a characteristic. (This is a rough statement, but it will suffice for the present.) On this type of theory to say that the penny looks elliptical to me is to say that a unique and not further analysable relation of 'appearing' holds between the penny, my mind, and the general characteristic of ellipticity. The essential point for us to notice at present about theories of this kind is that they do not imply that we are aware of *anything* that *really is* elliptical when we have the experience which we express by saying that the penny looks elliptical to us. . . .

Theories of the Object type are quite different. They do not involve a unique and unanalysable multiple relation of ‘*appearing*’, but a peculiar kind of object – an ‘*appearance*’. Such objects, it is held, actually *do have* the characteristics which the physical object *seems to have*. Thus the Object Theory analyses the statement that the penny looks to me elliptical into a statement which involves the actual existence of an elliptical object, which stands in a certain cognitive relation to me on the one hand, and in another relation, yet to be determined, to the round penny. This type of theory, though it has been much mixed up with irrelevant matter, and has never been clearly stated and worked out till our own day, is of respectable antiquity. The doctrine of ‘representative ideas’ is the traditional and highly muddled form of it. . . . In this book I shall deliberately confine myself to this type of theory, and shall try to state it clearly, and work it out in detail.

. . . On the theory that we are now going to discuss, whenever a penny looks to me elliptical, what really happens is that I am aware of an object which is, in fact, elliptical. This object is connected in some specially intimate way with the round physical penny, and for this reason is called an appearance *of* the penny. It really is elliptical, and for this reason the penny is said to look *elliptical*. We may generalise this theory of sensible appearance as follows: Whenever I truly judge that *x* appears to me to have the sensible quality *q*, what happens is that I am directly aware of a certain object *y*, which (*a*) really does have the quality *q*, and (*b*) stands in some peculiarly intimate relation, yet to be determined, to *x*. (At the present stage, for all that we know, *y* might sometimes be identical with *x*, or might be literally a part of *x*.) Such objects as *y* I am going to call *Sensa*. Thus, when I look at a penny from the side, what happens, on the present theory, is at least this: I have a sensation, whose object is an elliptical, brown sensum; and this sensum is related in some specially intimate way to a certain round physical object, viz., the penny.

Now I think it must at least be admitted that the sensum theory is highly plausible. When I look at a penny from the side I am certainly aware of *something*; and it is certainly plausible to hold that this something is elliptical in the same plain sense in which a suitably bent piece of wire, looked at from straight above, is elliptical. If, in fact, nothing elliptical is before my mind, it is very hard to understand why the penny should seem *elliptical* rather than of any other shape. I do not now regard this argument as absolutely conclusive, because I am inclined to think that the Multiple Relation theory can explain these facts also. But it is at least a good enough argument to make the sensum theory well worth further consideration.

Assuming that when I look at a penny from the side I am directly aware of something which is in fact elliptical, it is clear that this something cannot be identified with the penny, if the latter really has the characteristics that it is commonly supposed to have. The penny is supposed to be round, whilst the sensum is elliptical. Again, the penny is supposed to keep the same shape and size as we move about, whilst the sensa alter in shape and size. Now one and the same thing cannot, at the same time and in the same sense, be round and ellip-

tical. Nor can one and the same thing at once change its shape and keep its shape unaltered, if 'shape' be used in the same sense in both statements. Thus it is certain that, if there be sensa, they cannot in general be identified with the physical objects of which they are the appearances, if these literally have the properties commonly assigned to them. On the other hand, all that I ever come to know about physical objects and their qualities seems to be based upon the qualities of the sensa that I become aware of in sense-perception. If the visual sensa were not elliptical and did not vary in certain ways as I move about, I should not judge that I was seeing a round penny.

The distinction between sensum and physical object can perhaps be made still clearer by taking some wilder examples. Consider, e.g., the case of looking at a stick which is half in water and half in air. We say that it looks bent. And we certainly do not mean by this that we mistakenly judge it to be bent; we generally make no such mistake. We are aware of an object which is very much like what we should be aware of if we were looking at a stick with a physical kink in it, immersed wholly in air. The most obvious analysis of the facts is that, when we judge that a straight stick *looks* bent, we are aware of an object which really *is* bent, and which is related in a peculiarly intimate way to the physically straight stick. The relation cannot be that of identity; since the same thing cannot at once be bent and straight, in the same sense of these words. If there be *nothing* with a kink in it before our minds at the moment, why should we think then of kinks at all, as we do when we say that the stick looks bent? No doubt we can quite well mistakenly *believe* a property to be present which is really absent, when we are dealing with something that is only known to us indirectly, like Julius Cæsar or the North Pole. But in our example we are dealing with a concrete visible object, which is bodily present to our senses; and it is very hard to understand how we could seem to ourselves to *see* the property of bentness exhibited in a concrete instance, if in fact *nothing* was present to our minds that possessed that property.

As I want to make the grounds for the sensum theory as clear as possible, I will take one more example. Scientists often assert that physical objects are not 'really' red or hot. We are not at present concerned with the truth or falsehood of this strange opinion, but only with its application to our present problem. Let us suppose then, for the sake of argument, that it is true. When a scientist looks at a penny stamp or burns his mouth with a potato he has exactly the same sort of experience as men of baser clay, who know nothing of the scientific theories of light and heat. The visual experience seems to be adequately described by saying that each of them is aware of a red patch of approximately square shape. If such patches be not in fact red, and if people be not in fact aware of such patches, where could the notion of red or of any other colour have come from? The scientific theory of colour would have nothing to explain, unless people really are aware of patches under various circumstances which really do have different colours. The scientists would be in the position of Mr Munro's duchess, who congratulated herself that disbelief had become impossible, as the Liberal Theologians had left us nothing to disbelieve in. Thus we seem forced to the view that there are at least hot and coloured sensa; and, if we

accept the scientific view that physical objects are neither hot nor coloured, it will follow that sensa cannot be identified with physical objects.

The reader may be inclined to say, 'After all, these sensa are not real; they are mere appearances, so why trouble about them?' The answer is that you do not get rid of anything by labelling it 'appearance'. Appearances are as real in their own way as anything else. If an appearance were nothing at all, nothing would appear, and if nothing appeared, there would be nothing for scientific theories to account for. To put the matter in another way: Words like *real* and *reality* are ambiguous. A round penny and an elliptical visual sensum are not real in precisely the same sense. But both are real in the most general sense that a complete inventory of the universe must mention the one as much as the other. No doubt the kind of reality which is to be ascribed to appearances will vary with the particular type of theory as to the nature of sensible appearance that we adopt. On the present theory an appearance is a sensum, and a sensum is a particular existent, though it may be a short-lived one. On the Multiple Relation theory appearances have a very different type of reality. But *all* possible theories have to admit the reality, *in some sense*, of appearances; and therefore it is no objection to any particular theory that it ascribes a sort of reality to appearances.

I hope that I have now made fairly clear the grounds on which the sensum theory of sensible appearance has been put forward. Closely connected with it is a theory about the perception of physical objects, and we may sum up the whole view under discussion as follows: Under certain conditions I have states of mind called sensations. These sensations have objects, which are always concrete particular existents, like coloured or hot patches, noises, smells, etc. Such objects are called sensa. Sensa have properties, such as shape, size, hardness, colour, loudness, coldness, and so on. The existence of such sensa, and their presence to our minds in sensation, lead us to judge that a physical object exists and is present to our senses. To this physical object we ascribe various properties. These properties are not in general identical with those of the sensum which is before our minds at the moment. For instance, the *elliptical* sensum makes us believe in the existence of a *round* physical penny. Nevertheless, all the properties that we do ascribe to physical objects are based upon and correlated with the properties that actually characterise our sensa. The sensa that are connected with a physical object *x* in a certain specially intimate way are called the appearances of that object to those observers who sense these sensa. The properties which *x* is said to *appear to* have are the properties which those sensa that are *x*'s appearances *really do* have. Of course, the two properties may happen to be the same, e.g., when I look straight down on a penny, both the physical object and the visual appearance are round. Generally, however, there is only a correlation between the two. . . .

32 Qualities: an Excerpt from *Consciousness and Causality**

D. M. Armstrong

In this section I will argue that *perception*, as we experience it introspectively, is entirely qualityless. The only qualities involved are qualities, not of mental phenomena, but of the physical things perceived. In particular, I maintain that the so-called secondary qualities: colour, sound, taste, smell, heat and cold: qualities which have often been thought to be inner qualities, are in fact qualities of objective physical phenomena.

Philosophical discussions of perception instinctively gravitate towards colour, so let us yield to temptation and begin with it. I look out of my window and see the green leaves of a vine. Colour, once we give it our attention, is a very striking and arresting phenomenon. But it does not seem to be a mental phenomenon. If we consult perception, then its verdict is clear. The green colour is a property of the vine-leaves. Nor is it a property which involves any relations that the leaves have to any other object, whether objects in the field of view or the perceiver. It is an intrinsic property of the leaves. In what follows I will hold fast to this perceptual deliverance. It is not an indubitable given to be respected at all costs. There are no indubitable givens. But it is a plain deliverance of perception, and it is plausible that a true theory of perception will uphold this deliverance.

An opponent of the causal theory of the mind,¹ however, has the following interest in holding the deliverance false. Suppose it can be shown that the perceived greenness does not really qualify the leaves, but instead qualifies something mental. This will have the consequence that when we are aware of greenness we are *introspectively* aware of it. But the things, or properties of things, we are introspectively aware of are, presumably, all mental. Yet the concept of greenness, it seems perfectly clear, is not a causal concept. So there will be something mental which falsifies the causal theory of the mind.

There are two main lines of argument by which philosophers and others have sought to show that the greenness is not a quality of the leaves. The first turns upon the fact of perceptual illusion (and mental imagery), the second upon questions of scientific plausibility. I think that neither succeeds. But the second creates rather more worry for the causal theory, and requires relatively heroic measures to meet it.

For the present, however, let us consider perceptual illusion and imagery. It is possible for something to look green, although in fact the thing is not green

* From D. M. Armstrong and Norman Malcolm, *Consciousness and Causality* (Oxford: Blackwell, 1984). Reprinted by permission of the author.

at all. It is possible to have an hallucination, say of a green snake, although there is nothing in the physical world which corresponds, even inaccurately, to the hallucination. One can dream of, or form an image of, something green. In such cases, it is argued, greenness is present. At the same time, by hypothesis, it is not some ordinary physical object or surface which has the quality. It is then plausible to think that it is something mental which is green.

Once one has gone this far, it proves difficult to maintain that *anything* except mental things are green. The greenness of vine leaves is dismissed as a mere *façon de parler*. Vine-leaves are 'green' because they have the power to create in us mental phenomena which have the actual quality of greenness.

What are these mental phenomena? Here there is division. According to some philosophers, when, as we ordinarily say, we perceive a green leaf, perceiving it as green, it would be truer to say that it is the mental act of perceiving which has the green quality. We perceive greenly. According to another, historically more popular, view, we ought to postulate special mental objects, often called sense-data. It is sense-data which are the direct or immediate objects of perception and it is they which are the bearers of the green quality. (This second view seems very much preferable *phenomenologically*.)

But wherever we locate the quality of greenness in the mind, its presence there will falsify a purely causal theory of the mind.

In my view, what the causal theorist should do is to deny the first step in this argument. He should deny that when something physical looks green to somebody, but is not green, or where somebody images something green, then the sensory quality of greenness is present. The causal theorist can admit that there is a sense in which sensory illusion, or the having of such images, involves something green. But (a) the something is an ordinary physical something; and (b) it is a merely *intentional*, not a real, object.

Suppose that somebody believes that something which they cannot at the time see is green. This belief is compatible with the object existing, but not being green, or even with the object not existing at all. A natural view to take is that the belief is a structural state, a structure corresponding to the concepts involved in the belief. The state has an intentional object: the object's being green, or, if you like, the proposition that the object is green. It is the programme of the causal theory to give an account of this intentionality purely in terms of the causal role played by the mental constituents of the state and the internal organization of these constituents.

The obvious suggestion for the causal theory is to give a similar account of perception. A perception of something green will involve a green-sensitive element, that is to say, something which, in a normal environment, is characteristically brought into existence by green things, and which in turn permits the perceiver, if he should so desire, to discriminate by his behaviour the object from things which are not green. In the case of sensory illusion, a thing which is not green, but really is perceived, brings the green-sensitive element into existence in the mind. In hallucination, the 'thing perceived' has a merely intentional existence, and the green-sensitive element comes to be as a result of other causes. In having a mental image of something green, there is something in the

mind which resembles a perceptual green-sensitive element, but which lacks the causal powers associated with genuine green-sensitive elements.

A particular word about mental images before passing them by. An appealing thought is that having a perception (say that something is green) stands to having the corresponding image as holding a belief stands to merely entertaining the corresponding thought. It was suggested above [in a passage not included here] that, when a proposition is entertained without belief, a belief-like structure comes to be in the mind, but that the structure lacks the power to influence action which belief has. In being introspectively aware that we are having such a thought we are aware (i) of the resemblance to the corresponding belief-structure; yet also (ii) that *this* structure does not have the power to feed into our practical reasonings in the way that beliefs do. So, it may also be suggested, the having of an image is the having of a perception-like state, but where the ‘perception’ lacks the power to influence action. (Both thoughts and images are, or often are, under the control of the will in the way that beliefs and perceptions are not. So they are not suited to be a *guide* to the will, as beliefs and perceptions are.)

Whether we should actually *reduce* perceptions to a certain species of acquirings of beliefs (and so the having of images to a species of entertaining thoughts) is a further question. I incline to favour such a reduction, but it is controversial among philosophers of perception. What is perhaps a little less controversial, because weaker, and so may secure wider agreement, is this: perceptions are propositional in structure. To say this is not, of course, to say that they are in any way linguistic. But they do encode information (and misinformation) about the current state of the perceiver’s physical environment, his body, and the relations of the body to his environment. Whether this information acquired constitutes a set of beliefs (fading with extreme rapidity for the most part), or whether we should think of it as sub-doxastic, beneath belief, but capable of *causing* fully-fledged belief, is perhaps a minor matter which we can afford to bracket here.

Whether it be belief or not, this information (misinformation) gives the perceiver the capacity to react back upon the current environment and his own body if he should purpose so to react. Among the phenomena which support this informational account is the indeterminacy of perception. I can perceive that the vine-leaf is green, and even see what shade of green it is although I have no name for the shade. I can perceive its shape and size, and distance from me. But I cannot perceive its *exact* shade, its *exact* shape, size and distance. The perception is indeterminate in these respects. This naturally suggests the indeterminacy of most (all?) information.

(Those who postulate sense-data to explain the perception of the vine-leaf must either say that sense-data are indeterminate in nature, making sense-data paradoxical entities indeed, or else say that sense-data are perfectly determinate but that our apprehension of them is indeterminate. Against the latter way out the following may be said. Ordinary vine-leaves can be further investigated in individual cases and their colour, shape, size and so forth more exactly determined. Nobody has ever suggested how this may be done with sense-data. The

hypothesis of sense-data is therefore at a methodological disadvantage.)

What I have given is a very brief sketch of a theory of perception compatible with the causal theory of the mind. If it is along the right track, then there is no call to treat illusory or imaged sensible qualities, and in particular colour, sound, taste, smell, heat or cold, as actual qualities of actual entities. The case of colour was taken, but the argument goes through in the case of all the qualities. What we are dealing with here is misinformation, or, in the case of imaging, something like the entertaining of fantasies. The illusory qualities do not qualify anything.

From this perspective, therefore, sensory illusion and the having of images give us no reason to think that perceived or imaged colour, sound, taste, smell and so on qualify anything mental. We can be introspectively aware of the having of a perception (or image) of something green, or of a perception (or image) of something red, and introspectively aware of the difference between the two perceptions (or images). But why should we think that the introspective awareness is an awareness of two different *qualities*? Redness and greenness are two different qualities, of course, but they are qualities of things without. A green-sensitive element within need not be green, nor is it introspected as something green, nor indeed is it introspected as having any quality at all. It is introspected simply as something having sophisticated causal relationships to green things. A purely causal theory of the nature of perceptions (and images) can so far be sustained.

This result, however, is challenged by a more formidable set of arguments, drawn from scientific considerations, which seem to show that the secondary qualities cannot qualify external things. The importance of these arguments makes further comments desirable.

Physics and the Secondary Qualities

The traditional distinction between the primary and the secondary qualities became prominent in the sixteenth and seventeenth centuries with the rise of modern science. It is a product of that rise. Despite the attempt by Locke to give the distinction an *a priori* basis,² its foundation appears to be *a posteriori*. To thinkers such as Galileo, Newton and Boyle it seemed that some of the perceived qualities of objects played no role in determining the behaviour of the objects. That behaviour appeared to depend upon the shape, size, position, state of motion and mass of the objects. But such qualities as colour, sound, taste and smell seemed to play no role in 'the executive order of nature'. Boyle called the two groups of qualities the primary and the secondary qualities respectively. Among philosophers, at least, the terms have stuck.

The membership of the two groups has not always remained constant. The reason for this is that the distinction is always drawn relative to the science, in particular the physics, of the day. The situation may be illustrated by considering the case of degrees of heat. At a certain stage, many natural philosophers thought that degrees of heat in objects were different proportional quantities of

caloric fluid. Quantity of caloric was then for them a primary quality of objects. But this theory was superseded by the view that degrees of heat in the object are nothing but more or less violent motion of the constituent parts (in particular, the molecules) of the hot thing. So degrees of heat in objects can be explained without postulating any primary quality of the objects except motion of parts.

Apparently successful reductions of this sort effect an intellectual economy in the theory of the physical world, and to that extent are evidently to be desired. But they also create a problem. The perceived secondary qualities seem certainly to exist. It is difficult to treat them as being, in Berkeley's phrase, 'a false imaginary glare'.³ But if they exist, then they must qualify something. But what do they qualify? Nothing physical apparently. So they qualify something mental. They must be swept into the philosophical dustbin of the mind. They must qualify sense-data ('ideas'), or perceiving, or something of that kind.

In this way we are led to the view that the truth-condition for the statement that the vine-leaf is green is that it *looks* green to normal human perceivers, where 'looks green' is spelt out in such a way that the sensible quality of greenness qualifies something in the mind of the perceiver. And so for the rest of the secondary qualities.

This line of argument, rooted as it is in *a posteriori* scientific considerations, strikes me as having much greater force against the causal theory of the mind than the argument from sensory illusion and mental images. How should the causal theorist answer it? I will briefly consider two answers which I think are unsuccessful, and then propose another solution.

It has sometimes been suggested that we should accept the point made by physicists that such qualities as colour bestow no causal power upon the physical objects which have them, yet at the same time leave the qualities where they look to be: qualifying external objects. They are causally idle properties, getting a free ride from the objects which they qualify. This is a direct realist, non-reductive, but epiphenomenalist, theory of the secondary qualities.

This view leads to various difficulties. There are difficulties in conceiving how qualities such as colour attach to physical objects conceived of as modern physics conceives the objects. But the most serious difficulty for this view of the secondary qualities is its epiphenomenalism. If the qualities bestow no causal power, then our perceptions would be exactly the same whether or not these qualities existed. The green leaves stimulate my eyes and cause me to perceive them, but they do this because of the light-waves emitted from the object. It is the light-waves which take executive action. They in turn depend solely upon primary properties of the surfaces of the leaves. Why should we think that the leaves have any other properties? Our perceptions would be exactly the same whether or not the causally idle external greenness was there.

So much for a first attempt to escape the dilemma posed by the physicist's assault upon the secondary qualities. Can we defend the objectivity of the secondary qualities by advancing a causal theory of these properties? To do this, we must first say that for a physical surface to be green is for it to have the power to furnish green sensations to normal perceivers in standard conditions. This was more or less the course favoured by Locke.⁴ Powers, however, are very close to

dispositions. Indeed, we could also speak of a disposition to furnish green sensations to normal perceivers in standard conditions. As a result we can go on to give the same account of the power which I earlier gave [in a passage not included here] of dispositions. The power can be identified with the property or properties of the object in virtue of which the object exerts its power. This in turn permits an identification of the colour of the object with whatever physically respectable properties of the object are the causes of the sensation. In this way, sounds can be identified with sound-waves, heat with motion of molecules, and so on.

The bulge in the carpet is here pushed down in a very satisfactory manner. The problem, however, is that it immediately appears elsewhere. What account is to be given of sensations of green? This was no special problem for Locke, for whom immediately or directly apprehended greenness was a property of 'ideas' in the mind, 'ideas' being the forerunners of the modern sense-data. But what is an upholder of the causal theory of the mind to say about sensations of green, after adopting a Lockean theory of objective greenness? Only, apparently, that a sensation of green is that which is apt to be produced in us by a green surface. We thus have a tight circle of two concepts, each defined in terms of the other. This would not be a vicious circularity if we have an independent way of introducing the two together. But what is this way? The Lockean approach gives us no help.

It has been suggested to me that the difficulty here lies in developing the causal theory as a *conceptual* theory. Suppose, instead, that we develop the causal theory as a scientific theory. . . . We then define greenness as that which is apt for the production of sensations of greenness. And for a causal theory of sensations of greenness we simply pin our hopes on future science.

This policy of the promissory note would perhaps be justified if we could see some possible way in which a non-circular causal theory of sensations of greenness could then be developed. Since I at least am unable to see this, I will try to develop an alternative.

I believe that this alternative should still hold to the idea of reducing the secondary qualities to scientifically respectable primary qualities: reflectance properties of surfaces, properties of air-waves, motions of constituent molecules, and so on. This reduction has two great advantages. The first is a phenomenological advantage. The secondary qualities can be located where they appear to be. It is the surface of the leaf which is green, sounds can fill a room, smells hang around in them, tastes can inhere in the tasty body, water can be hot or cold, just as perception delivers. The second advantage is that the qualities can be treated as causally active, as bestowing active and passive power upon the objects which possess them. As a result, there is no epistemological problem how it is that we come to be aware of them. Coloured objects act upon us *in virtue of their colour*, and create in us a perception of their colour.

At the same time, however, a causal theory of the secondary qualities seems unsatisfactory, for the reasons which have just been given. We want to see how we can get a grip upon the secondary qualities quite independently of our grip upon sensations of such qualities. I propose, to this end, what might be tagged

a *Gestalt* theory. When in perception we are aware of the colour, sound, smell, taste, etc., of the physical things, then the qualities which we are aware of are complexes of physical properties. The perceived secondary qualities are primary qualities! But we are aware of them in a unified, *Gestalt*, manner, a manner which fails to reveal the primary nature of these properties.

As a preliminary model, consider that it is often possible to recognize, say, a certain complex shape without being able to analyse the shape and give the shape-formula. (The shapes might be shown for only a very short time.) A certain shape is instantaneously recognized as the same shape again, but the recognition-process does not appear, at least, to involve recognizing the shape-formula. Contemporary psychologists have spoken of the way that a stimulus from without is fitted to a template within. There could be knowledge that the right fit had been achieved, without knowing just what shape the thing perceived has which makes it fit the template rightly. Might not the secondary qualities be objective complexes of primary qualities which are recognized in this primitive manner?

However, there is a difficulty in this model. When complex shapes are shown to us quickly, and we then prove able to recognize that it is the same, or not the same, shape which is shown again, yet are unable to give the shape-formula, still we do recognize that it is *shapes* that we are dealing with. We attribute a shape to the object perceived, even if indeterminately. But when a surface looks green to me, I am not attributing the corresponding primary quality or, indeed, any primary quality. Yet I am attributing some property to it. So am I not attributing a property to the surface which is *different* from the corresponding primary quality?

The only way, I think, that this difficulty can be met is to have recourse to the topic-neutral⁵ manoeuvre. Why should it not be that we attribute a property of *some sort* to the surface of the leaf, a property detected by the eyes, but without any specification of the sort, thus leaving it open that the property is in fact a primary property? It may be added to this, along the lines of the Headless Woman illusion,⁶ that our visual inability to pick out that the quality involved is a primary quality inevitably generates the illusion that it is not a primary quality. Again, our inability to pick out that the property is a structured property inevitably generates the illusion that it is not a structured property.

I believe that this suggestion is along the right lines. But, at first sight, it faces an enormous phenomenological difficulty. Earlier, I spoke of the phenomenological advantages of locating the colour where it appears to be: on the surface of a leaf. A topic-neutral account of the *mental* is reasonably plausible, at any rate if the sensible qualities are extruded from the mind. But the sensible qualities themselves are the paradigms of concrete perceived qualities. How can a subclass of these qualities, the secondary qualities, be treated as qualities we know not what, later identified, as a result of scientific considerations, with the primary qualities?

My suggestion is that the illusion of concrete secondary quality is created in the following way. Phenomenologically, the secondary qualities lack structure, they do not appear to have any 'grain' as Wilfrid Sellars puts it. Nevertheless,

they have a huge multitude of systematic resemblances and differences *to each other*. Each secondary quality has a position in complex dimensional arrays of qualities. The phenomenon is most clearly evident in the relations of the colours to each other, but is present in the case of each sense. There even seem to be classificatory relationships which cross the senses: for instance, red is to hot as blue is to cold. It is this immensely complex network of perceived resemblances and differences which largely creates the impression that our acquaintance with the secondary qualities is acquaintance with definite qualities which are other than the primary properties. (I say ‘largely’, because the failure to perceive the identity with complexes of primary properties, and the failure to perceive any ‘grain’ in the secondary qualities also have their influence, in the style of the Headless Woman.)

Resemblance is an internal relation: it depends upon the nature of the things which resemble each other. (Some philosophers have denied this. I believe that they are wrong, but it is at least profoundly natural to treat resemblance as an internal relation.) We can perceive resemblance without perceiving the respect of resemblance, and if my view of the secondary qualities is correct that is what occurs in their case. The immensely complex dimensional classification of the secondary qualities, with all its degrees of resemblance, is a matter of perception of resemblances without grasping the basis of the resemblance in the primary qualities. (A basis which scientific investigation increasingly uncovers.) However, given our instinctive taking of resemblance to be an internal relation, a mere perception of resemblance suffices to generate the illusion that we have a concrete acquaintance with the qualities which sustain the resemblance. *A perception of the internal relation of resemblance generates the illusion of a perception of intrinsic quality.*

It is worth noticing that our awareness of the resemblances and differences of colour are sharper and clearer than in the case of the other secondary qualities. (We see straightaway that orange is between red and yellow.) And it is colour which gives us the strongest impression of acquaintance with the concrete nature of the quality involved.

I conclude, then, that it is possible to uphold an identification of the secondary qualities with primary qualities. If so, there is still no reason to bring the secondary qualities within the mind, or give an analysis of secondary qualities in terms of sensations of such qualities.

Two final points before concluding. First, as in the case of the identification of the mental with the physical, there may be problems in a straightforward identification of secondary qualities with primary qualities. It has been argued that the primary qualities correlated with the one shade of colour form an irreducibly disjunctive set. (This situation does not hold for all ranges of the secondary qualities, but it may for colour.) There seems to be no *a priori* reason why this should not be so. The human perceptual system might classify together surfaces which a mature science would not treat as having anything significant in common. If this were so, and provided the evidence did not impugn the whole project of the reduction of secondary qualities to primary qualities, then the colour when possessed by *this* surface, and no doubt the same colour when possessed by surfaces of the same sort, would be identified with a certain

primary property. But that colour *simpliciter* could not be identified with that primary property.

If all this were so, I think that one could simply accept it as a complication, but not as a refutation, of the account of the secondary qualities which I have proposed.

The second point is that one might wonder whether primary properties are in any more secure position than secondary qualities. Might it not turn out that they, too, are only known in terms of their resemblances and differences to the other primary qualities? If so, a topic-neutral analysis will have to be given of them, too. We will not be acquainted with intrinsic quality at any point.

Here again I would simply allow this speculation. Contemporary physics suggests that we should give an account of colour, sound, taste, smell, heat and cold in terms of the ‘executive’ primary properties. But who knows if the latter are fundamental? (Why should middle-sized creatures like ourselves be in perceptual touch with the fundamental properties of the world, if there are any?) A deeper physics might give an account of the current list of primary qualities in terms of properties which we can neither perceive nor image.

Notes

- 1 In the essay from which this excerpt is taken, Armstrong defends ‘the causal theory of the mind’, the view that mental concepts – such as those of belief, desire, sensation – are to be defined in terms of their causal roles. A causal theorist might hold, for instance, that it is part of the essence of pain that it be a state typically caused by bodily damage and typically causing a desire to be rid of it. Armstrong draws an analogy with the elasticity of a certain rubber band: its elasticity is that property of the band which plays a certain causal role – namely, being responsible for its stretching when pulled, and its returning to its original shape when the force is removed. The property that plays this role is some sort of micro-structural feature of the band. Similarly, the property that plays the ‘pain-role’ in human beings is (according to Armstrong) some neurological state. – [Eds]
- 2 John Locke, *Essay*, book II, ch. VIII, Section 9.
- 3 George Berkeley, *Second Dialogue*, in *Berkeley’s Philosophical Writings*, ed. D. M. Armstrong (Collier-Macmillan, 1965), p. 174.
- 4 Locke, *Essay*, book II, ch. VIII, Section 10.
- 5 Armstrong’s causal theory of the mind is ‘topic-neutral’ in that ‘it does not specify the nature of that which plays the causal role’. So the state which plays the ‘pain-role’, for instance, could be a neurological state (as the materialist supposes); but, at least in principle, the same role could have been played by some sort of ‘spiritual state’ within an immaterial soul. Armstrong thinks that ‘the topic neutral nature of our knowledge of mental phenomena’ helps explain why ‘the mental, as we actually experience it introspectively, is elusive, hard to pin down, as it were transparent or diaphanous’. ‘What is grasped only as something which plays a certain causal role is grasped transparently and inconclusively’ (*Consciousness and Causality*, p. 158). – [Eds]
- 6 ‘To produce this illusion, a woman is placed on a suitably illuminated stage draped all in black, and a black cloth is placed over her head. In these circumstances, the spectators do not perceive the woman’s head. But this real failure to see gives rise to their seeming to see that the woman lacks a head.’

Now it is perfectly plain that in introspection we are not aware of mental phenomena as material states and processes. The materialist can agree with the dualist about this. But it can be predicted that, as in the case of the Headless Woman, it will *seem* that what we are aware of are *not* material states and processes. Failure to be aware of materiality will naturally be interpreted as awareness of immateriality. The introspective implausibility of materialism is therefore no argument against materialism' (Armstrong, *Consciousness and Causality*, p. 158). – [Eds]

33 The Status of Appearances: an Excerpt from *Theory of Knowledge* (1st edition)*

Roderick M. Chisholm

The Problem of Democritus

When a man sees an external thing, say, a tree, his perception is the result of a complex physiological and psychological process. Light reflected from the thing stimulates the rod cells and cone cells in his eyes; in consequence of this stimulation, there is a further effect within the brain which, in turn, produces a visual sensation. Perception by means of the other sense organs is similar. In each case, the sensation (also referred to as the "sense impression," "appearance," "idea," or "sense datum") would seem to depend for its existence upon the state of the perceiving subject. Or to proceed somewhat more cautiously, the ways in which the things that we perceive *appear* to us when we perceive them depend in part upon our own psychological and physiological condition. This fact has led to some of the most puzzling questions of the theory of knowledge.

Democritus took it to imply not only that we do not perceive what it is that we think we perceive, but also that external things are not at all what we tend to believe that they are. The appearances of things, he said, "change with the condition of our body and the influences coming toward it or resisting it."¹ The question as to whether any particular thing will appear white, black, yellow, red, sweet, or bitter, he noted, cannot be answered merely by reference to the nature of the thing; one must also refer to the nature of the person or animal who is perceiving the thing. And from these premises, which are undeniable, Democritus then went on to infer (1) that no one ever *perceives* any external thing to be white, black, yellow, red, sweet, or bitter, and also (2) that no unperceived external thing *is*, in fact, white, black, yellow, red, sweet, or bitter.

* From Roderick M. Chisholm, *Theory of Knowledge*, © 1966. Reprinted by permission of the author and Prentice-Hall, Inc., Upper Saddle River, N.J.

The same premises have also been used to support other, equally extreme, conclusions. Oversimplifying slightly, we may say that Democritus reasoned in this way: “The wine that tastes sweet to me tastes sour to you; therefore, I do not perceive that it is sweet and you do not perceive that it is sour, and the wine itself is neither sweet nor sour.” Protagoras, however, reasoned in a somewhat different way: “The wine that tastes sweet to me tastes sour to you; hence, I perceive that it is sweet and you perceive that it is sour; and therefore, one cannot say absolutely either that the wine is sweet or that the wine is sour; one can only say relativistically that whereas it is true for me that the wine is sweet, it is true for you that the wine is sour.”² And some of the American New Realists, in defense of the view that “things *are* just what they *seem*,” drew still another conclusion: “The wine that tastes sweet to me tastes sour to you; therefore, one must say (absolutely and not relativistically) that there are contradictions in nature; one must say of the wine not only that it is both sweet and not sweet, but also that it is both sour and not sour.”³

Variants of these arguments may be found not only in writings on popular science (“Physics and psychology teach us that the world is not at all like what we perceive”), but also in the works of distinguished psychologists and philosophers. Some philosophers, in order to avoid such extreme conclusions, have been led to question the premises. It has been suggested, for example, that the appearances of things may only *appear* to change with the condition of our body.⁴ It has also been suggested that things may not actually appear in different ways – that it is a mistake to suppose that by altering either our perceptual apparatus or the conditions of observation, we can produce anything that might properly be called a change in the way in which a physical thing appears.⁵ But such extreme measures are not at all necessary. We can accept the premises that Democritus used and, at the same time, reject his conclusions, for the conclusions do not follow from the premises. This would also hold true for the other versions of the argument.

Aristotle’s Solution

Referring to Democritus, Aristotle wrote: “The earlier students of nature were mistaken in their view that without sight there was no white or black, without taste no savour. This statement of theirs is partly true, partly false. ‘Sense’ and ‘the sensible object’ are ambiguous terms; i.e., they may denote either potentialities or actualities. The statement is true of the latter, false of the former. This ambiguity they wholly failed to notice.”⁶

In suggesting that the terms “white” and “black” are ambiguous, Aristotle is taking note of the fact that in certain uses, these terms are intended to refer to ways of appearing and that in other uses they are intended to refer to certain properties or dispositions of physical things – those properties or dispositions in virtue of which the things appear in the ways in which they do appear. If a physical thing *is* white, if it has the properties or dispositions to which Aristotle referred, then it is such that, when it is viewed by an ordinary observer under

favorable lighting conditions, it will appear white to that observer. The physicist can tell us in detail just what the conditions are that a thing must satisfy if it is to have this property; that is to say, he can tell us just what characteristics a physical surface must have if it is to appear white to a normal observer in ordinary light. Let us say of such terms as "white," "black," "yellow," "red," "bitter," and "sweet," that when they are used to refer to these properties or dispositions, they have a *dispositional* use, and that when they are used to refer to ways of appearing, to ways in which things may appear, they have a *sensible* use. Aristotle is telling us, then, that the statement "Without sight, there is no white or black, without taste, no savour" is true if the terms "white," "black," and "savour" have a sensible use, and false if they have a dispositional use. Democritus, therefore, seems to have committed the fallacy of equivocation: Having established that the statement is true when it is taken in the first of these two ways, he goes on to infer fallaciously that it is also true when it is taken in the second.

And it is clear that in the passages referred to, Democritus does not establish his thesis about perception – his thesis that no one ever *perceives* any object to be white, black, yellow, red, bitter, or sweet. For the only argument that he presents in favor of this thesis is the fallacious argument in favor of his thesis concerning the nature of physical things.

Similar objections apply to the other versions of the argument. In each of the three versions considered, the terms "sweet" and "sour" have their sensible use in the premise ("The wine that tastes sweet to me tastes sour to you") and their dispositional use in the conclusion.

Sense-Datum Fallacies

The deceptive character of all three versions of the argument might be said to lie in the fact that certain truths about appearances are mistaken for truths about the things that present those appearances. From the fact that a thing's *appearing* white depends upon the condition of the perceiver, one infers mistakenly that the thing's *being* white is also something that depends upon the condition of the perceiver.

It is also possible to err in the other direction. One may make the mistake of supposing, with respect to certain truths about the things that appear to us, that they are also truths that hold of the appearances that those things present.

One such mistake, very frequently made, is that of supposing that if we perceive a physical thing, then we also *perceive* its appearances – that we see its visual appearances, hear its auditory appearances, feel its tactful appearances. But this is to misconceive the nature of perception. We perceive a thing when the thing as stimulus object has acted upon our sense organs, thereby causing us to be appeared to. The appearances of things, however, are not stimulus objects that affect our sense organs and therefore they are not themselves anything that we perceive. We do not see, hear, or feel the appearances of things.

Another such mistake may be more pernicious. From the fact that a physical thing *appears* white, for example, one might infer mistakenly that the thing

presents an appearance which *is* white, and hence, that there are certain physical things and certain appearances which are alike in color. If this inference were sound, one could also say that, under favorable conditions of observation, the appearances of things have the same color as do the things themselves, in which case appearances could be said to *resemble* their objects in important respects. Thus, Lucretius suggested that when a man perceives a tree, a *simulacrum* – a small physical object having the characteristics that the tree is seen to have – is produced inside the head.⁷ Subsequent philosophers have said that the appearance may “picture” or even “duplicate” the thing that appears.⁸ And why *not* say that if a physical thing appears white, then it presents an appearance which *is* white?

For one thing, it is clear that the inference from “Something appears *F*” to “Something presents an appearance that *is F*” is not in general valid. For there are adjectives which are such that, if we replace “*F*” by any of those adjectives, then “Something appears *F*” will be true and “Something presents an appearance which is *F*” will be false. From “The man appears tubercular,” we may not infer “The man presents an appearance which is tubercular,” and from “The books appear worn and dusty and more than two hundred years old,” we may not infer “The books present appearances which are worn and dusty and more than two hundred years old.”

Moreover, there is an absurdity inherent in saying that an appearance and a physical thing may have the same color. If we say of a physical thing that it is white, we are saying that the thing is such that, when it is viewed by a normal observer under favorable conditions, then it will appear white. Suppose, then, that “It will appear white” does imply “It will present an appearance which is white,” where “is white” has the sense that it has when it is applied to a physical thing. In such a case, a white, physical thing would be something such that, when it is viewed by a normal observer under favorable conditions, it will present an appearance which is such that, when *it* – the appearance – is viewed under favorable conditions, then it will present a (second-order) appearance which is white; the (second-order) appearance will therefore be such that, when it is viewed under favorable conditions, then it will present a (third-order) appearance which is such that . . . and so on, *ad infinitum*.

If we thus assimilate appearances to substances or concrete things, we multiply entities – and problems – beyond necessity. We find ourselves confronted, for example, with such strange questions as: If the appearance can be white in the sense in which a rose can be white, does it also have a certain weight, an inside, and a backside? Could it be that the backside of the white appearance, the side that (somehow) faces away, is green, or blue, or yellow?

But what is the appearance if it is not a substance or concrete thing?

The Adverbial Theory

When we say “The appearance of the thing is white,” our language suggests that we are attributing a certain property to a substance. But we could just as

well have said “The thing appears white,” using the verb “appears” instead of the substantive “appearance.” And in “The thing appears white,” as already noted, the word “white” functions as an adverb.⁹ Ordinarily, the point of an adverb is not to attribute a property to a substance, but to attribute a property to another property (“He is exceptionally tall”) or to attribute a property to an event, process, or state of affairs (“He is walking slowly”). We might say, then, that the word “white,” in what we have called its sensible use, tells us something about that state of affairs which is an object’s appearing; it tells us something about the *way* in which the object appears, just as “slowly” may tell us something about the way in which an object moves.

We have noted, however, that a man may be presented with a “white appearance” when no object is appearing (say, when he is thinking about a possible white object). Hence, if we are to speak more strictly, we should not say that “white,” in its sensible use, always refers to the way in which an object appears; it refers, rather, to the way in which one is *appeared to* – whether or not an object appears. Or if we introduce an active verb such as “sensing” or “experiencing” as a synonym for the passive “is appeared to,” we could say that “white,” in its sensible use, refers to the way in which a man may sense or experience.

No longer needing such expressions as “white appearance,” we need not countenance the question as to whether the white appearance has a certain weight, or a backside, or an inside. And thus, we need not wonder whether the backside of a white appearance might be green, or blue, or yellow. We need not ask whether appearances might exist unsensed – whether, in Bertrand Russell’s terms, there are “unsensed sensibilia.”¹⁰ And we need not ask whether appearances might be identical with parts of the external physical things that we perceive – whether the white appearance that we sense might be identical with the surface of the white object that we see. For in saying “He is appeared to white,” or “He senses whitely,” we are not committed to saying that there *is* a thing – an appearance – of which the word “white,” in its sensible use, designates a property. We are saying, rather, that there is a certain state or process – that of being appeared to, or sensing, or experiencing – and we are using the adjective “white,” or the adverb “whitely,” to describe more specifically the way in which that process occurs.

The Phenomenological Problem

One may feel, however, that this “adverbial” theory leaves something out. Even if the appearance is not a *simulacrum* of the object that appears, the relation between the appearance and the object may seem to be more intimate than the “adverbial” theory, as we have it so far, would allow. The problem of saying just what this relationship is may be called the phenomenological problem of appearances. The facts are familiar to everyone, but it is difficult to describe them without either overestimating or underestimating the role of appearances and without drawing unwarranted philosophical conclusions. The principal facts, I believe, are four.

(1) We perceive the object to have the characteristics we do perceive it to have, partly *because* of the way in which it appears to us. If the objects that we now perceive happened to appear in certain ways *other* than those in which they are now appearing, then we would not be perceiving them to be the objects that we are now perceiving them to be. It does not follow from these facts, however, that to perceive something to be, say, a tree, is to “make a causal inference” or “to frame the hypothesis” that a tree is one of the causes of the way in which one is being appeared to. Perceiving no more consists in deducing the causes of appearing, than reading consists in deducing the causes of ink marks.

(2) As we emphasized earlier, the appearance of a physical object – the way of being appeared to which the object as stimulus serves to cause – plays a fundamental role in the context of *justification*. If I ask myself Socratically what my justification is for thinking that it is a *tree* that I see, and [if I then ask myself what justification I have for accepting the statements I gave in answer to the previous question, and so on], I will reach a point at which I will justify my claim about the tree by appeal to a proposition about the way in which I am appeared to.

(3) A point of a rather different sort follows from one of the familiar features of perception. Whenever we *see* a physical object, then we also see certain parts of that object and fail to see certain other parts of that object. (But from the fact that we fail to perceive certain parts of the object, it does not follow that we fail to perceive the object. Verbs of perception are like “to be located in” and unlike “to contain.” If one object contains another, then it contains every part of the other; hence, New Hampshire contains every part of Jaffrey. But one object may be located in another without being located in every part of the other; what is in New Hampshire need not also be in Jaffrey.¹¹) As the use of a microscope may suggest, every part that we see has parts of its own that we do not see. Similar remarks apply to perception by means of any of the other senses: Whenever we perceive an object by means of any one of the senses, there are certain parts of that object that we perceive and certain other parts of it that we do not perceive. With reference to these facts, we may now make our third point concerning the relation between perceiving and being appeared to: Whenever we perceive an object, then the object appears to us in a certain way; each of the parts that we perceive also appears to us in a certain way; and those parts that we do not perceive do not appear to us in any way.

(4) Using, for the moment, the terminology of “appearances,” we may also say that the appearances of the parts of the object are included in the appearances of the whole. If, for example, a man is looking at a hen, then we may say of the hen itself, and of those parts of the hen that the man happens to see, that each of these objects presents an appearance. We can say of the hen that it is a whole in which these various parts (among others) are contained; we can also say of the appearance of the hen, that it, too, is a whole in which the appear-

ances of the various parts are contained. Indeed we might say of the appearance of each part, that it is a part of the appearance of the whole. The appearance of the outer part of the tip of one of the feathers is a part of the appearance of the feather; the appearance of the feather is a part of the appearance of the wing; the appearance of the wing is a part of the appearance of the side of the hen; and the appearance of the side of the hen is a part of the appearance of the hen. And these facts, it must be conceded, are difficult to formulate, either in the terminology of "appearing" or in the terminology of "sensing" or "being appeared to."

If we use the terminology of "appearing," we might express the facts in question as follows: "The way in which a thing appears to a man includes ways in which some, but not all, of its parts appear, and the way in which any part of a thing appears is included in the way in which the whole appears." If we use the terminology of "being appeared to," we might say: "The way in which a man is appeared to by a thing includes ways in which he is appeared to by some, but not all, of the parts of the thing, and the way in which he is appeared to by any part of the thing is included in the way in which he is appeared to by the thing." And if we use the terminology of "sensing," then we shall have to replace the "by" by some other preposition or phrase – possibly, "with respect to" – and say: "The way in which a man senses with respect to a thing includes ways in which he senses with respect to some, but not all, of the parts of the thing, and the way in which he senses with respect to any part of the thing is included in the way in which he senses with respect to the thing." It is clear that the terminology of "appearances," whatever its theoretical limitations, has a practical advantage at this point. But if I am not mistaken, the facts of the matter can be put in the terminology of "being appeared to."

Apearances and Brain Processes

According to what is sometimes called the "identity theory," appearances may be identified with something that is to be found in the brain, and therefore, they may be subsumed under the category of what is material or physical. The theory is defended on the ground (1) that there is known to be at least a close correlation between appearances and what is cerebral or neurological, and (2) that in order not to multiply entities beyond necessity it is reasonable to suppose that a strict identity is involved rather than a mere correlation between entities that are distinct. (Prehistoric astronomers, noting the close correlations obtaining between the wanderings of the evening star and those of the morning star, may have reasoned similarly in behalf of the thesis that the evening star and the morning star are one and the same.) It is commonly believed that if the identity theory could be shown to be true, then, so far as what we know about appearances is concerned, there would be no need to assume the existence of any entities other than physical bodies and their properties, states, and processes; what we know about appearances could be accommodated to the

assumption that there is “nothing in the world but increasingly complex arrangements of physical constituents.”¹²

To evaluate the identity theory and the claims that have been made in its behalf, we must first decide just what it is that is being identified with what.

If we were to reject the “adverbial theory” of appearing, or being appeared to, and were to accept a substantival theory of *appearances* in its place, then our formulation of the identity theory could be reasonably straightforward. Taken substantively, the sentence “Jones experiences a red appearance” could be said to be like “Jones eats a red tomato” in that it describes an intimate relation between Jones and a certain other substance. We could thus formulate the identity theory by saying that appearances are *parts* of the brain – chunks of grey matter, say, or cells, or strips of nervous tissue. And this is what Thomas Case, a nineteenth-century advocate of “physical realism,” seems to have said.

According to Case, appearances are to be identified with “physical parts of the nervous system, tactile, optic, auditory, etc., sensibly affected in various manners.” Assuming that people *perceive* appearances, he was then able to say that they perceive the insides of their own bodies, not external, physical things. “The hot felt is the tactile nerves heated, the white seen is the optic nerves so coloured.” He then argued that on the basis of what people perceive about their nervous systems, they make inferences and hypotheses about what goes on outside: “From the hot within we infer a fire without.”¹³

Case thus seems to have committed the “sense-datum fallacy”; for he assumes that when the “fire without” appears hot, then there is an appearance “within” which actually has the property that the fire appears to have. He does not distinguish the sensible and dispositional uses of property words. He assumes that people perceive appearances and not external physical things. And he assumes that the process we ordinarily call perceiving, is really just a matter of framing hypotheses and making inferences, and thus he is able to conclude that we come to know external things by first examining the insides of our heads. His “physical realism,” therefore, was easily parodied. (F. H. Bradley remarked that according to Case’s theory, when he was offended by an unpleasant smell, what he was really aware of was “the stinking state of my own nervous system.”¹⁴)

But the identity theory need not involve the various errors that have been attributed to Case. Contemporary versions of it are considerably more difficult to criticize.

J. J. C. Smart has suggested that appearances “are nothing over and above brain processes”.¹⁵ His view thus presupposes an adverbial theory of appearing rather than a substantival theory of appearances; he is concerned with the *process* of appearing and not with certain *substances* called “appearances.” Given his view, such sentences as “Jones experiences a red appearance” are misleading, for “appearance” should be replaced by “appearing.” But if we are to avoid multiplying entities beyond necessity, we will not say “Jones experiences a red *appearing*.” For “Jones experiences a red appearing” suggests that there are *two* processes – one, the *experiencing*, the other, the *appearing*. We may suppose that the experiencing and the appearing – or rather, the experiencing and the

being appeared to – are one and the same. Hence, we could make use of a locution similar to our earlier “Jones is appeared red to.” But since we do not want to say that the word “red,” in application to a *process*, has the same meaning that it has in application to a concrete thing or substance, our locution will be even less misleading if we express it, once again, as “Jones is appeared to *redly*.” This awkward locution, as we have emphasized, has the theoretical advantage of suggesting that appearing is a process, that the adverb “redly” designates a property of a process (just as “swiftly” and “slowly” designate properties of processes), and that the process of being appeared to does not involve a *second* process which is the *experiencing* of the process of being appeared to.

What now does this second version of the “identity theory” tell us? What is involved in saying that that process which is Jones’s being appeared to redly is really something that is to be found in his brain?

Let us consider how the theory might be applied to a single case – to just one occasion of Jones’s being appeared to redly. The theory would tell us that on this occasion (1) there is going on in Jones’s head a certain process – some kind of vibration, say – which a neurologist might be able to identify independently, and (2) that this neurological process is the very same process as the one that we are now describing as Jones’s being appeared to redly. One professed to give a “physical” account of the appearances that a man may experience but *not* of that process or event which is the man’s experiencing of those appearances. The present view dispenses with the appearances and professes to give a physical account of experiencing – a physical account of that process or event which is Jones’s being appeared to. . . .

Notes

- 1 Fragment quoted from Milton Nahm, *Selections from Early Greek Philosophy* (New York: Appleton-Century-Crofts, 1934), p. 209; cf. pp. 173–87, 194–5.
- 2 See the discussion of Protagoras’ view in Plato’s *Theaetetus*.
- 3 Cf. E. B. Holt, Ralph Barton Perry, and others, *The New Realism* (New York: Macmillan, 1912), pp. 2, 365. . . .
- 4 Cf. G. E. Moore, *Philosophical Studies* (London: Routledge & Kegan Paul, 1922), p. 245.
- 5 This suggestion seems to be presupposed by passages in J. L. Austin’s *Sense and Sensibilia* (New York: Oxford University Press, 1962). . . .
- 6 *De Anima*, book III, ch. 2, p. 426a; see also *Metaphysics*, book IV, ch. 5, 1010b.
- 7 *On the Nature of Things*, book IV.
- 8 “No man doubts that when he brings to mind the look of a dog he owned when a boy, there is something of a canine sort immediately present to and therefore compresent with his consciousness, but that it is quite certainly not that dog in the flesh.” A. O. Lovejoy, *The Revolt against Dualism* (New York: W. W. Norton, 1930), p. 305. . . .
- 9 The point is developed in detail by C. J. Ducasse, *Nature, Mind and Death* (La Salle, Ill.: Open Court Publishing, 1949), ch. 13. This general view of appearing is suggested by Thomas Reid, in his *Essays on the Intellectual Powers of Man*, Essay I, ch. 1, section 12, and by G. F. Stout, in “Are Presentations Mental or Physical?” *Proceedings of the Aristotelian Society*, n.s., vol. ix (1909).

- 10 See the essay “The Relation of Sense-Data to Physics,” in Russell’s *Mysticism and Logic* (New York: W. W. Norton, 1929); this book was first published in 1918.
- 11 C. D. Broad once argued that, inasmuch as we do not see every part of the bell on any of those occasions on which, as we like to think, we see a bell, therefore, strictly speaking, we never see a bell at all; see his *Mind and its Place in Nature* (New York: Harcourt, Brace & World, 1925), pp. 149–50. This is like saying that, since the butcher doesn’t cut every part of the roast, therefore, strictly speaking, he doesn’t cut the roast at all.
- 12 J. J. C. Smart, “Sensations and Brain Processes,” in *The Philosophy of Mind*, ed. V. C. Chappell (Englewood Cliffs, N.J.: Prentice-Hall, 1962), p. 161
- 13 Thomas Case, *Physical Realism* (London: Longmans, Green 1888), pp. 24, 25, 33.
- 14 Quoted by H. H. Price, *Perception* (New York: Robert M. McBride 1933), p. 127.
- 15 *The Philosophy of Mind*, p. 163.

What is the Relation between Mind and Body?

34 Which Physical/Thing Am I? An Excerpt from “Is There a Mind–Body Problem?”*

Roderick M. Chisholm

... The “double aspect theory” tells us this: There are certain things which have physical properties and therefore are physical objects; some of these things also have certain mental or intentional properties; and persons – you and I – are such things as these.

C. A. Strong put this last point clearly. He wrote:

I am to outer appearance physical but to inner perception psychical; there is therefore no contradiction in a thing being at once physical, that is, extended, composed of parts, productive of effects, and psychical, that is of the nature of feeling.¹

Strong is not here saying that “my mind” is an aspect of a physical thing, much less that *I am* an aspect of a physical thing. What he says is that there *is* a certain physical thing which has inner and outer aspects and that that physical thing is identical with me.

If we were to accept this theory, then we could ask: “*Which* physical thing am I?” I am afraid we could not provide a precise answer to this question.

If I am in fact a physical thing, then, it should be obvious, that physical thing is either this gross physical body now standing before you or it is some proper part of this gross physical body. There are, of course, many philosophical arguments professing to show that the person cannot be identical with his gross macroscopic physical body. Some of these arguments, I think, are sound – in particular those appealing to certain facts about persistence through time.

The body that persists through time – the one I have been carrying with me, so to speak – is an *ens successivum*. That is to say, it is an entity which is made up of different things at different times. The set of things that make it up today is not identical with the set of things that made it up yesterday or with the set of

* From Roderick M. Chisholm, “Is There a Mind–Body Problem?” *The Philosophic Exchange*, 2, no. 4 (1978), pp. 25–34. Reprinted by permission of the author and the Center for Philosophic Exchange. Also includes material from Roderick M. Chisholm, “Self Profile,” in Radu J. Bogdan, ed., *Roderick M. Chisholm* (Dordrecht: D. Reidel Publishing, 1986); © 1986 by D. Reidel Publishing Company. Reprinted with kind permission of the author and Kluwer Academic Publishers.

things that made it up the day before. Now one could say that an *ens successivum* has different “stand-ins” at different times and that these stand-ins do duty for the successive entity at the different times. Thus the thing that does duty for my body today is other than the thing that did duty for it yesterday and other than the thing that will do duty for it tomorrow. But what of me?

Am I an entity such that different things do duty for *me* at different days? Is it *one* thing that does my feeling depressed for me today and *another* thing that did it yesterday and still another thing that will do it tomorrow? If I happen to be feeling sad, then, surely, there is no *other* thing that is doing my feeling sad for me. We must reject the view that persons are thus *entia successiva*.

Our reasoning can be summarized. Suppose (i) that I am now sad. Now (ii) if there is an *ens successivum* that bears my name and is now sad, then it is sad in virtue of the fact that one of its stand-ins is now sad. But (iii) I am not sad in virtue of the fact that some *other* thing is doing my feeling sad for me. Therefore (iv) I am not an *ens successivum*.

What would be an *ens nonsuccessivum*? If an individual thing were a non-successive entity, what would it be like? If an *ens successivum* is an individual thing that is made up of different things at different times, then an *ens nonsuccessivum* would be an individual thing that is *not* made up of different things at different times. This means that, at any moment of its existence, it has precisely the same parts it has at any other moment of its existence; at no time during which it exists, does it have a part it does not have at any other time during which it exists.

It is tempting to reason, in Leibnizian fashion: “There are *entia successiva*. Therefore there are *entia nonsuccessiva*.” I believe this reasoning is sound. I would add, moreover, that every extended period of time, however short, is such that some *ens nonsuccessivum* exists during some part of that time. For I believe it is only by presupposing this thesis that we can make sense of the identity or persistence of *any* individual thing through time.

Might I not be, then, such an *ens nonsuccessivum*? Leibniz mentions – and rejects – a theory which is similar to this. “The soul,” he says, “does not dwell in certain atoms appropriated to itself, nor in a little incorruptible bone such as the *Luz* of the Rabbis.”² Of course, the hypothesis I have suggested, if filled in by reference to such a material thing as the *Luz* bone, would not imply that “the soul” dwells there – if the soul is understood to be something *other* than the person, still another thing that the person “has.” We would be saying rather that the person dwells there. And to say that he “dwells” there would be to say that the person *is* the *Luz* bone or some proper part of it.

If we accept this theory, then, of course, we part company with personalism. The doctrine that persons are physical things – even intactly persisting physical things – would not have been taken seriously by Borden Parker Bowne and his followers. Yet, if we view the person in the way I have suggested, we may go on to affirm many of the *other* philosophical theses that the personalists felt to be important. Thus we could say, as Bishop Butler did, that “our gross organized bodies with which we perceive the objects of sense, and with which we act, are no part of ourselves. . . . We see with our eyes in the same way we see with our

glasses.”³ The eyes are the *organs* of sight, not the *subject* of sight. We could say, as Butler and the personalists did, that the destruction of the gross physical body does not logically imply the destruction of the person. And we could accept the view that St Thomas attributes to Plato: the person is “in a body in somewhat the same way as a sailor is in a ship.”⁴

Some Objections Considered

To understand the view that is being proposed, let us formulate certain objections that readily come to mind and then attempt to reply to them. I will consider four such objections.

(1) “The hypothesis you are considering implies, then, that there is a kind of matter that is incorruptible and that the person is a material thing of that sort? But this is hardly adequate to the facts of physics.”

The reply is that the theory does not imply that there is certain matter that is incorruptible. It implies rather that there are certain material things – in all probability, certain material particles or subparticles – that are incorrupted and remain incorrupted as long as the person survives.

The theory would be, then, that I am literally identical with some proper part of this macroscopic body, some intact, nonsuccessive part that has been in this larger body all along. This part is hardly likely to be the *Luz* bone, of course; more likely, it would be something of a microscopic nature, and presumably something that is located within the brain.

(2) “Persons, being thinking things, must have a complex structure. But no microscopic entity that is known to physics has the equipment that is necessary for thinking. After all, you can’t think unless you have a brain. And *those* little things don’t have brains!”

The hypothesis being criticized is the hypothesis that *I* am such a microscopic entity. But note that I do have a brain. And therefore, according to the hypothesis in question, the microscopic entity has one, too – the same one that I have, the one that is inside my head. It is only a confusion to suppose that the microscopic entity – which may in fact be inside my brain – has *another* brain which is in fact inside of it.⁵

The brain is the *organ* of consciousness, not the *subject* of consciousness – unless I am myself my brain.⁶ The nose, similarly, is the organ of smell and not the subject of smell – unless I am myself my nose. But if I am one or the other – the brain or the nose – then, I the subject, will have some organs that are spatially outside me.

The hypothesis in question, then, is that I am a certain proper part of my brain. This would imply that the subject of consciousness is a proper part of the organ of consciousness.

(3) “You say I’m identical with some microscopic particle or some

subparticle. But I am 6 feet tall and weigh 175 pounds. Therefore your theory would imply that there is a certain microscopic particle which is 6 feet tall and weighs 175 pounds. But this is absurd and therefore your theory is absurd."

The argument, of course, errs in taking too literally the premise expressed by saying "I am 6 feet tall and weigh 175 pounds." For what the premise actually tells us is that I have a body which is 6 feet tall and weighs 175 pounds.

(4) "Do you mean to suggest seriously, then, that instead of weighing 175 pounds, you may weigh less than a milligram?" The answer has to be yes. We must be ready, therefore, to be ridiculed, for, in this case, even those who know better may be unable to resist the temptation. But those who do know better will realize that a person can truly say, in *one* sense, that he weighs 175 pounds, and in *another* sense, that he weighs less than a milligram. The formulation of the first statement would be more nearly accurate (I say "more nearly *accurate*," not "more nearly correct") if it read: "I have a body that weighs 175 pounds."

Speaking in a loose and popular sense, I may attribute to myself certain properties of my gross macroscopic body. (And speaking to a filling station attendant I may attribute certain properties of my automobile to myself: "I'm down there on the corner of Jay Street without any gasoline.") The response needn't be: "How, then, can you be standing here?" One might say that the property of being down there is one I have "borrowed" from my automobile.) But if I am a microscopic part of my gross body, then, strictly and philosophically, one cannot attribute to *me* the properties of *it*. The properties of weighing 175 pounds and being 6 feet tall are properties I "borrow" from my body. Strictly and philosophically, *it* has them and I do not.⁷

(5) "You say that you might be a small physical part that uses the main-frame brain as its organ of thought – it thinks by means of the brain. Theoretically, then, there is the possibility that you might exchange brains with another person – either by transferring brains from one body to another or by transferring persons from one body to another. But what makes you the person you are is your *consciousness*: your present beliefs, desires, memories, and perceptions. Recall what Locke said:

It being the same consciousness that makes a man be himself to himself, personal identity depends on that only, whether it be annexed to one individual substance, or can be continued in a succession of several substances. (*Essay Concerning Human Understanding*, book II, ch. xxiii, section 10)

But our consciousness is dependent upon our brains. We have the beliefs, memories and perceptions we do have because of the present make-up of our brains. Therefore, if you and I exchange brains, we will also exchange consciousnesses. And this means that you will become me and I will become you. But isn't that absurd?"

That is absurd. But the absurd consequence follows from the assumption

that personal identity is a function of the nature of one's consciousness. The objection confuses the *criteria* of personal identity with its *truth-conditions*. The *criteria* of personal identity are simply the *means* by which one *identifies* any given person. We may say, if we choose, that they are the means by which one "determines the identity" of the person. But they do not determine the identity of the person in the sense of *making* that person the person that he is. Compare the criteria by means of which we decide whether a certain event occurred in the past. If we decide that it rained yesterday, we do so by means of certain *traces* which we find today (puddles, testimonies, pictures, recordings). These traces are not truth-conditions of yesterday's rain – it is logically possible that they occur even though it did not rain yesterday. They enable us to determine whether or not it rained – but they do not themselves determine the rain.

Conclusion

What are the possibilities, after all? There *are* persons. Therefore either the person is a physical thing or, as Lovejoy suggests, the person is a nonphysical thing. But does anything we know about persons justify us in assuming that persons are *nonphysical* individual things?

What if we suppose that the concept of an extended thing presupposes the concept of ultimate nonextended things which, somehow, make up the extended thing? Could we then identify the person with such an unextended thing? I believe that this hypothesis would contradict the assumption that persons are *entia per se*. For I would say that the unextended things (boundaries, lines, points, surfaces) that are said to be presupposed by extended things are ontological parasites and not instances of *entia per se*: they depend for their own properties upon the extended things which are said to presuppose them.⁸

What point would there be in the hypothesis that certain individual things have the property of being nonphysical? How could that help us in explaining anything?⁹

If I *am* a physical thing, then the most plausible hypothesis would seem to be that I am a proper part of this gross macroscopic body, even if there is no way of telling from the "outside" which proper part I happen to be.

I would suggest that, if this philosophic hypothesis seems implausible to you, you try to formulate one that is less implausible.

Notes

- 1 C. A. Strong, "Final Observations," *Journal of Philosophy*, xxxviii (1941), pp. 233–43; the quotation is on page 237.
- 2 *New Essays Concerning Human Understanding*, book II, ch. xxviii (La Salle, Ill., Open Court Publishing, 1916), p. 242. Alfred Langley, editor of this edition of Leibniz, quotes an ancient discussion of the *Luz* bone: "The old Rabbis of blessed memory have not only seen this bone, but have found it actually so strong and hard that their hammer and rock flew in pieces before this bone was injured in the least" (p. 242n).

- 3 Joseph Butler, *The Analogy of Religion*, part I, chapter 1 ("Of a Future Life"); see *The Whole Works of Joseph Butler, LL.D.* (London: Thomas Tegg, 1839), p. 7.
 - 4 St Thomas Aquinas, *On Spiritual Creatures*, Article II (Milwaukee: Marquette University Press, 1949), p. 35.
 - 5 I have illustrated this confusion in Richard Taylor's *Action and Purpose* (Englewood Cliffs, N.J., Prentice-Hall, 1966), p. 137.
 - 6 Compare Franz Brentano, *Religion und Philosophie* (Bern: A. Francke Verlag).
 - 7 Strawson emphasizes that persons have both psychological and physical properties. But, if what I say is true, most of the physical properties that we ordinarily attribute to the person are "borrowed" in this sense from the person's body.
 - 8 Compare Brentano, *Religion und Philosophie*, p. 224.
 - 9 For a clear formulation of this point, see Richard Taylor, *Metaphysics*, 2nd edition (Englewood Cliffs, N.J., Prentice-Hall, 1974), pp. 34–35.
-

35 Personal Identity: a Materialist Account*

Sydney Shoemaker

1 Introduction

From earliest times people have found intelligible, and sometimes believable, the idea that persons are capable of surviving death, either in disembodied form or through bodily resurrection or reincarnation. And many a piece of popular fiction relies on the idea that a person might have different bodies at different times. We are also familiar, both from fiction and from the annals of psychiatric medicine, with the idea of two or more distinct 'personalities' successively manifesting themselves in one and the same body. Yet another such idea is that two distinct minds or consciousnesses might simultaneously inhabit the same body – and recent studies of 'split-brain patients' have suggested to some investigators not only that this is conceivable but that it actually happens.¹ One way of raising the problem of personal identity is by asking whether, or to what extent, such ideas are coherent, and what it is about the nature of personal identity, or our concept of it, which permits, or forbids, such envisioned departures from the normal course of events.

The problem of personal identity can be viewed as an aspect of the 'mind – body problem'. For a variety of reasons we are inclined to resist the view, so strongly suggested by the current scientific world view, that mental states and processes are nothing over and above certain highly complex physical and chemical processes. One reason is the 'special access' we have to our own mental

* From Sydney Shoemaker and Richard Swinburne, *Personal Identity* (Oxford: Blackwell, 1984). Reprinted by permission of the author.

states. One comes to have knowledge of these states without observing, or gathering evidence about, the physical states of one's own body; and possession of the knowledge seems compatible with total ignorance of one's own inner physiological states, and, more generally, the condition of one's body. And if one reflects on what one knows in having this self-knowledge – the existence of intentional states like believing that Argentina's inflation rate is higher than Brazil's, and qualitative states like seeing blue and having an itch – it is difficult at best to see how this could be reducible to any facts about one's behaviour or neurophysiology. Puzzlement about the nature of mental states is bound to give rise to puzzlement about the nature of persons, the pre-eminent subjects of such states. And this in turn manifests itself in puzzlement about personal identity – for a central part of understanding the nature of a kind of things (like persons) is understanding the identity conditions for things of that sort. The considerations that make it seem that mental states cannot be physical states also make it seem that persons cannot simply be physical bodies, and that personal identity must consist in something other than bodily identity.

Among the things to which persons have a 'special access' are facts about their own identity over time; they have this in their memory knowledge of their own past histories. One's memory knowledge of one's own past differs strikingly from one's knowledge, including memory knowledge, of the past histories of other persons. If I claim to remember *you* doing something yesterday, it is at least a theoretical possibility that my claim is in error, not because my memory is mistaken, but because the person I remember doing that thing is not you but someone who looks just like you and whom I have misidentified as you. But if I claim to remember that I did such-and-such yesterday, it is absurd to suppose that I could be mistaken in *that* way. And whatever may be said of my judgements about the identity of others, it is certainly not the case that I ground such judgements about myself on evidence of bodily identity. Here again the nature of self-knowledge raises questions about personal identity, in part by calling into question the natural view that the identity of a person is simply the identity of a living human body.

A rather different source of perplexity about personal identity has to do with the special concern persons have for their own continued existence and their own future welfare. Imagine that a wizard demonstrates to you his ability to reduce any object to a pile of dust by a wave of his wand and then, with another wave, to create an exact duplicate of that thing out of another pile of dust. If one really believes that he can do this, one probably would not be too averse to letting him do it to one's kitchen stove. But only a monster would offer his wife or child as a subject for the wizard's trick, and only a madman (or a suicide) would offer himself. Or so it initially strikes us. Our concern for personal identity, the kind of importance it has for us, seems totally different in kind from the concern we have for the identity of other sorts of things. And this is linked to the special concern each person has for his or her own future welfare. It is this that gives point to many of our moral, social and legal practices, and explains the significance they attach to considerations of personal identity. If a person does an action, it is that same person who can later be held responsible for the

action, and whom it is appropriate to punish or reward for doing it. If someone buys something, it is that person who is subsequently entitled to the use of the item purchased. These principles, which are constitutive of the institutions of punishment and property and the concept of moral responsibility, are intelligible only against the background of a conception of human motivation in which a central role is played by the special concern each person has for his own future well-being.

An account of personal identity ought to make intelligible the knowledge we have of personal identity, including the special access each of us has, in memory to his own identity, and it ought to make intelligible the special sort of importance personal identity has for us. It ought also to cohere with the rest of what we know about the world. In my own view, this last requirement means that an account of personal identity ought to be compatible with a naturalistic, or materialistic, account of mind. To a large extent, the mind–body problem, including the problem of personal identity, arises because of considerations that create the appearance that no naturalistic account could be true; and I think that solving the problem has got to consist in large part in dispelling that appearance (while acknowledging and explaining the facts that give rise to it). Finally, our account of personal identity must be compatible with the logical principles that govern the notion of identity itself. It is to these that we now turn.

2 The Concept of Identity

Logicians characterize identity as an ‘equivalence relation’, meaning by this a relation that is transitive (if a has it to b , and b has it to c , then a has it to c), symmetrical (if a has it to b , b has it to a), and reflexive (everything has it to itself). It is marked off from other equivalence relationships by its conformity to Leibniz’s Law (the principle of the indiscernibility of identicals), which says that if a is identical to b , then whatever is true of a is true of b , and conversely. Identity is even more briefly characterizable as the relation which everything has, necessarily, to itself, and which nothing has to anything other than itself (but the last clause makes this definition circular, since it means ‘and which nothing has to anything not identical to itself’).

It is important to distinguish the relation of identity we are here concerned with from another relation that bears the same name. In the baggage claim areas of some airports are signs reading ‘Careful: many suitcases are identical’.² This is a perfectly correct use of ‘identical’, but not in the sense of it relevant to the problem of personal identity. In the airport sign, ‘identical’ means ‘exactly alike’; it expresses what is sometimes called ‘qualitative identity’ (or ‘qualitative sameness’). This must be distinguished from the sense of ‘identical’ in which it means ‘one and the same’, and expresses ‘numerical identity’ (or ‘numerical sameness’). It is the latter, the sense of ‘identical’ in which ‘identical twins’ (and identical suitcases) are not identical, that concerns us here.

Confusion of these two senses of ‘identical’ (and ‘same’) is one source of the idea that identity over time is incompatible with change. If something changes,

then in some respect it is no longer the same as it was; the thing at the earlier time and the thing at the later time are not ‘identical’. Indeed, we may seem to be driven into contradiction here: if I say ‘A is not the same as *it was*’, I seem to imply identity with the pronoun ‘it’ while denying it by saying ‘not the same’. But the contradiction is only apparent; the identity that is implied is numerical identity, while that which is denied is qualitative identity. Change is incompatible with qualitative identity between the successive states of the changing thing; but it not only allows, but logically requires, that the successive states be states of numerically the same thing.

Another source of the view that identity over time is incompatible with change is a misunderstanding of Leibniz’s Law. If a leaf is green in the summer and yellow in the fall, then in a certain sense something is true of it in the summer which is not true of it in the fall, and vice versa. But this is no violation of Leibniz’s Law. If A is a leaf in the summer and B is a leaf in the fall, what Leibniz’s Law holds to follow from ‘A is identical to B’ is not that if A is green in the summer then B is green in the fall; it is rather that if A is green in the summer then B is green in the summer (and if B is yellow in the fall, A is yellow in the fall). More generally, it tells us that whatever property A has at a time, B must have at that time, and conversely; and this is entirely compatible with A (= B) having different properties at different times.

More common than the view that identity over time is incompatible with any sort of change is the view that it is incompatible with one particular sort of change – change of composition. This stems in part from the confusions already mentioned, but has another source as well. If over a period of time some of the molecules in a tree are replaced by others, then in one sense we no longer have the same ‘substance’ as we had before. Now there is a well established philosophical sense of ‘substance’, going back to Aristotle, in which a tree *is* a substance. And to say in *this* sense that we no longer have the same substance is to say (in the case at hand) that we no longer have the same tree – which is the view Bishop Butler took of the case in which, over time, all of the matter in a tree is replaced.³ But surely we should distinguish these senses of ‘substance’. When we say that we no longer have the same substance in this case, what counts as a substance is a particular portion or quantity of matter. But it is not in *this* sense that a tree is a substance. The tree is, at a particular time, composed of a particular portion of matter, but that is not to say that it is identical to that portion of matter and that it could not at some later time be composed of some quite different matter. And of course we regularly do take things like trees to survive the gradual (and in the case of things like rivers and waterfalls, not so gradual) replacement of the matter of which they are composed. To argue that such replacement is impossible on the grounds that being the same tree requires being the same substance either equivocates on the word ‘substance’ or begs the question. I think that we can see from this that it is either false or vacuous to say that the identity of things like trees consists in their being or having the same substance; it is false if ‘substance’ means ‘portion of stuff’, and vacuous if ‘substance’ means ‘persisting subject of properties’ (for then the claim comes to: being the same tree consists in being the same tree).

Some writers like to speak of the identity over time of a person or a table as consisting in the occurrence of a succession of momentary ‘person-stages’ or ‘table-stages’ that are related to one another in certain ways. (Some speak of ‘time-slices’ rather than of ‘stages’). So, for example, if I say ‘The table I am sitting at now is the table I was sitting at yesterday’, I am asserting a relation to hold between a table-stage occurring now and one occurring yesterday. But it is important to be clear that the relation I assert to hold between the stages is not itself identity; today’s table-stage and yesterday’s table-stage cannot be the same, since stages are individuated by the times at which they occur. Let us pretend, for the moment, that different table-stages are stages of one and the same table if they belong to a single spatiotemporally continuous succession of table-stages – or, for short, if they are ‘table-linked’. Here we can say, following John Perry, that the relation of being table-linked is the *unity relation* for tables.⁴ The unity relation is not identity; it is rather the relation that holds between different table-stages when they are stages of one and the same table. Nevertheless, to specify the unity relation for tables is to say what the identity over time of tables consists in. And one way of formulating the problem of personal identity is to ask what the unity relation for persons is, i.e., what is the relation between person-stages occurring at different times, in virtue of which they are stages of one and the same person.

What sort of thing is a ‘person-stage’ or ‘table-stage’? Some philosophers who use this terminology think of persons and other continuants as four-dimensional objects which have temporal as well as spatial parts. For them momentary stages will be either temporally very small parts of continuants or temporally unextended cross-sections of them taken at particular moments of time. But one need not be committed to the four-dimensional view of ordinary continuants in order to use this terminology. Person-stages can be thought of as ‘temporal slices’, not of persons, but of the histories or careers of persons. One might think of a momentary stage as a set of property instantiations; if C is a continuant existing at time t , and P is the set of properties possessed by C at t , then C’s stage at t will be the set consisting of the instantiations in C at t of the properties in P. Or one can think of a momentary stage as an ordered pair consisting of a thing and a time; C’s stage at t will be just the ordered pair $\langle C, t \rangle$.

Questions about identity over time can be said to be questions about the *diachronic* unity of continuants of some kind, e.g., persons or tables. Questions can also be raised about the *synchronic* unity of such things. In some cases the latter are questions about ‘identity across space’; for example, if we were concerned (with Heraclitus) about the identity of rivers we might ask why it is that the river (or river part) at Minneapolis and that at New Orleans count as (parts of) the same river. In the case of persons, questions about synchronic unity are more likely to be asked about momentary experiences and other mental states than about spatial parts. The question will be: in virtue of what do different experiences or mental states occurring at the same time count as belonging to one and the same person? This is sometimes posed as the problem of the ‘unity of consciousness’. A useful term for the unity relation for persons (both diachronic and synchronic) is Bertrand Russell’s term ‘copersonal’.⁵

There is a tradition, going back to Bishop Butler, of holding that personal identity is indefinable and unanalysable, that no non-trivial account can be given of the identity conditions for persons, and that personal identity does not ‘consist’ in anything. Butler (and likewise Thomas Reid) seemed to think that this is a consequence of personal identity being identity in a ‘strict and philosophical sense’ – as contrasted with the ‘loose and popular sense’ he believed to be invoked in our ascription of identity over time to such things as trees and ships. There are contemporary philosophers who think that Butler was right. But we will begin with the working assumption that an account of personal identity can be given. Indeed, we will begin with the view Butler was primarily attacking, that of John Locke.

3 The Memory Theory

Locke’s central thesis was that personal identity consists, not in sameness of substance, but in ‘sameness of consciousness’.⁶ It is by no means uncontroversial what Locke meant by this. But it is clear enough that ‘consciousness’ for Locke includes memory, and that it is primarily memory he has in mind when he speaks of consciousness in his discussion of personal identity. Rightly or wrongly, Locke has been taken as the founder of the view that the identity over time of a person consists in facts about memory and the capacity to remember. He seems, in fact, to have held a fairly extreme version of that view: that a person A existing at a time t_2 is the same as a person B existing at an earlier time t_1 if and only if A remembers, or ‘can remember’, at t_2 actions or experiences of B occurring at t_1 .

Before we consider the objections that have been raised against this view we must try to understand its initial appeal. We have already noted (section 1) that the way a person knows of his own past on the basis of memory is different from that in which he knows of the past of any other person. When I claim to remember *your* doing such-and-such yesterday, a question can arise whether the person I remember doing that thing was really you, and not someone else who looked like you (and this is so even if the accuracy of my memory of the incident is conceded); when I claim to remember *my* doing such-and-such a thing yesterday, no such question can arise. There seems to be a way of remembering past experiences and actions – I call it ‘remembering from the inside’ – such that if someone remembers X (an action or experience) in that way, it follows that X was an experience or action of that person.⁷ Already, then, we seem to have an intimate connection between memory and personal identity: the person who remembers from the inside must be identical to the person who earlier had the remembered experience or did the remembered action. What I am calling ‘remembering from the inside’ is the kind of memory a person has of a past action when he remembers *doing* it, or of an experience when he remembers *having* it. But there also seems to be a more general connection between personal identity and memory which is not restricted to remembering from the inside: if someone remembers any event whatever,

he must be identical to one of those who witnessed that event, or otherwise knew of it in a direct way, at the time of its occurrence. I can remember *that* the Battle of Hastings occurred in 1066, but no one now alive can be said to remember the Battle of Hastings.

It is not only from the first-person point of view that the memory theory, and the allied view that personal identity is independent of bodily identity, can seem attractive. Locke remarks that ‘should the soul of a prince, carrying with it the consciousness of the prince’s past life, enter and inform the body of a cobbler, as soon deserted by his own soul, everyone sees he would be the same person with the prince, accountable only for the prince’s actions.’⁸ It is easy enough to develop this into a case in which we could have rather compelling evidence, based on considerations having to do with memory, that someone had ‘changed bodies’. For those who are sceptical about ‘souls’, it may help to imagine a case in which what are switched are not souls but brains. Suppose, then, that by a surgical blunder (of rather staggering proportions!) Brown’s brain gets into Robinson’s head.⁹ When the resulting person, call him ‘Brownson’, regains consciousness, he claims to be Brown, and exhibits detailed knowledge of Brown’s past life, always reporting Brown’s deeds and experiences in the first person. It is hard to resist the conclusion that we, viewing the case from the outside, ought to accept Brownson’s claim to be Brown, precisely on the basis of the evidence that he remembers Brown’s life from the inside. This gives *prima facie* support to the Lockean view that personal identity consists in part in facts having to do with memory.

A variety of objections have been raised against the memory theory, and some of these are clearly telling against the version apparently held by Locke. What I want to do next is to consider how Locke’s view might be modified to meet these objections, with a view to seeing whether a modified version of it can provide an acceptable theory of personal identity.

4 Objections and Revisions

The most famous objections to Locke’s account are those raised in the eighteenth century by Bishop Butler and Thomas Reid. Butler charged that the account is circular: ‘one should really think it self-evident, that consciousness of personal identity presupposes, and therefore cannot constitute, personal identity, any more than knowledge, in any other case, can constitute truth, which it presupposes’.¹⁰ Reid charged that the account is self-contradictory, and sought to show this with his ‘brave officer’ example.¹¹ At a certain time a boy is flogged for robbing an orchard. Years later the same person, now a young officer, performs a valiant deed in battle, remembering still his boyhood flogging. Many years later our man is an elderly general, who remembers the valiant deed in battle, but no longer remembers the flogging. Reid charges that on Locke’s theory the old general both is and is not the same person as the small boy; he is the same because he is identical to the young officer who is identical to the small boy (and because identity is transitive); he is not the same because he has

no memory of the flogging (and, let us suppose, has no memory at all of that period of the boy's life).

Let us begin with Reid's objection. Plainly the objection is decisive if the memory theory makes it a necessary and sufficient condition of someone's being the person who did a past action that he should remember that action. A defender of Locke might try to parry the objection by pointing out that what Locke requires for personal identity is that one *can* remember the action of the earlier person, not that one *does* remember it, and that it is plausible that under some possible circumstances (hypnosis, psychoanalysis) Reid's old general would remember the childhood incident, and so in that sense 'can' remember it. What Reid says, however, is that the old general had 'absolutely lost consciousness of the flogging', and it seems plausible to take the 'absolutely' as implying that the memory was lost without any possibility of recall. The example still seems possible under that interpretation.

Plainly the simple Lockean theory must be revised. The standard revision to meet this difficulty is most conveniently put in the person-stage terminology. Take the simple Lockean theory to hold that two person-stages belong to the same person if and only if the later contains memories (from the inside) of experiences, etc., contained in the earlier one. Here we should allow that one's current person-stage contains a memory of something even if one has temporarily forgotten that thing, as long as one has the potentiality of remembering it. In such a case the stage will retain a 'memory trace' that is the basis of that potentiality. Let us say that two person-stages that are so related are 'memory-connected'. The revised Lockean account says that the unity relation for persons is not the relation of being memory-connected but the 'ancestral' of this relation. This comes to saying that two stages belong to the same person if and only if they are the end-points of a series of stages such that each member of the series is memory-connected with the preceding member. One such series consists of the stage of the boy at the time of his flogging, the stage of the young officer at the time of his valiant deed, and a stage of the old general at a time at which he remembers the valiant deed but not the flogging. What this account makes necessary for identity with a 'past self' is not that one remember the actions and experiences of that past self but that one have 'memory continuity' with that past self – memory continuity consisting in the occurrence of a chain of memory-connected person-stages of the sort just described.¹²

Rather than address myself directly to Butler's objection, which seems to me to attack something Locke never said, I shall consider some circularity objections, perhaps descendants of Butler's, which seem to me more fundamental. It is arguable, first of all, that, far from personal identity being definable in terms of memory, memory must be defined in terms of personal identity. A definition of what it is for a person S to remember a particular event E could be expected to include the provisions (1) that S now has a state (which could be dispositional) which could be called an apparent memory, and (2) that the content of that apparent memory 'matches' in an appropriate way the nature of the past event E. But it is obvious that these conditions are not sufficient; if your first haircut was exactly like mine, I do not automatically remember yours in remembering

mine. The obvious remedy is to supplement conditions (1) and (2) with the additional requirement (3) that S was appropriately related to E at the time of its occurrence, i.e., he witnessed it, underwent it (if it was a haircut), performed it (if it was an action), and so on. But if the definition of ‘S remembers E’ contains condition (3), then we cannot without circularity use the notion of event memory to define personal identity, since (3) implicitly invokes the notion of personal identity – it implies that the person S who now satisfies condition (1) is *the same person* as someone in the past who was involved in a certain way with event E. It may further be argued (and this is perhaps closer to what Butler had in mind) that the particular sort of memory I have called ‘remembering from the inside’ cannot be characterized without the use of the notion of personal identity. For it may be claimed that to say that someone remembers *doing* an action, or *having* an experience, is elliptical for saying that he remembers *himself* doing the action or having the experience, and that for this reason these locutions (and the notion of ‘remembering from the inside’ which is explained in terms of them) cannot without circularity be used in an account of personal identity.

The first step towards answering these objections is to see that the addition of condition (3) to conditions (1) and (2) does not give us a sufficient condition for the truth of ‘S remembers E’. There can be memory illusions, and it is perfectly possible for an illusory memory to happen to correspond to something that happened to its subject in the past. For example, a hypnotist induces in me an apparent memory of having visited Yellowstone Park as a child, and it just so happens (unbeknownst to the hypnotist) that I did visit Yellowstone Park as a child, but have completely forgotten about it. The case I want is not one in which the hypnotist brings to consciousness a latent memory which was already present, but rather one in which my apparent memory of the visit to Yellowstone is entirely due to the hypnotist’s suggestion, and not at all due to my childhood visit to the Park – i.e., it is such that I would have had the very same apparent memory even if I had never visited Yellowstone. In this case I clearly do not remember the visit to Yellowstone, even though conditions (1)–(3) are all satisfied. What this brings out is that the notion of memory is a *causal* notion; it is a necessary condition of a person’s remembering a past event that his apparent memory of that event should be caused, in an appropriate way, by that event itself.¹³

What this may seem to call for is the replacement of condition (3) by something like this: (3’) S’s apparent memory (mentioned in (1)) was caused, in an appropriate way, by his experiencing (or otherwise being involved in) E at the time of its occurrence. Of course, (3’) implicitly invokes the notion of personal identity in the same way that (3) does, and so does not get us out of the circularity. But it is not obvious that we cannot formulate the causal requirement, and make it do the work done by (3) and (3’), without invoking the notion of personal identity. The requirement might be: (3’’) S’s apparent memory (mentioned in (1)) was caused, in an appropriate way, by someone’s experiencing (or otherwise being involved in) E at the time of its occurrence. Replacing ‘his’ (or ‘S’s’) with ‘someone’s’ eliminates the circularity – or rather, it does so *if* the

notion of ‘being caused in an appropriate way’ can be spelled out without invoking the notion of personal identity.

That some such phrase as ‘in an appropriate way’ is needed in the causal condition can be seen from a modification of my Yellowstone Park example. Suppose that at some time in my life a traumatic experience completely obliterated my memory of that incident. Prior to that time, however, I told someone about the visit, on the basis of the accurate memory I then had of it, and that person subsequently told the hypnotist about it. What the hypnotist did was instill in me an apparent memory which corresponded to the account he was given of my visit, an account which originally came from me. But he did not restore or revive my memory – it was irretrievably gone. Rather, he instilled in me a memory illusion which he could just as easily have instilled in someone who had never been to Yellowstone. Yet my apparent memory not only ‘matches’ the past event but is traceable back to it by a causal chain that goes through the hypnotist and his informant back to my earlier memory of the event, and via that back to the event itself. Here we have a causal connection, but not one of the ‘appropriate sort’.

If ‘causal connection of the appropriate sort’ could only be explained as meaning something like ‘connection via a causal chain that does not go outside the states of a single person’, obviously (3”) would invoke the notion of personal identity as much as (3) did, and the circularity objection would stand. However, there are reasons for thinking that this is not so.

One thing that seems clear is that we can *know* that there is a causal connection of the appropriate sort, and therefore that a person remembers a past event, without *first* knowing the relevant fact of personal identity (i.e., that the rememberer is identical to someone who experienced the remembered past event). For consider again the brain transfer example described earlier. If Brownson does indeed manifest apparent memories of Brown’s past life, the fact that he has Brown’s brain would seem to provide sufficient reason for thinking that these memories are ‘caused in an appropriate way’ by Brown’s past actions and experiences, and thus that Brownson really does remember those actions and experiences, which in turn can serve as a basis for saying that Brownson is Brown. If we had to settle the question of identity prior to discovering whether Brownson’s apparent memories really are memories of Brown’s life, we could not use an affirmative answer to the latter question as a basis for an affirmative answer to the former. Yet it seems that we can do this.

Moreover, there seem to be conceivable cases in which we can have the appropriate sort of causal connection for memory in the absence of identity between the rememberer and the person who experienced the remembered event. It may seem that if such cases are possible, they get us out of the circularity problem only at the cost of falsifying the memory theory; but we will see that the memory theory can easily be modified so as to accommodate such cases. The possible cases I have in mind are what have been called cases of ‘fission’ – cases (entirely imaginary, of course) in which a person somehow divides into two persons. The most realistic such case is one described by David Wiggins.¹⁴ He imagines a complex brain-transplant in which the two hemispheres of some-

one's brain are transplanted, separately, into the skulls of two different bodies. The result of the operation, we will suppose, is that both offshoots have memories from the inside of the life of the original person. It would be difficult to maintain that each offshoot is identical to the original person, since plainly they are not identical to each other (after the operation they go their separate ways, and soon can be distinguished on psychological grounds as well as by spatial location and physical properties).¹⁵ And it would seem arbitrary to suggest that one is identical to the original person and the other not.¹⁶ Yet it would be hard to deny that the apparent memories both have of the original person's life are genuine memories, and are related to events in that past life by causal connections 'of the appropriate sort'.

It is easy to see how the Lockean memory theory can be modified to allow for this case. We will simply say that memory continuity is sufficient for personal identity as long as there is no 'branching' in the chain of person-stages – where the fission case illustrates what is meant by 'branching'. I will have more to say about the fission example later on. Its importance here is that its possibility seems to count against the claim that the 'appropriate sort' of causal connection for memory can only be characterized as one that involves personal identity, and thus helps to defuse the circularity objection.

The fission case also helps with the objection that memory 'from the inside' can only be characterized as the sort of memory a person has of his own past, and thus that any attempt to define personal identity in terms of it will be circular. If the offshoots in the fission case are not identical to the original person, then their memories of that person's actions and experiences will not be memories of their own actions and experiences. Yet, for all that, they will remember those actions and experiences in the way in which normally one can remember only one's own actions and experiences. This is remembering from the inside, and it is something which in principle we can have in the absence of personal identity.

It will be seen that I have abandoned the claims, tentatively made in section 3, that it is necessarily the case that one can remember a past event only if one was a witness to it, and can remember a past experience or action from the inside only if it was an experience or action of one's own. An alternative approach, which differs only verbally from that presented here, is to hold on to these claims and to introduce a technical term, 'quasi-remember', for the notion defined in terms of conditions (1), (2) and (3''). The definition of personal identity will then be in terms of quasi-remembering (quasi-memory) rather than in terms of remembering (memory). Remembering will now be a special case of quasi-remembering; it will be quasi-remembering in which the rememberer is identical to the person who experienced or underwent the quasi-remembered event.¹⁷

There remains, however, an objection to the memory theory which is simpler and more direct than those considered so far. This is just that it seems conceivable that someone should survive total amnesia, total loss of memory.¹⁸ If that happens, there will be no chain of memory-connected person-stages going from stages prior to the onset of the amnesia to stages subsequent to it. So this pos-

sibility counts against the modified Lockean theory as well as against the simple theory.

We must be careful here about what is meant by ‘amnesia’. What is ordinarily meant is a total or partial loss of memory which can be recovered from; it is a condition that is treatable by hypnosis and in other ways. The possibility of total amnesia in this sense is no threat to the memory theory, since the amnesia victim can be presumed to retain memories in latent form. What is needed to refute the memory theory is the possibility of what might be called ‘philosophical amnesia’, i.e., the irreversible loss of all memory of the past.

But in addition to distinguishing different sorts of amnesia we must distinguish different sorts of memory. What we have meant by memory up until now is what is sometimes called ‘event memory’ – memory of particular events in the past. But there are other sorts of memory that are equally important. There is ‘factual memory’ – remembering that De Soto discovered the Mississippi River, that sulphuric acid is H_2SO_4 , that there will be an eclipse tomorrow. There is remembering how to do something – ride a bicycle, tie a bow tie, etc. There is remembering the meaning of ‘soigner’, the smell of lilacs, and so on. If the claim that philosophical amnesia is possible means that someone can suffer total and irreversible loss of memories of all past events, so that from a certain time onwards the person has, and can have, no memory of events prior to that time, then I think that we must allow that this is possible. But if ‘philosophical amnesia’ is taken to mean total and irretrievable loss of all memories of all kinds, then the claim that a person can survive such amnesia is far more questionable. For what we are now imagining is something close to what has been called a ‘brain zap’ – the total destruction of all of the effects of the person’s past experience, learning, reasoning, deliberation, and so on.¹⁹ Whether it is physiologically possible that a human body should survive a brain zap and remain alive and capable of realizing a mental life of a human sort seems questionable, to say the least. But let us suppose this is possible. Suppose that in a terrible accident a person suffers brain damage amounting to a total brain zap, and that somehow the surgeons manage to repair the brain in such a way that its possessor is able to start again, as it were, as if he were an infant. Eventually that body is again the body of someone with the mental life of a mature human being; but it is someone whose conception of the world, along with his personality and character, was formed by the experience of that body since the time of the accident and the reconstitution of the brain. It is anything but obvious that this person would be the person who had the body prior to the accident. So if total amnesia means this sort of brain zap, it is far from uncontroversial – indeed it seems just false – that it is something a person could survive.

But perhaps a total and irreversible loss of all memories need not amount to a brain zap. A person’s personality, character, tastes, interests and so on are the product (at least in part) of his past experience, and it is not obvious that the loss of all memories would necessarily involve the loss of all such traits as these. To be sure, to a certain extent personality and character traits do seem inseparable from memories of certain kinds. It is hard to see how someone’s pacifism could survive his loss of all of his beliefs about the effects of warfare. However,

let us assume that there are some traits of personality that could, at least in principle, survive a total and irreversible loss of all memories. If that is so, such a loss of memory would not necessarily amount to a total brain zap; and then it becomes more plausible to suppose that such a loss of memory is something a person could survive – in which case the memory theory, even as revised, is false.

5 Personal Identity as Psychological Continuity

This requires us to consider something we would have had to consider anyhow, namely the role *vis-à-vis* personal identity of kinds of psychological continuity other than memory continuity – I mean continuity with respect to the sorts of traits just mentioned: interests, tastes, talents, and traits of personality and character. Let us return to the Brown–Brownson case. If Brownson's possession of Brown's brain makes it plausible that he will have memories from the inside of Brown's past life, it makes it equally plausible that he will resemble Brown psychologically in all of the ways one expects a person on one day to resemble himself as he was the day before, and this resemblance would certainly be part of our reason for regarding Brownson as the same person as Brown. Suppose just for the moment that while Brownson's memories-from-the-inside are all of Brown's past, his personality and character traits are those of the old Robinson; I think that in this case (which would be physiologically unintelligible, and perhaps psychologically unintelligible as well) we would be much more hesitant about identifying Brownson with Brown.

We know, of course, that different people can share personality and character traits. And this may seem a reason for saying that Brownson's similarity to Brown with respect to such traits could not be part of what constitutes his identity with Brown, even though it might be evidence for it. This may suggest that, conceptually speaking, memory continuity is much more intimately related to personal identity than is similarity and continuity of personality. But all of this ignores the fact that what we have in the Brown–Brownson case is not merely similarity of personality and character. Brownson does not merely have the same personality traits as Brown did; he has those traits *because* Brown's life was such as to lead him to acquire such traits. The fact that Brownson has Brown's brain gives us reason to suppose that there is a relationship of causal or counterfactual dependence between Brownson's traits subsequent to the brain transfer and Brown's traits prior to it – we have reason to think that if Brown's traits had been different, Brownson's traits would have been different in corresponding ways. It is precisely when the circumstances are such that evidence of similarity is evidence of such a causal or counterfactual dependence that evidence of similarity is evidence of identity. Indeed, it is for the same reason that the nature of Brownson's memories is evidence that he is Brown; we have reason to think that if Brown's life had been different, Brownson's memories would have been correspondingly different, and thus that Brownson's memories are causally and counterfactually dependent on Brown's past life. Thus the status of similarity

and continuity of personality traits as evidence of personal identity seems no different than that of memory continuity; both are evidence only in so far as they include, or are evidence for, causal relations between earlier and later states.

Henceforth I shall use the term ‘psychological continuity’ to cover both of these sorts of causally grounded continuity. The memory continuity account of personal identity thus gives way to a more general psychological continuity account.²⁰ Memory continuity is now seen as just a special case of psychological continuity, and it is in psychological continuity that personal identity is now held to consist. Reverting to the ‘person-stage’ terminology, two person-stages will be directly connected, psychologically, if the later of them contains a psychological state (a memory impression, personality trait, etc.) which stands in the appropriate relation of causal dependence to a state contained in the earlier one; and two stages belong to the same person if and only if (1) they are connected by a series of stages such that each member of the series is directly connected, psychologically, to the immediately preceding member, and (2) no such series of stages which connects them ‘branches’ at any point, i.e., contains a member which is directly connected, psychologically, to two different stages occurring at the same time.

It is not peculiar to persons that their identity over time involves there being relationships of causal or counterfactual dependence between successive stages. The same is true of continuants generally. It is, I think, a point in favour of the psychological continuity account of personal identity that it can be seen as applying to the special case of personal identity an account of identity through time – call it the ‘causal continuity account’ – which holds of continuants generally.²¹ . . .

Notes

- 1 Thomas Nagel, ‘Brain Bisection and the Unity of Consciousness’, in Nagel, *Mortal Questions* (Cambridge, 1979).
- 2 Reported by Saul Kripke, in a lecture.
- 3 John Perry, *Personal Identity* (Berkeley, 1975), pp. 100–1.
- 4 See Perry, ‘The Problem of Personal Identity’, in his *Personal Identity*.
- 5 Russell, ‘The Philosophy of Logical Atomism’, in R. C. Marsh (ed.), *Logic and Knowledge* (London, 1956), p. 277.
- 6 Locke’s *Essay* was first published in 1690, but the chapter on ‘Identity and Diversity’ was added in the second edition, which appeared in 1694.
- 7 See Shoemaker, ‘Persons and their Pasts’, *American Philosophical Quarterly*, 7 (1970), pp. 269–85; see esp. p. 180.
- 8 Locke, *Essay Concerning Human Understanding*, ed. P. H. Nidditch, p. 340.
- 9 See S. Shoemaker, *Self-Knowledge and Self-Identity* (Ithaca, N.Y., 1963), pp. 22–5.
- 10 Perry, *Personal Identity*, p. 100.
- 11 Ibid., pp. 114–15.
- 12 For more precise formulations of such modified Lockean accounts, see H. P. Grice, ‘Personal Identity’, *Mind*, 50 (1941), pp. 330–50; and Perry, ‘The Problem of Personal Identity’.

- 13 See Max Deutscher and C. B. Martin, 'Remembering', *The Philosophical Review*, 75 (1966), pp. 161–97.
 - 14 David Wiggins, *Identity and Spatio-temporal Continuity* (Oxford, 1967), p. 53.
 - 15 For attempts to circumvent this difficulty, see John Perry, 'Can the Self Divide?', *Journal of Philosophy*, 69 (1972), pp. 463–88; and David Lewis, 'Survival and Identity', in Amélie Rorty (ed.), *The Identities of Persons* (Berkeley, Cal., 1976).
 - 16 It is really only for materialists (or anti-dualists) that this would be arbitrary. A believer in indivisible immaterial souls would of course insist that at most one of the offshoots would inherit the soul of the original person.
 - 17 This was my approach in Shoemaker, 'Persons and their Pasts'.
 - 18 See David Wiggins, *Sameness and Substance* (Oxford, 1980), p. 167 and pp. 176–7.
 - 19 Perry, review of Bernard Williams, *Problems of the Self*, *Journal of Philosophy*, 73 (1976), pp. 416–18; see esp. p. 421.
 - 20 A psychological continuity account is given in Anthony Quinton, 'The Soul', *Journal of Philosophy*, 59 (1962), pp. 393–403.
 - 21 See Sydney Shoemaker, 'Identity, Properties and Causality', *Midwest Studies in Philosophy*, no. 4 (Minneapolis, 1979), pp. 321–42.
-

36 Divided Minds and the Nature of Persons*

Derek Parfit

It was the split-brain cases which drew me into philosophy. Our knowledge of these cases depends on the results of various psychological tests, as described by Donald MacKay.¹ These tests made use of two facts. We control each of our arms, and see what is in each half of our visual fields, with only one of our hemispheres. When someone's hemispheres have been disconnected, psychologists can thus present to this person two different written questions in the two halves of his visual field, and can receive two different answers written by this person's two hands.

Here is a simplified imaginary version of the kind of evidence that such tests provide. One of these people looks fixedly at the centre of a wide screen, whose left half is red and right half is blue. On each half in a darker shade are the words, 'How many colours can you see?' With both hands the person writes, 'Only one'. The words are now changed to read, 'Which is the only colour that you can see?' With one of his hands the person writes 'Red', with the other he writes 'Blue'.

If this is how such a person responds, I would conclude that he is having two

* From Derek Parfit, 'Divided Minds and the Nature of Persons', in Colin Blakemore and Susan Greenfield, eds, *Mindwaves* (Oxford: Blackwell, 1987). Reprinted by permission of the author and publisher.

visual sensations – that he does, as he claims, see both red and blue. But in seeing each colour he is not aware of seeing the other. He has two streams of consciousness, in each of which he can see only one colour. In one stream he sees red, and at the same time, in his other stream, he sees blue. More generally, he could be having at the same time two series of thoughts and sensations, in having each of which he is unaware of having the other.

This conclusion has been questioned. It has been claimed by some that there are not *two* streams of consciousness, on the ground that the subdominant hemisphere is a part of the brain whose functioning involves no consciousness. If this were true, these cases would lose most of their interest. I believe that it is not true, chiefly because, if a person's dominant hemisphere is destroyed, this person is able to react in the way in which, in the split-brain cases, the sub-dominant hemisphere reacts, and we do not believe that such a person is just an automaton, without consciousness. The sub-dominant hemisphere is, of course, much less developed in certain ways, typically having the linguistic abilities of a three-year-old. But three-year-olds are conscious. This supports the view that, in split-brain cases, there *are* two streams of consciousness.

Another view is that, in these cases, there are two persons involved, sharing the same body. Like Professor MacKay, I believe that we should reject this view. My reason for believing this is, however, different. Professor MacKay denies that there are two persons involved because he believes that there is only one person involved. I believe that, in a sense, the number of persons involved is none.

The Ego Theory and the Bundle Theory

To explain this sense I must, for a while, turn away from the split-brain cases. There are two theories about what persons are, and what is involved in a person's continued existence over time. On the *Ego Theory*, a person's continued existence cannot be explained except as the continued existence of a particular *Ego*, or *subject of experiences*. An Ego Theorist claims that, if we ask what unifies someone's consciousness at any time – what makes it true, for example, that I can now both see what I am typing and hear the wind outside my window – the answer is that these are both experiences which are being had by me, this person, at this time. Similarly, what explains the unity of a person's whole life is the fact that all of the experiences in this life are had by the same person, or subject of experiences. In its best-known form, the *Cartesian view*, each person is a persisting purely mental thing – a soul, or spiritual substance.

The rival view is the *Bundle Theory*. Like most styles in art – Gothic, baroque, rococo, etc. – this theory owes its name to its critics. But the name is good enough. According to the Bundle Theory, we can't explain either the unity of consciousness at any time, or the unity of a whole life, by referring to a person. Instead we must claim that there are long series of different mental states and events – thoughts, sensations, and the like – each series being what we call one life. Each series is unified by various kinds of causal relations, such as the

relations that hold between experiences and later memories of them. Each series is thus like a bundle tied up with string.

In a sense, a Bundle Theorist denies the existence of persons. An outright denial is of course absurd. As Reid protested in the eighteenth century, ‘I am not thought, I am not action, I am not feeling; I am something which thinks and acts and feels.’ I am not a series of events, but a person. A Bundle Theorist admits this fact, but claims it to be only a fact about our grammar, or our language. There are persons or subjects in this language-dependent way. If, however, persons are believed to be more than this – to be separately existing things, distinct from our brains and bodies, and the various kinds of mental states and events – the Bundle Theorist denies that there are such things.

The first Bundle Theorist was Buddha, who taught ‘anatta’, or the *No Self view*. Buddhists concede that selves or persons have ‘nominal existence’, by which they mean that persons are merely combinations of other elements. Only what exists by itself, as a separate element, has instead what Buddhists call ‘actual existence’. Here are some quotations from Buddhist texts:

At the beginning of their conversation the king politely asks the monk his name, and receives the following reply: ‘Sir, I am known as ‘Nagasena’; my fellows in the religious life address me as ‘Nagasena’. Although my parents gave me the name . . . it is just an appellation, a form of speech, a description, a conventional usage. ‘Nagasena’ is only a name, for no person is found here.’

A sentient being does exist, you think, O Mara? You are misled by a false conception. This bundle of elements is void of Self. In it there is no sentient being. Just as a set of wooden parts Receives the name of carriage, So do we give to elements The name of fancied being.

Buddha has spoken thus: ‘O Brethren, actions do exist, and also their consequences, but the person that acts does not. There is no one to cast away this set of elements, and no one to assume a new set of them. There exists no Individual, it is only a conventional name given to a set of elements.’²

Buddha’s claims are strikingly similar to the claims advanced by several Western writers. Since these writers knew nothing of Buddha, the similarity of these claims suggests that they are not merely part of one cultural tradition, in one period. They may be, as I believe they are, true.

What We Believe Ourselves to Be

Given the advances in psychology and neurophysiology, the Bundle Theory may now seem to be obviously true. It may seem uninteresting to deny that there are separately existing Egos, which are distinct from brains and bodies and the various kinds of mental states and events. But this is not the only issue. We may be convinced that the Ego Theory is false, or even senseless. Most of us, however, even if we are not aware of this, also have certain beliefs about what is

involved in our continued existence over time. And these beliefs would only be justified if something like the Ego Theory was true. Most of us therefore have false beliefs about what persons are, and about ourselves.

These beliefs are best revealed when we consider certain imaginary cases, often drawn from science fiction. One such case is *teletransportation*. Suppose that you enter a cubicle in which, when you press a button, a scanner records the states of all of the cells in your brain and body, destroying both while doing so. This information is then transmitted at the speed of light to some other planet, where a replicator produces a perfect organic copy of you. Since the brain of your Replica is exactly like yours, it will seem to remember living your life up to the moment when you pressed the button, its character will be just like yours, and it will be in every other way psychologically continuous with you. This psychological continuity will not have its normal cause, the continued existence of your brain, since the causal chain will run through the transmission by radio of your 'blueprint'.

Several writers claim that if you chose to be teletransported, believing this to be the fastest way of travelling, you would be making a terrible mistake. This would not be a way of travelling, but a way of dying. It may not, they concede, be quite as bad as ordinary death. It might be some consolation to you that, after your death, you will have this Replica, which can finish the book that you are writing, act as parent to your children, and so on. But, they insist, this Replica won't be you. It will merely be someone else, who is exactly like you. This is why this prospect is nearly as bad as ordinary death.

Imagine next a whole range of cases, in each of which, in a single operation, a different proportion of the cells in your brain and body would be replaced with exact duplicates. At the near end of this range, only 1 or 2 per cent would be replaced; in the middle, 40 or 60 per cent; near the far end, 98 or 99 per cent. At the far end of this range is pure teletransportation, the case in which all of your cells would be 'replaced'.

When you imagine that some proportion of your cells will be replaced with exact duplicates, it is natural to have the following beliefs. First, if you ask, 'Will I survive? Will the resulting person be me?', there must be an answer to this question. Either you will survive, or you are about to die. Second, the answer to this question must be either a simple 'Yes' or a simple 'No'. The person who wakes up either will or will not be you. There cannot be a third answer, such as that the person waking up will be half you. You can imagine yourself later being half-conscious. But if the resulting person will be fully conscious, he cannot be half you. To state these beliefs together: to the question, 'Will the resulting person be me?', there must always be an answer, which must be all-or-nothing.

There seem good grounds for believing that in the case of teletransportation, your Replica would not be you. In a slight variant of this case, your Replica might be created while you were still alive, so that you could talk to one another. This seems to show that, if 100 per cent of your cells were replaced, the result would merely be a Replica of you. At the other end of my range of cases, where only 1 per cent would be replaced, the resulting person clearly *would* be you. It therefore seems that, in the cases in between, the resulting person must

be either you, or merely a Replica. It seems that one of these must be true, and that it makes a great difference which is true.

How We are Not What We Believe

If these beliefs were correct, there must be some critical percentage, somewhere in this range of cases, up to which the resulting person would be you, and beyond which he would merely be your Replica. Perhaps, for example, it would be you who would wake up if the proportion of cells replaced were 49 per cent, but if just a few more cells were also replaced, this would make all the difference, causing it to be someone else who would wake up.

That there must be some such critical percentage follows from our natural beliefs. But this conclusion is most implausible. How could a few cells make such a difference? Moreover, if there is such a critical percentage, no one could ever discover where it came. Since in all these cases the resulting person would believe that he was you, there could never be any evidence about where, in this range of cases, he would suddenly cease to be you.

On the Bundle Theory, we should reject these natural beliefs. Since you, the person, are not a separately existing entity, we can know exactly what would happen without answering the question of what will happen to you. Moreover, in the cases in the middle of my range, it is an empty question whether the resulting person would be you, or would merely be someone else who is exactly like you. These are not here two different possibilities, one of which must be true. These are merely two different descriptions of the very same course of events. If 50 per cent of your cells were replaced with exact duplicates, we could call the resulting person you, or we could call him merely your Replica. But since these are not here different possibilities, this is a mere choice of words.

As Buddha claimed, the Bundle Theory is hard to believe. It is hard to accept that it could be an empty question whether one is about to die, or will instead live for many years.

What we are being asked to accept may be made clearer with this analogy. Suppose that a certain club exists for some time, holding regular meetings. The meetings then cease. Some years later, several people form a club with the same name, and the same rules. We can ask, ‘Did these people revive the very same club? Or did they merely start up another club which is exactly similar?’ Given certain further details, this would be another empty question. We could know just what happened without answering this question. Suppose that someone said: ‘But there must be an answer. The club meeting later must either be, or not be, the very same club.’ This would show that this person didn’t understand the nature of clubs.

In the same way, if we have any worries about my imagined cases, we don’t understand the nature of persons. In each of my cases, you would know that the resulting person would be both psychologically and physically exactly like you, and that he would have some particular proportion of the cells in your brain and body – 90 per cent, or 10 per cent, or, in the case of teletransportation, 0 per

cent. Knowing this, you know everything. How could it be a real question what would happen to you, unless you are a separately existing Ego, distinct from a brain and body, and the various kinds of mental state and event? If there are no such Egos, there is nothing else to ask a real question about.

Accepting the Bundle Theory is not only hard; it may also affect our emotions. As Buddha claimed, it may undermine our concern about our own futures. This effect can be suggested by redescribing this change of view. Suppose that you are about to be destroyed, but will later have a Replica on Mars. You would naturally believe that this prospect is about as bad as ordinary death, since your Replica won't be you. On the Bundle Theory, the fact that your Replica won't be you just consists in the fact that, though it will be fully psychologically continuous with you, this continuity won't have its normal cause. But when you object to teletransportation you are not objecting merely to the abnormality of this cause. You are objecting that this cause won't get *you* to Mars. You fear that the abnormal cause will fail to produce a further and all-important fact, which is different from the fact that your Replica will be psychologically continuous with you. You do not merely want there to be psychological continuity between you and some future person. You want to *be* this future person. On the Bundle Theory, there is no such special further fact. What you fear will not happen, in this imagined case, *never* happens. You want the person on Mars to be you in a specially intimate way in which no future person will ever be you. This means that, judged from the standpoint of your natural beliefs, even ordinary survival is about as bad as teletransportation. *Ordinary survival is about as bad as being destroyed and having a Replica.*

How the Split-Brain Cases Support the Bundle Theory

The truth of the Bundle Theory seems to me, in the widest sense, as much a scientific as a philosophical conclusion. I can imagine kinds of evidence which would have justified believing in the existence of separately existing Egos, and believing that the continued existence of these Egos is what explains the continuity of each mental life. But there is in fact very little evidence in favour of this Ego Theory, and much for the alternative Bundle Theory.

Some of this evidence is provided by the split-brain cases. On the Ego Theory, to explain what unifies our experiences at any one time, we should simply claim that these are all experiences which are being had by the same person. Bundle Theorists reject this explanation. This disagreement is hard to resolve in ordinary cases. But consider the simplified split-brain case that I described. We show to my imagined patient a placard whose left half is blue and right half is red. In one of this person's two streams of consciousness, he is aware of seeing only blue, while at the same time, in his other stream, he is aware of seeing only red. Each of these two visual experiences is combined with other experiences, like that of being aware of moving one of his hands. What unifies the experiences, at any time, in each of this person's two streams of consciousness? What unifies his awareness of seeing only red with his awareness of moving one hand? The

answer cannot be that these experiences are being had by the same person. This answer cannot explain the unity of each of this person's two streams of consciousness, since it ignores the disunity between these streams. This person is now having all of the experiences in both of his two streams. If this fact was what unified these experiences, this would make the two streams one.

These cases do not, I have claimed, involve two people sharing a single body. Since there is only one person involved, who has two streams of consciousness, the Ego Theorist's explanation would have to take the following form. He would have to distinguish between persons and subjects of experiences, and claim that, in split-brain cases, there are *two* of the latter. What unifies the experiences in one of the person's two streams would have to be the fact that these experiences are all being had by the same subject of experiences. What unifies the experiences in this person's other stream would have to be the fact that they are being had by another subject of experiences. When this explanation takes this form, it becomes much less plausible. While we could assume that 'subject of experiences', or 'Ego', simply meant 'person', it was easy to believe that there are subjects of experiences. But if there can be subjects of experiences that are not persons, and if in the life of a split-brain patient there are at any time two different subjects of experiences – two different Egos – why should we believe that there really are such things? This does not amount to a refutation. But it seems to me a strong argument against the Ego Theory.

As a Bundle Theorist, I believe that these two Egos are idle cogs. There is another explanation of the unity of consciousness, both in ordinary cases and in split-brain cases. It is simply a fact that ordinary people are, at any time, aware of having several different experiences. This awareness of several different experiences can be helpfully compared with one's awareness, in short-term memory, of several different experiences. Just as there can be a single memory of just having had several experiences, such as hearing a bell strike three times, there can be a single state of awareness both of hearing the fourth striking of this bell, and of seeing, at the same time, ravens flying past the bell-tower.

Unlike the Ego Theorist's explanation, this explanation can easily be extended to cover split-brain cases. In such cases there is, at any time, not one state of awareness of several different experiences, but two such states. In the case I described, there is one state of awareness of both seeing only red and of moving one hand, and there is another state of awareness of both seeing only blue and moving the other hand. In claiming that there are two such states of awareness, we are not postulating the existence of unfamiliar entries, two separately existing Egos which are not the same as the single person whom the case involves. This explanation appeals to a pair of mental states which would have to be described anyway in a full description of this case.

I have suggested how the split-brain cases provide one argument for one view about the nature of persons. I should mention another such argument, provided by an imagined extension of these cases, first discussed at length by David Wiggins.³

In this imagined case a person's brain is divided, and the two halves are transplanted into a pair of different bodies. The two resulting people live quite sepa-

rate lives. This imagined case shows that personal identity is not what matters. If I was about to divide, I should conclude that neither of the resulting people will be me. I will have ceased to exist. But this way of ceasing to exist is about as good – or as bad – as ordinary survival.

Some of the features of Wiggins's imagined case are likely to remain technically impossible. But the case cannot be dismissed, since its most striking feature, the division of one stream of consciousness into separate streams, has already happened. This is a second way in which the actual split-brain cases have great theoretical importance. They challenge some of our deepest assumptions about ourselves.⁴

Notes

- 1 See Donald MacKay, 'Divided Brains – Divided Minds?', chapter 1 of *Mindwaves*, ed. Colin Blakemore and Susan Greenfield (Oxford: Blackwell, 1987).
- 2 For the sources of these and similar quotations, see my *Reasons and Persons* (Oxford: Oxford University Press, 1984), pp. 502–3, 532.
- 3 At the end of his *Identity and Spatio-temporal Continuity* (Oxford: Blackwell, 1967).
- 4 I discuss these assumptions further in part 3 of my *Reasons and Persons*.

37 Personal Identity: the Dualist Theory*

Richard Swinburne

1 Empiricist Theories

There are two philosophical questions about personal identity. The first is: what are the logically necessary and sufficient conditions for a person P_2 at a time, t_2 being the same person as a person P_1 at an earlier time t_1 ,¹ or, loosely, what does it mean to say that P_2 is the same person as P_1 ? The second is: what evidence of observation and experience can we have that a person P_2 at t_2 is the same person as a person P_1 at t_1 (and how are different pieces of evidence to be weighed against each other)? Many writers about personal identity have, however, needed to give only one account of personal identity, because their account of the logically necessary and sufficient conditions of personal identity was in terms of the evidence of observation and experience which would establish or oppose claims of personal identity. They have made no sharp distinction between the meaning of such claims and the evidence which supported them. Theories of this kind we may call empiricist theories.

* From Sydney Shoemaker and Richard Swinburne, *Personal Identity* (Oxford: Blackwell, 1984). Reprinted by permission of the author.

In this section I shall briefly survey the empiricist theories which have been offered and argue that they are ultimately unsatisfactory, and so go on to argue that my two questions have very different answers. What we mean when we say that two persons are the same is one thing; the evidence which we may have to support our claim is something very different.

The most natural theory of personal identity which readily occurs to people, is that personal identity is constituted by bodily identity. P_2 is the same person as P_1 if P_2 's body is the same body as P_1 's body. The person to whom you are talking now and call 'John' is the same person as the person to whom you were talking last week and then called 'John' if and only if he has the same body. To say that the two bodies – call them B_1 and B_2 – are the same is not to say that they contain exactly the same bits of matter. Bodies are continually taking in new matter (by people eating and drinking and breathing in) and getting rid of matter. But what makes the bodies the same is that the replacement of matter is only gradual. The matter which forms my body is organized in a certain way, into parts – legs, arms, heart, liver, etc. – which are interconnected and exchange matter and energy in regular ways. What makes my body today the same body as my body yesterday is that most of the matter is the same (although I may have lost some and gained some) and its organization has remained roughly the same.

This bodily theory of personal identity gives a somewhat similar account of personal identity to the account which it is natural to give of the identity of any material object or plant, and which is due ultimately to Aristotle (*Metaphysics*, Book 7). Aristotle distinguished between substances and properties. Substances are the individual things, like tables and chairs, cars and plants, which have properties (such as being square or round or red). Properties are 'universals', that is they can be possessed by many different substances; many different substances can be square or red. Substances are the individual substances which they are because of the matter out of which they are made and the form which is given to that matter. By 'the form' is meant those properties (normally of shape and organization) the possession of which is essential if a substance is to be the substance in question, the properties which it cannot lose without ceasing to exist. We thus distinguish between the essential properties of a substance – those which constitute its form – and the accidental properties of a substance. It is among the essential properties of a certain oak tree that it has, under normal conditions, a certain general shape and appearance, a certain life cycle (of producing leaves in spring and acorns in autumn); but its exact height, its position, and the distribution of leaves on its tallest branch are accidental properties. If the matter of the oak tree is reduced to a heap of planks, the oak tree, lacking its essential properties, has ceased to exist. We think of substances as belonging to different kinds, natural – e.g., oak trees or ferns; or artificial – e.g., cars or desks; and the defining properties of a kind constitute the form of a substance which belongs to it. . . .

What makes a substance the same substance as an earlier substance is that its matter is the same, or obtained from the matter of the former substance by gradual replacement, while continuing to possess the essential properties which

constitute its form. The table at which I am writing today is the same table at which I was writing yesterday because it consists of the same matter (or at any rate, most of the same matter), organized in the same way – into the form of a table. For inanimate things, however, too much replacement of matter, however gradual, will destroy identity. If I replace the drawer of my desk by another drawer, the desk remains the same desk. But if, albeit gradually, I replace first the drawers and then the sides and then the top, so that there is none of the original matter left, we would say that the resulting desk was no longer the same desk as the original desk. For living things, such as plants, total replacement of matter – so long as it is gradual, and so long as physiology and anatomy also change only gradually if at all – will not destroy identity. The oak tree is the same as the sapling out of which it has grown, because replacement of matter has been gradual, and form (i.e., shape, physiology, and behaviour) has been largely preserved while any changes in it have been gradual. . . .

Persons too are substances. (Men, or human beings, are persons of a certain kind – viz., those with similar anatomy, physiology, and evolutionary origin to ourselves. There may be persons, e.g., on another planet, who are not human beings.) If we apply Aristotle's general account of the identity of substances to persons, it follows that for a person to be the same person as an earlier person, he has to have the same matter (or matter obtained from that earlier person by gradual replacement) organized into the form of a person. The essential properties which make the form of a person would include, for Aristotle, not merely shape and physiological properties, but a kind of way of behaving and a capacity for a mental life of thought and feeling. For P_2 at t_2 to be the same person as P_1 at t_1 , both have to be persons (to have a certain kind of body and mental life) and to be made of the same matter (i.e., to be such that P_2 's body is obtained from P_1 's by gradual replacement of parts). Such is the bodily theory of personal identity. It does not deny that persons have a mental life, but insists that what makes a person the same person as an earlier person is sameness of body.

The difficulty which has been felt by those modern philosophers basically sympathetic to a bodily theory of personal identity is this. One part of the body – viz., the brain – seems to be of crucial importance for determining the characteristic behaviour of the rest. The brain controls not merely the physiology of the body but the way people behave and talk and think. If a man loses an arm or a leg, we do not think that the subsequent person is in any way different from the original person. If a man has a heart transplant or a liver transplant, again we do not think that the replacement makes a different person. On the other hand, if the brain of a person P_1 were removed from his body B_1 and transplanted into the skull of a body B_2 of a person P_2 , from which the brain was removed and then transplanted into the empty skull of B_1 (i.e., if brains were interchanged), we would have serious doubt whether P_1 had any more the same body. We would be inclined to say that the person went where his brain went – viz., that P_1 at first had body B_1 , and then, after the transplant, body B_2 . The reason why we would say this is that (we have very good scientific reason to believe) the person with B_2 's body would claim to be P_1 , to have done and experienced the things which we know P_1 to have done, and would have the character, beliefs,

and attitudes of P_1 . What determines my attitude towards a person is not so much the matter out of which his body is made, but who he claims to be, whether he has knowledge of my past life purportedly on the basis of previous acquaintance with me, and more generally what his beliefs about the world are and what are his attitudes towards it. Hence a philosopher seeking a materialist criterion of personal identity, will come to regard the brain, the core of the body, rather than the rest of the body as what matters for personal identity. So this modified bodily theory states: that P_2 is the same person as P_1 if and only if P_2 has the same central organ controlling memory and character, viz., same brain, as P_1 . Let us call it the brain theory of personal identity. A theory along these lines (with a crucial qualification, to be discussed shortly) was tentatively suggested by David Wiggins in *Identity and Spatiotemporal Continuity* (Oxford, 1967).²

The traditional alternative to a bodily theory of personal identity is the memory-and-character theory. This claims that, given the importance for our attitude towards persons of their memory claims and character, continuity in respect of these would constitute personal identity – whether or not this continuity is caused by continuity of some bodily organ, such as the brain; and the absence of continuity of memory and character in some particular case involves the absence of personal identity, even if there is continuity in respect of that bodily organ which produces such continuity between other persons on other occasions.

The simplest version of this theory was that given by John Locke. According to Locke, memory alone (or ‘consciousness’, as he often calls it) constitutes personal identity. Loosely – P_2 at t_2 is the same person as P_1 at an earlier time t_1 , if and only if P_2 remembers having done and experienced various things, where these things were in fact done and experienced by P_1 .

Before expounding Locke’s theory further we need to be clear about the kind of memory which is involved. First, it is what is sometimes called personal memory, i.e., memory of one’s own past experiences. It is thus to be distinguished from factual memory, which is memory of some fact known previously; as when I remember that the battle of Hastings was fought in 1066. This is not a memory of a past experience. . . . Secondly, it is personal memory in the weak sense. In the normal or strong sense of ‘remember’, one can only remember doing something if one really did it. I may say that I ‘remember’ going up the Eiffel Tower, but if I didn’t do it, it seems natural to say that I cannot really remember having done it. In this sense, just as you can only know what is true, so you can only remember what you really did. However, there is also a weak sense of ‘remember’ in which a man remembers whatever he believes that he remembers in the strong sense. One’s weak memories are not necessarily true ones. Now if the memory criterion defined personal identity in terms of memory in the strong sense, it would not be very useful; for to say that P_2 remembers having done what P_1 did would already entail their being the same person, and anyone in doubt as to whether P_2 was the same person as P_1 , would have equal doubt whether P_2 really did remember doing what P_1 did. What the criterion as stated is concerned with is memory in the weak sense, which (because the strong

sense is the more natural one) I shall henceforward call apparent memory.

So Locke's theory can now be rephrased as follows: P_2 at t_2 is the same person as P_1 at an earlier time t_1 , if and only if P_2 apparently remembers having done and experienced various things when those things were in fact done and experienced by P_1 . A person is who he thinks that he is. . . .

Locke's theory needs tidying up if we are to avoid absurdity. Consider, first, the following objection made by Thomas Reid:

Suppose a brave officer to have been flogged when a boy at school for robbing an orchard, to have taken a standard from the enemy in his first campaign, and to have been made a general in advanced life; suppose also, which must be admitted to be possible, that, when he took the standard, he was conscious of his having been flogged at school, and that, when made a general, he was conscious of his taking the standard, but had absolutely lost the consciousness of his flogging. These things being supposed, it follows, from Mr Locke's doctrine, that he who was flogged at school is the same person who took the standard, and that he who took the standard is the same person who was made a general. Whence it follows if there be any truth in logic, that the general is the same person with him who was flogged at school. But the general's consciousness does not reach so far back as his flogging; therefore according to Mr Locke's doctrine, he is not the same person who was flogged. Therefore the general is, and at the same time is not, the same person with him who was flogged at school. (Reid, *Essays on the Intellectual Powers of Man*, book III, ch. 6)

The objection illustrates the important point that identity is a transitive relation; if a is identical with b and b is identical with c , then necessarily a is identical with c . We can meet the objection by reformulating Locke's theory as follows: P_2 at t_2 is the same person as P_1 at an earlier time t_1 if and only if either P_2 apparently remembers what P_1 did and experienced, or he apparently remembers what some person P' at an intermediate time t' did and experienced, when P' apparently remembers what P_1 did and experienced, or they are linked by some longer intermediate chain. (That is, P_2 apparently remembers what P' did and experienced, P' apparently remembers what P'' did and experienced, and so on until we reach a person who apparently remembers what P_1 did and experienced.) If P_1 and P_2 are linked by such a chain, they are, we may say, linked by continuity of memory. Clearly, the apparent memories of the deeds and experiences of the previous person at each stage in the chain need not be completely accurate memories of what was done and experienced. But they do need to be fairly accurate memories of what was done and experienced, if the later person is to be the person who did and experienced those things. . . .

Many advocates of a memory theory have not always been very clear in their exposition about whether the apparent memories which form the links in the chain of memory need to be actual memories, or whether they need only to be hypothetical memories. By 'actual memories' I mean actual recallings of past experiences. The trouble with the suggestion that actual memories are required is that we do not very often recall our past, and it seems natural to suppose that the deeds and experiences of some moments of a person's life never get re-

called. Yet the memory theory, as stated so far, rules out that possibility. If I am not connected by a chain of memories with the deeds and experiences done by a person at a certain time, then I am not identical with that person. It is perhaps better if the theory claims that the apparent memories which form the links need only be hypothetical memories – i.e., what a person would apparently remember if he were to try to remember the deeds and experiences in question, e.g., in consequence of being prompted.

There is, however, a major objection to any memory theory of personal identity, arising from the possibility of duplication. The objection was made briefly by Reid and at greater length in an influential article by Bernard Williams. Williams imagines the case of a man whom he calls Charles who turns up in the twentieth-century claiming to be Guy Fawkes:

All the events he claims to have witnessed and all the actions he claims to have done point unanimously to the life-history of some one person in the past – for instance Guy Fawkes. Not only do all Charles' memory-claims that can be checked fit the pattern of Fawkes' life as known to historians, but others that cannot be checked are plausible, provide explanations of unexplained facts, and so on.³

The fact that memory claims which ‘cannot be checked are plausible, provide explanations of unexplained facts, and so on’ is evidence that Charles is not merely claiming to remember what he has in fact read in a book about Guy Fawkes, and so leaves us back with the supposition, natural to make in normal cases, that he is reporting honestly his apparent memories. So, by a memory theory Charles would be Guy Fawkes. But then suppose, Williams imagines, that another man Robert turns up, who satisfies the memory criteria for being Guy Fawkes equally well. We cannot say that they are both identical with Guy Fawkes, for if they were, they would be identical with each other – which they are not since they currently live different lives and have different thoughts and feelings from each other. So apparent memory cannot constitute personal identity, although it may be fallible evidence of it.

The objection from the possibility of duplication, together with other difficulties which will be mentioned in later chapters, have inclined the majority of contemporary writers to favour a theory which makes some sort of bodily continuity central to personal identity. As we have seen, the brain theory takes into account the insight of memory-and-character theory into the importance of these factors for personal identity, by selecting the brain, as the organ causally responsible for the continuity of memory and character, as that part of the body the continuity of which constitutes the continuity of the person.

The trouble is that any brain theory is also open to the duplication objection. The human brain has two very similar hemispheres – a left and a right hemisphere. The left hemisphere plays a major role in the control of limbs of, and processing of sensory information from, the right side of the body (and from the right sides of the two eyes); and the right hemisphere plays a major role in the control of limbs of, and processing of sensory information from, the left side of the body (and from the left sides of the two eyes). The left hemisphere

plays a major role in the control of speech. Although the hemispheres have different roles in the adult, they interact with each other; and if parts of a hemisphere are removed, at any rate early in life, the roles of those parts are often taken over by parts of the other hemisphere. Brain operations which remove substantial parts of the brain are not infrequent. It might be possible one day to remove a whole hemisphere, without killing the person. There are no logical difficulties in supposing that we could transplant one of P_1 's hemispheres into one skull from which a brain had been removed, and the other hemisphere into another such skull, and that both transplants should take, and it may well be practically possible to do so. It is certainly more likely to occur than the Guy Fawkes story told by Williams! If these transplants took, clearly each of the resulting persons would behave to some extent like P_1 , and indeed both would probably have some of the apparent memories of P_1 . Each of the resulting persons would then be good candidates for being P_1 .

After all, if one of P_1 's hemispheres had been destroyed and the other remained intact and untransplanted, and the resulting person continued to behave and make memory claims somewhat like those of P_1 , we would have had little hesitation in declaring that person to be P_1 . The same applies, whichever hemisphere was preserved – although it may well be that the resulting person would have greater capacities (e.g., speech) if one hemisphere was preserved than if the other one was preserved. We have seen earlier, good reason for supposing that the person goes where his brain goes, and if his brain consists only of one hemisphere, that should make no difference. So if the one remaining hemisphere is then transplanted, we ought to say that the person whose body it now controls is P_1 . Whether that person is P_1 can hardly be affected by the fact that instead of being destroyed, the other hemisphere is also transplanted so as to constitute the brain of person. But if it is, that other person will be just as good a candidate for being P_1 . So a Wiggins-type account might lead us to say that both resulting persons are P_1 . But, for the reason given earlier in connection with the Guy Fawkes examples, that cannot be – since the two later persons are not identical with each other. Hence, Wiggins adds to his tentative definition a clause stating that P_2 who satisfies his criterion stated earlier is the same person as P_1 , only if there is no other later person who also satisfies the criterion.⁴

But the introduction into any theory, whether a memory theory, a brain theory, or whatever, of a clause stating that a person who satisfies the criterion in question for being the same as an earlier person is the same, only so long as there is no other person who satisfies the criterion also or equally well, does have an absurd consequence. Let us illustrate this for the brain theory. Suppose P_1 's left hemisphere is transplanted into some skull and the transplant takes. Then, according to the theory, whether the resulting person is P_1 , i.e., whether P_1 survives, will depend on whether the other transplant takes. If it does, since both resulting persons will satisfy the memory and brain continuity criteria equally well, neither will be P_1 . But if the other transplant does not take, then since there is only one person who satisfies the criterion, that person is P_1 . So whether I survive an operation will depend on what happens in a body entirely different

from the body which will be mine, if I do survive. But how can who I am depend on what happens to you? A similar absurd consequence follows when a similar clause forbidding duplication is added to a memory theory.

Yet if we abandon the duplication clause, we are back with the original difficulty – that there may be more than one later person who satisfies any memory criterion or brain criterion, or combination thereof, for being the same person as an earlier person. Our discussion brings to our attention also the fact that both these criteria are criteria which may be satisfied to varying degrees. P_2 can have 90 per cent, or 80 per cent, or less than 50 per cent of the brain of P_1 ; and likewise the similarity of apparent memory and character may vary along a spectrum. Just how well do criteria have to be satisfied for the later person to be the same person as the earlier person? Any line one might draw seems totally artificial. One might think that it was non-arbitrary to insist on more than 50 per cent of the original brain matter – for only one later person could have more than 50 per cent of the original brain matter (whereas if our criterion demands only a smaller proportion, more than one later person could satisfy it). But would we really want to say that P_6 was the same person as P_1 if P_2 was obtained from P_1 by a transplant of 60 per cent (and so more than half) of P_1 's brain matter, P_3 was obtained from P_2 by a transplant of 60 per cent of P_2 's brain matter, and so on until we came to P_6 . By the criterion of 'more than half of the brain matter', P_6 would be the same person as P_5 , P_5 as P_4 and so on, and so by the transitivity of identity P_6 would be the same person as P_1 – although he would have very little of P_1 's brain matter. Any criterion of the proportion of brain matter transferred, to be plausible, would have to take account of whether there had been similar transplants in the past, and the length of the interval between them. And then the arbitrariness of the criterion would stare us in the face.

This problem pushes the thinker towards one of two solutions. The first solution is to say that personal identity is a matter of degree. P_2 is the same person as P_1 to the extent to which there is sameness of brain matter and continuity of memory. After all, survival for inanimate things is a matter of degree. As we gradually replace bits of a desk with new bits, the resulting desk is only more or less the same as the original desk. And if my car is taken to pieces and some of the bits are used to make one new car, and some of the bits used to make another new car, both cars are partly the same as and partly different from the old car. Why cannot we say the same of people? Normally we are not inclined to talk thus, because brain operations are rare and brain hemisphere transplants never happen. Hence there is normally at most only one candidate for being the same person as an earlier person, and he is normally a very strong candidate indeed – having a more or less identical brain and very great similarities of apparent memory and character. So we tend to think of personal identity as all or nothing. But it is not thus in its logic, the argument goes. There is the logical possibility, which could become an empirical possibility, of intermediate cases – of persons who are to some extent the same as and to some extent different from original persons.

This view has been advocated by Derek Parfit.⁵ When a person divides, as a

result of a split brain transplant, he ‘survives’ in part, Parfit holds, as each of two persons. They constitute his later ‘selves’, neither of whom, to speak strictly, are identical with the original person.

This theory, which Parfit calls the complex view,⁶ does, however, run up against a fundamental difficulty that it commits him to substantial empirical claims which to all appearance could very easily be false. I can bring this out by adopting Bernard Williams’s famous mad surgeon story.⁷ Suppose that a mad surgeon captures you and announces that he is going to transplant your left cerebral hemisphere into one body, and your right one into another. He is going to torture one of the resulting persons and free the other with a gift of a million pounds. You can choose which person is going to be tortured and which to be rewarded, and the surgeon promises to do as you choose. You believe his promise. But how are you to choose? You wish to choose that you are rewarded, but you do not know which resultant person will be you. Now on the complex theory each person will be you to the extent to which he has your brain and resembles you in his apparent memories and character. It would be in principle empirically ascertainable whether and to what extent persons with right hemisphere transplants resemble their originals in apparent memories and character more or less than persons with left hemisphere transplants. But clearly the difference is not going to be great. So Parfit must say that your choice does not greatly matter. Both subsequent persons will be in part you – although perhaps to slightly different degrees. And so you will – although perhaps to slightly different degrees – in part suffer and in part enjoy what each suffers and enjoys. So you have reason both for joyous expectation and for terrified anticipation. But one problem is: how could you have reason for part joyous expectation and part terrified anticipation, when no one future person is going to suffer a mixed fate?

But even if this notion of partial survival does make sense, the more serious difficulty remains, which is this. We can make sense of the supposition that the victim makes the wrong choice, and has the experience of being tortured and not the experience of being rewarded; or the right choice, and has the experience of being rewarded and not the experience of being tortured. A mere philosophical analysis of the concept of personal identity cannot tell you which experiences will be yours tomorrow. To use Bernard Williams’s telling word, any choice would be a ‘risk’. But on Parfit’s view no risk would be involved – for knowing the extent of continuity of brain, apparent memory, and character, you would know the extent to which a future person would be you and so the extent to which his experiences would be yours. Although it *may* be the case that if my cerebral hemispheres are transplanted into different bodies, I survive partly as the person whose body is controlled by one and partly as the person whose body is controlled by the other, it may not be like that at all. Maybe I go where the left hemisphere goes; and when my right hemisphere is separated from the left hemisphere and comes to control a body by itself, either a new person is formed, or the resulting organism, although behaving to some extent like a person, is really a very complicated non-conscious machine. As we have noted, the fate of some parts of my body, such as my arms and legs, is quite

irrelevant to the fate of me. And plausibly the fate of some parts of my brain is irrelevant – can I not survive completely a minor brain operation which removes a very small tumour? But then maybe it is the same with some larger parts of the brain too. We just don't know. If the mad surgeon's victim took the attitude that it didn't matter which way he chose, we must, I suggest, regard him as taking an unjustifiably dogmatic attitude.

The alternative way out of the duplication problem is to say that although apparent memory and brain continuity are, as they obviously are, evidence of personal identity, they are fallible evidence and personal identity is something distinct from them. Just as the presence of blood stains and fingerprints matching those of a given man are evidence of his earlier presence at the scene of the crime, and the discovery of Roman-looking coins and buildings is evidence that the Romans lived in some region, so the similarity of P_2 's apparent memory to that of P_1 and his having much the same brain matter, is evidence that P_2 is the same person as P_1 . Yet blood stains and fingerprints are one thing and a man's earlier presence at the scene of the crime another. His presence at the scene of the crime is not analysable in terms of the later presence of blood stains and fingerprints. The latter is evidence of the former, because you seldom get blood stains and fingerprints at a place, matching those of a given man, unless he has been there leaving them around. But it might happen. So, the suggestion is, personal identity is distinct from, although evidenced by, similarity of memory and continuity of brain.

This account, which for the moment I will follow Parfit in calling the simple view, can meet all the difficulties which have beset the other theories which we have discussed. The difficulty for the complex view was that it seemed very peculiar to suppose that mere logic could determine which of the experiences had by various persons, each of which was to some extent continuous with me in apparent memory and brain matter, would be mine. There seemed to be a further truth – that I would or would not have those experiences – beyond any truths about the extent of similarity in apparent memory and matter of future persons to myself. The simple view claims explicitly that personal identity is one thing, and the extent of similarity in matter and apparent memory another. There is no contradiction in supposing that the one should occur without the other. Strong similarity of matter and apparent memory is powerful evidence of personal identity. I and the person who had my body and brain last week have virtually the same brain matter and such similar apparent memory, that it is well-nigh certain that we are the same person. But where the brain matter is only in part the same and the memory connection less strong, it is only fairly probable that the persons are the same. Where there are two later persons P_2 and P_2^* , each of whom had some continuity with the earlier person P_1 , the evidence supports to some extent each of the two hypotheses – that P_2 is the same person as P_1 , and that P_2^* is the same person as P_1 . It may give more support to one hypothesis than to the other, but the less well supported hypothesis might be the true one, or maybe neither hypothesis is true. Perhaps P_1 has ceased to exist, and two different persons have come into existence. So the simple view fully accepts that mere logic cannot determine which experiences

will be mine, but it allows that continuity of apparent memory and brain provides fallible evidence about this. And of course the duplication objection that they allow for the two subsequent persons being the same person, which we brought against the brain and the memory theories, has no force against the simple theory. For although there can be equally good evidence that each of two later persons is the same person as an earlier person, that evidence is fallible; and since clearly only one person at one time can be strictly the same person as some person at an earlier time, it follows that in one case the evidence is misleading – although we may not know in which case.

. . . In the next section I will expound and develop the simple view, and show that it amounts to the same as Cartesian dualism – the view that a person consists of two parts, soul, and body. . . .

2 The Dualist Theory

The brain transplant considerations of the first section leading to the simple view of personal identity showed that significant continuity of brain and memory was not enough to ensure personal identity. They did not show that continuity of brain or memory were totally dispensable; that P_2 at time t_2 could be the same person as P_1 at an earlier time t_1 , even though P_2 had none of the brain matter (or other bodily matter) of P_1 and had no apparent memory of P_1 's actions and experiences. A number of more extravagant thought-experiments do, however, show that there is no contradiction in this latter supposition.

There seems no contradiction in the supposition that a person might acquire a totally new body (including a completely new brain) – as many religious accounts of life after death claim that men do. To say that this body, sitting at the desk in my room is my body is to say two things. First it is to say that I can move parts of this body (arms, legs, etc.), just like that, without having to do any other intentional action and that I can make a difference to other physical objects only by moving parts of this body. By holding the door handle and turning my hand, I open the door. By bending my leg and stretching it I kick the ball and make it move into the goal. But I do not turn my hand or bend my leg by doing some other intentional action; I just do these things. Secondly, it is to say that my knowledge of states of the world outside this body is derived from their effects on this body – I learn about the positions of physical objects by seeing them, and seeing them involves light rays reflected by them impinging on my eyes and setting up nervous impulses in my optic nerve. My body is the vehicle of my agency in the world and my knowledge of the world. But then is it not coherent to suppose that I might suddenly find that my present body no longer served this function, that I could no longer acquire information through these eyes or move these limbs, but might discover that another body served the same function? I might find myself moving other limbs and acquiring information through other eyes. Then I would have a totally new body. If that body, like my last body, was an occupant of the Earth, then we would have a case of reincarnation, as Eastern religions have understood that. If that body was an occupant

of some distant planet, or an environment which did not belong to the same space as our world, then we would have a case of resurrection as, on the whole, Western religions (Christianity, Judaism and Islam) have understood that. . . .

Equally coherent, I suggest, is the supposition that a person might become disembodied. A person has a body if there is one particular chunk of matter through which he has to operate on and learn about the world. But suppose that he finds himself able to operate on and learn about the world within some small finite region, without having to use one particular chunk of matter for this purpose. He might find himself with knowledge of the position of objects in a room (perhaps by having visual sensations, perhaps not), and able to move such objects just like that, in the ways in which we know about the positions of our limbs and can move them. But the room would not be, as it were, the person's body; for we may suppose that simply by choosing to do so he can gradually shift the focus of his knowledge and control, e.g., to the next room. The person would be in no way limited to operating and learning through one particular chunk of matter. Hence we may term him disembodied. The supposition that a person might become disembodied also seems coherent.

I have been arguing so far that it is coherent to suppose that a person could continue to exist with an entirely new body or with no body at all. . . . Could a person continue to exist without any apparent memory of his previous doings? Quite clearly, we do allow not merely the logical possibility, but the frequent actuality of amnesia – a person forgetting all or certain stretches of his past life. Despite Locke, many a person does forget much of what he has done. But, of course, we normally only suppose this to happen in cases where there is the normal bodily and brain continuity. Our grounds for supposing that a person forgets what he has done are that the evidence of bodily and brain continuity suggests that he was the previous person who did certain things which he now cannot remember having done. And in the absence of both of the main kinds of evidence for personal identity, we would not be justified in supposing that personal identity held. . . . For that reason I cannot describe a case where we would have good reason to suppose that P_2 was identical with P_1 even though there was neither brain continuity nor memory continuity between them. However, only given verificationist dogma is there any reason to suppose that the only things which are true are those of whose truth we can have evidence, and I shall suggest in section 3 [not included here] that there is no good reason for believing verificationism to be true. We can make sense of states of affairs being true, of which we can have no evidence that they are true. And among them surely is the supposition that the person who acquires another body loses not merely control of the old one, but memories of what he did with its aid. . . .

Those who hope to survive their death, despite the destruction of their body, will not necessarily be disturbed if they come to believe that they will then have no memory of their past life on Earth; they may just want to survive and have no interest in continuing to recall life on Earth. Again, apparently, there seems to be no contradiction involved in their belief. It seems to be a coherent belief (whether or not true or justified). Admittedly, there may be stories or beliefs which involve a hidden contradiction when initially they do not seem to do so.

But the fact that there seems (and to so many people) to be no contradiction hidden in these stories is good reason for supposing that there is no contradiction hidden in them – until a contradiction is revealed. If this were not a good reason for believing there to be no contradiction, we would have no good reason for believing any sentence at all to be free of hidden contradiction. . . .

In section 1, I set out Aristotle's account of the identity of substances: that a substance at one time is the same substance as a substance at an earlier time if and only if the later substance has the same form as, and continuity of matter with, the earlier substance. On this view a person is the same person as an earlier person if he has the same form as the earlier person (i.e., both are persons) and has continuity of matter with him (i.e., has the same body).

Certainly, to be the same person as an earlier person, a later person has to have the same form – i.e., has to be a person. If my arguments for the logical possibility of there being disembodied persons are correct, then the essential characteristics of a person constitute a narrower set than those which Aristotle would have included. My arguments suggest that all that a person needs to be a person are certain mental capacities – for having conscious experiences (i.e., thoughts or sensations) and performing intentional actions. Thought-experiments of the kind described earlier allow that a person might lose his body, but they describe his continuing to have conscious experiences and his performing or being able to perform intentional actions, i.e., to do actions which he means to do, bring about effects for some purpose.

Yet if my arguments are correct, showing that two persons can be the same, even if there is no continuity between their bodily matter, we must say that in the form stated the Aristotelian account of identity applies only to inanimate objects and plants and has no application to personal identity.⁸ We are then faced with a choice either of saying that the criteria of personal identity are different from those for other substances, or of trying to give a more general account than Aristotle's of the identity of substances which would cover both persons and other substances. It is possible to widen the Aristotelian account so that we can do the latter. We have only to say that two substances are the same if and only if they have the same form and there is continuity of the stuff of which they are made, and allow that there may be kinds of stuff other than matter. I will call this account of substance identity the wider Aristotelian account. We may say that there is a stuff of another kind, immaterial stuff, and that persons are made of both normal bodily matter and of this immaterial stuff but that it is the continuity of the latter which provides that continuity of stuff which is necessary for the identity of the person over time.

This is in essence the way of expressing the simple theory which is adopted by those who say that a person living on Earth consists of two parts – a material part, the body; and an immaterial part, the soul. The soul is the essential part of a person, and it is its continuing which constitutes the continuing of the person. While on Earth, the soul is linked to a body (by the body being the vehicle of the person's knowledge of and action upon the physical world). But, it is logically possible, the soul can be separated from the body and exist in a disembodied state (in the way described earlier) or linked to a new body. This way of

expressing things has been used in many religious traditions down the centuries, for it is a very natural way of expressing what is involved in being a person once you allow that a person can survive the death of his body. Classical philosophical statements of it are to be found in Plato and, above all, in Descartes. I shall call this view classical dualism.

I wrote that ‘in essence’ classical dualism is the view that there is more stuff to the person than bodily matter, and that it is the continuing of this stuff which is necessary for the continuing of the person, because a writer such as Descartes did not distinguish between the immaterial stuff, let us call it soul-stuff, and that stuff being organized (with or without a body) as one soul. Descartes and other classical dualists however did not make this distinction, because they assumed (implicitly) that it was not logically possible that persons divide – i.e., that an earlier person could be in part the same person as each of two later persons. Hence they implicitly assumed that soul-stuff comes in essentially indivisible units. That is indeed what one has to say about soul-stuff, if one makes the supposition (as I was inclined to do, in section 1) that it is not logically possible that persons divide. There is nothing odd about supposing that soul-stuff comes in essentially indivisible units. Of any chunk of matter, however small, it is always logically, if not physically, possible that it be divided into two. Yet it is because matter is extended, that one can always make sense of it being divided. For a chunk of matter necessarily takes up a finite volume of space. A finite volume of space necessarily is composed of two half-volumes. So it always makes sense to suppose that part of the chunk which occupies the left half-volume of space to be separated from that part of the chunk which occupies the right half-volume. But that kind of consideration has no application to immaterial stuff. There is no reason why there should not be a kind of immaterial stuff which necessarily is indivisible; and if the supposition of section 1 is correct, the soul-stuff will have that property. . . .

Given that for any present person who is currently conscious, there is no logical impossibility, whatever else may be true now of that person, that that person continue to exist without his body, it follows that that person must now actually have a part other than a bodily part which can continue, and which we may call his soul – and so that his possession of it is entailed by his being a conscious thing. For there is not even a logical possibility, that if I now consist of nothing but matter and the matter is destroyed, I should nevertheless continue to exist. From the mere logical possibility of my continued existence there follows the actual fact that there is now more to me than my body; and that more is the essential part of myself. A person’s being conscious is thus to be analysed as an immaterial core of himself, his soul being conscious.⁹

So Descartes argues, and his argument seems to me correct – given the wider Aristotelian framework. If we are prepared to say that substances can be the same, even though none of the stuff (in a wide sense) of which they are made is the same, the conclusion does not follow. The wider Aristotelian framework provides a partial definition of ‘stuff’ rather than a factual truth.

To say that a person has an immaterial soul is not to say that if you examine him closely enough under an acute enough microscope you will find some very

rarefied constituent which has eluded the power of ordinary microscopes. It is just a way of expressing the point within a traditional framework of thought that persons can – it is logically possible – continue, when their bodies do not. It does, however, seem a very natural way of expressing the point – especially once we allow that persons can become disembodied. Unless we adopt a wider Aristotelian framework, we shall have to say that there can be substances which are not made of anything, and which are the same substances as other substances which are made of matter.

It does not follow from all this that a person's body is no part of him. Given that what we are trying to do is to elucidate the nature of those entities which we normally call 'persons', we must say that arms and legs and all other parts of the living body are parts of the person. My arms and legs are parts of me. . . .

As we have seen, classical dualism is the way of expressing the simple view of personal identity within what I called the wider Aristotelian framework. However, this framework is a wider one than Aristotle himself would have been happy with, allowing a kind of stuff other than Aristotle would have countenanced. There has been in the history of thought a different and very influential way of modifying Aristotle, to take account of the kind of point made by the simple view. This way was due to St Thomas Aquinas (see, e.g., *Summa contra Gentiles*). Aquinas accepted Aristotle's general doctrine that substances are made of matter, organized by a form; the desk is the desk which it is because of the matter of which it is made and the shape which is imposed upon it. The form was normally a system of properties, universals which had no existence except in the particular substances in which they were instantiated. However, Aquinas claimed that for man the form of the body, which he called the soul, was separable from the body and capable of independent existence. The soul of man, unlike the souls of animals or plants, was, in Aquinas's terminology, an 'intellectual substance'.

However, if we are going to modify Aristotle to make his views compatible with the simple theory of personal identity, this seems definitely the less satisfactory way of doing so. Properties seem by their very nature to be universals and so it is hard to give any sense to their existing except when conjoined to some stuff. Above all, it is hard to give sense to their being individual – a universal can be instantiated in many different substances. What makes the substances differ is the different stuff of which they are composed. The form of man can be instantiated in many different men. But Aquinas wants a form which is a particular, and so could only be combined with one body. All of this seems to involve a greater distortion of Aristotle's system than does classical dualism. Aquinas's system does have some advantages over classical dualism – for example, it enables him to bring out the naturalness of a person being embodied and the temporary and transitory character of any disembodiment – but the disadvantages of taking Aristotle's approach and then distorting it to this unrecognizable extent are in my view very great. Hence my preference for what I have called classical dualism. . . .

Notes

- 1 The logically necessary and sufficient conditions for something being so are those conditions such that if they are present, that thing must be so; and if they are absent, that thing cannot be so – all this because of considerations of logic.
- 2 Wiggins is even more tentative in the amended version of the book, *Sameness and Substance* (Oxford, 1980).
- 3 Bernard Williams, ‘Personal Identity and Individuation’, *Proceedings of the Aristotelian Society*, (1956–57), p. 332.
- 4 He suggests analysing ‘person’ in such a way that ‘coincidence under the concept person logically required the continuance in one organized parcel of all that was causally sufficient and causally necessary to the continuance of essential and characteristic functioning, no autonomously sufficient part achieving autonomous and functionally separate existence’ (Wiggins, *Identity and Spatiotemporal Continuity*, p. 55).
- 5 ‘Personal Identity’, *Philosophical Review*, 80 (1971), pp. 3–27.
- 6 He introduces this terminology in his paper, ‘On the Importance of Self-Identity’, *Journal of Philosophy*, 68 (1971), pp. 683–90.
- 7 Bernard Williams, ‘The Self and the Future’, *Philosophical Review*, 79 (1970), pp. 161–80.
- 8 I do not discuss the difficult issue of whether the Aristotelian account applies to animals other than man, e.g., whether continuity of matter and form is necessary and sufficient for the identity of a dog at a later time with a dog at an earlier time.
- 9 It may be useful, in case anyone suspects the argument of this paragraph of committing some modal fallacy, to set it out in a more formal logical shape. I use the usual logical symbols – ‘.’ means ‘and’, ‘~’ means ‘not’, ‘◊’ means ‘it is logically possible’. I then introduce the following definitions:

$p =$ I am a conscious person, and I exist in 1984

$q =$ My body is destroyed at the end of 1984

$r =$ I have a soul in 1984

$s =$ I exist in 1985

x ranges over all consistent propositions compatible with $(p.q)$ and describing 1984 states of affairs

(‘ (x) ’ is to be read in the normal way as ‘for all states $x \dots$ ’)

The argument may now be set out as follows:

$$\begin{array}{l} p \\ (x) \Diamond (p.q.x.s) \\ \sim \Diamond (p.q.\sim r.s) \end{array}$$

Premise (1)

Premise (2)

Premise (3)

$\therefore \sim r$ is not within the range of x .

But since $\sim r$ describes a state of affairs in 1984, it is not compatible with $(p.q)$. But q can hardly make a difference to whether or not r . So p is incompatible with $\sim r$.

$\therefore r$

The argument is designed to show that r follows from p ; and so, more generally, that every conscious person has a soul. Premise (3) is justified by the wider Aristotelian

principle that if I am to continue, some of the stuff out of which I am made has to continue. As I argued in the text, that stuff must be non-bodily stuff. The soul is defined as that non-bodily part whose continuing is essential for my continuing.

Premise (2) relies on the intuition that whatever else might be the case in 1984, compatible with (p.q), my stream of consciousness could continue thereafter.

If you deny (2) and say that r is a state of affairs not entailed by (p.q), but which has to hold if it is to be possible that s, you run into this difficulty. There may be two people in 1984, Oliver, who has a soul, and Fagin, who does not. Both are embodied and conscious, and to all appearances indistinguishable. God (who can do all things logically possible, compatible with how the world is up to now), having forgotten to give Fagin a soul, has, as he annihilates Fagin's body at the end of 1984, no power to continue his stream of thought. Whereas he has the power to continue Oliver's stream of thought. This seems absurd.

38 The Puzzle of Conscious/Experience*

David J. Chalmers

Conscious experience is at once the most familiar thing in the world and the most mysterious. There is nothing we know about more directly than consciousness, but it is extraordinarily hard to reconcile it with everything else we know. Why does it exist? What does it do? How could it possibly arise from neural processes in the brain? These questions are among the most intriguing in all of science.

From an objective viewpoint, the brain is relatively comprehensible. When you look at this page, there is a whir of processing: photons strike your retina, electrical signals are passed up your optic nerve and between different areas of your brain, and eventually you might respond with a smile, a perplexed frown or a remark. But there is also a subjective aspect. When you look at the page, you are conscious of it, directly experiencing the images and words as part of your private, mental life. You have vivid impressions of colored flowers and vibrant sky. At the same time, you may be feeling some emotions and forming some thoughts. Together such experiences make up consciousness: the subjective, inner life of the mind.

For many years consciousness was shunned by researchers studying the brain and the mind. The prevailing view was that science, which depends on objectivity, could not accommodate something as subjective as consciousness. The behaviorist movement in psychology, dominant earlier in this century, concentrated on external behavior and disallowed any talk of internal mental processes.

* From David Chalmers, "The Puzzle of Conscious Experience," *Scientific American*, 273 (1995), pp. 80-6. Reprinted by permission of the author.

Later, the rise of cognitive science focused attention on processes inside the head. Still, consciousness remained off-limits, fit only for late-night discussion over drinks.

Over the past several years, however, an increasing number of neuroscientists, psychologists and philosophers have been rejecting the idea that consciousness cannot be studied and are attempting to delve into its secrets. As might be expected of a field so new, there is a tangle of diverse and conflicting theories, often using basic concepts in incompatible ways. To help unsnarl the tangle, philosophical reasoning is vital.

The myriad views within the field range from reductionist theories, according to which consciousness can be explained by the standard methods of neuroscience and psychology, to the position of the so-called mysterians, who say we will never understand consciousness at all. I believe that on close analysis both of these views can be seen to be mistaken and that the truth lies somewhere in the middle.

Against reductionism I will argue that the tools of neuroscience cannot provide a full account of conscious experience, although they have much to offer. Against mysterianism I will hold that consciousness might be explained by a new kind of theory. The full details of such a theory are still out of reach, but careful reasoning and some educated inferences can reveal something of its general nature. For example, it will probably involve new fundamental laws, and the concept of information may play a central role. These faint glimmerings suggest that a theory of consciousness may have startling consequences for our view of the universe and of ourselves.

The Hard Problem

Researchers use the word “consciousness” in many different ways. To clarify the issues, we first have to separate the problems that are often clustered together under the name. For this purpose, I find it useful to distinguish between the “easy problems” and the “hard problem” of consciousness. The easy problems are by no means trivial – they are actually as challenging as most in psychology and biology – but it is with the hard problem that the central mystery lies.

The easy problems of consciousness include the following: How can a human subject discriminate sensory stimuli and react to them appropriately? How does the brain integrate information from many different sources and use this information to control behavior? How is it that subjects can verbalize their internal states? Although all these questions are associated with consciousness, they all concern the objective mechanisms of the cognitive system. Consequently, we have every reason to expect that continued work in cognitive psychology and neuroscience will answer them.

The hard problem, in contrast, is the question of how physical processes in the brain give rise to subjective experience. This puzzle involves the inner aspect of thought and perception: the way things feel for the subject. When we

see, for example, we experience visual sensations, such as that of vivid blue. Or think of the ineffable sound of a distant oboe, the agony of an intense pain, the sparkle of happiness or the meditative quality of a moment lost in thought. All are part of what I am calling consciousness. It is these phenomena that pose the real mystery of the mind.

To illustrate the distinction, consider a thought experiment devised by the Australian philosopher Frank Jackson. Suppose that Mary, a neuroscientist in the twenty-third century, is the world's leading expert on the brain processes responsible for color vision. But Mary has lived her whole life in a black-and-white room and has never seen any other colors. She knows everything there is to know about physical processes in the brain – its biology, structure and function. This understanding enables her to grasp everything there is to know about the easy problems: how the brain discriminates stimuli, integrates information and produces verbal reports. From her knowledge of color vision, she knows the way color names correspond with wavelengths on the light spectrum. But there is still something crucial about color vision that Mary does not know: what it is like to experience a color such as red. It follows that there are facts about conscious experience that cannot be deduced from physical facts about the functioning of the brain.

Indeed, nobody knows why these physical processes are accompanied by conscious experience at all. Why is it that when our brains process light of a certain wavelength, we have an experience of deep purple? Why do we have any experience at all? Could not an unconscious automaton have performed the same tasks just as well? These are questions that we would like a theory of consciousness to answer.

I am not denying that consciousness arises from the brain. We know, for example, that the subjective experience of vision is closely linked to processes in the visual cortex. It is the link itself that perplexes, however. Remarkably, subjective experience seems to emerge from a physical process. But we have no idea how or why this is.

Is Neuroscience Enough?

Given the flurry of recent work on consciousness in neuroscience and psychology, one might think this mystery is starting to be cleared up. On closer examination, however, it turns out that almost all the current work addresses only the easy problems of consciousness. The confidence of the reductionist view comes from the progress on the easy problems, but none of this makes any difference where the hard problem is concerned.

Consider the hypothesis put forward by neurobiologists Francis Crick of the Salk Institute for Biological Studies in San Diego and Christof Koch of the California Institute of Technology. They suggest that consciousness may arise from certain oscillations in the cerebral cortex, which become synchronized as neurons fire 40 times per second. Crick and Koch believe the phenomenon might explain how different attributes of a single perceived object (its color and

shape, for example), which are processed in different parts of the brain, are merged into a coherent whole. In this theory, two pieces of information become bound together precisely when they are represented by synchronized neural firings.

The hypothesis could conceivably elucidate one of the easy problems about how information is integrated in the brain. But why should synchronized oscillations give rise to a visual experience, no matter how much integration is taking place? This question involves the hard problem, about which the theory has nothing to offer. Indeed, Crick and Koch are agnostic about whether the hard problem can be solved by science at all.

The same kind of critique could be applied to almost all the recent work on consciousness. In his 1991 book *Consciousness Explained*, philosopher Daniel C. Dennett laid out a sophisticated theory of how numerous independent processes in the brain combine to produce a coherent response to a perceived event. The theory might do much to explain how we produce verbal reports on our internal states, but it tells us very little about why there should be a subjective experience behind these reports. Like other reductionist theories, Dennett's is a theory of the easy problems.

The critical common trait among these easy problems is that they all concern how a cognitive or behavioral function is performed. All are ultimately questions about how the brain carries out some task – how it discriminates stimuli, integrates information, produces reports and so on. Once neurobiology specifies appropriate neural mechanisms, showing how the functions are performed, the easy problems are solved. The hard problem of consciousness, in contrast, goes beyond problems about how functions are performed. Even if every behavioral and cognitive function related to consciousness were explained, there would still remain a further mystery: Why is the performance of these functions accompanied by conscious experience? It is this additional conundrum that makes the hard problem hard.

The Explanatory Gap

Some have suggested that to solve the hard problem, we need to bring in new tools of physical explanation: nonlinear dynamics, say, or new discoveries in neuroscience, or quantum mechanics. But these ideas suffer from exactly the same difficulty. Consider a proposal from Stuart R. Hameroff of the University of Arizona and Roger Penrose of the University of Oxford. They hold that consciousness arises from quantum-physical processes taking place in microtubules, which are protein structures inside neurons. It is possible (if not likely) that such a hypothesis will lead to an explanation of how the brain makes decisions or even how it proves mathematical theorems, as Hameroff and Penrose suggest. But even if it does, the theory is silent about how these processes might give rise to conscious experience. Indeed, the same problem arises with any theory of consciousness based only on physical processing.

The trouble is that physical theories are best suited to explaining why systems

have a certain physical structure and how they perform various functions. Most problems in science have this form; to explain life, for example, we need to describe how a physical system can reproduce, adapt and metabolize. But consciousness is a different sort of problem entirely, as it goes beyond the explanation of structure and function.

Of course, neuroscience is not irrelevant to the study of consciousness. For one thing, it may be able to reveal the nature of the neural correlate of consciousness – the brain processes most directly associated with conscious experience. It may even give a detailed correspondence between specific processes in the brain and related components of experience. But until we know why these processes give rise to conscious experience at all, we will not have crossed what philosopher Joseph Levine has called the explanatory gap between physical processes and consciousness. Making that leap will demand a new kind of theory.

A True Theory of Everything

In searching for an alternative, a key observation is that not all entities in science are explained in terms of more basic entities. In physics, for example, space-time, mass and charge (among other things) are regarded as fundamental features of the world, as they are not reducible to anything simpler. Despite this irreducibility, detailed and useful theories relate these entities to one another in terms of fundamental laws. Together these features and laws explain a great variety of complex and subtle phenomena.

It is widely believed that physics provides a complete catalogue of the universe's fundamental features and laws. As physicist Steven Weinberg puts it in his 1992 book *Dreams of a Final Theory*, the goal of physics is a “theory of everything” from which all there is to know about the universe can be derived. But Weinberg concedes that there is a problem with consciousness. Despite the power of physical theory, the existence of consciousness does not seem to be derivable from physical laws. He defends physics by arguing that it might eventually explain what he calls the objective correlates of consciousness (that is, the neural correlates), but of course to do this is not to explain consciousness itself. If the existence of consciousness cannot be derived from physical laws, a theory of physics is not a true theory of everything. So a final theory must contain an additional fundamental component.

Toward this end, I propose that conscious experience be considered a fundamental feature, irreducible to anything more basic. The idea may seem strange at first, but consistency seems to demand it. In the nineteenth century it turned out that electromagnetic phenomena could not be explained in terms of previously known principles. As a consequence, scientists introduced electromagnetic charge as a new fundamental entity and studied the associated fundamental laws. Similar reasoning should apply to consciousness. If existing fundamental theories cannot encompass it, then something new is required.

Where there is a fundamental property, there are fundamental laws. In this case, the laws must relate experience to elements of physical theory. These laws

will almost certainly not interfere with those of the physical world; it seems that the latter form a closed system in their own right. Rather the laws will serve as a bridge, specifying how experience depends on underlying physical processes. It is this bridge that will cross the explanatory gap.

Thus, a complete theory will have two components: physical laws, telling us about the behavior of physical systems from the infinitesimal to the cosmological, and what we might call psychophysical laws, telling us how some of those systems are associated with conscious experience. These two components will constitute a true theory of everything.

Searching for a Theory

Supposing for the moment that they exist, how might we uncover such psychophysical laws? The great hindrance in this pursuit will be a lack of data. As I have described it, consciousness is subjective, so there is no direct way to monitor it in others. But this difficulty is an obstacle, not a dead end. For a start, each one of us has access to our own experiences, a rich trove that can be used to formulate theories. We can also plausibly rely on indirect information, such as subjects' descriptions of their experiences. Philosophical arguments and thought experiments also have a role to play. Such methods have limitations, but they give us more than enough to get started.

These theories will not be conclusively testable, so they will inevitably be more speculative than those of more conventional scientific disciplines. Nevertheless, there is no reason why they should not be strongly constrained to account accurately for our own first-person experiences, as well as the evidence from subjects' reports. If we find a theory that fits the data better than any other theory of equal simplicity, we will have good reason to accept it. Right now we do not have even a single theory that fits the data, so worries about testability are premature.

We might start by looking for high-level bridging laws, connecting physical processes to experience at an everyday level. The basic contour of such a law might be gleaned from the observation that when we are conscious of something, we are generally able to act on it and speak about it – which are objective, physical functions. Conversely, when some information is directly available for action and speech, it is generally conscious. Thus, consciousness correlates well with what we might call “awareness”: the process by which information in the brain is made globally available to motor processes such as speech and bodily action.

The notion may seem trivial. But as defined here, awareness is objective and physical, whereas consciousness is not. Some refinements to the definition of awareness are needed, in order to extend the concept to animals and infants, which cannot speak. But at least in familiar cases, it is possible to see the rough outlines of a psychophysical law: where there is awareness, there is consciousness, and vice versa.

To take this line of reasoning a step further, consider the structure present in

the conscious experience. The experience of a field of vision, for example, is a constantly changing mosaic of colors, shapes and patterns and as such has a detailed geometric structure. The fact that we can describe this structure, reach out in the direction of many of its components and perform other actions that depend on it suggests that the structure corresponds directly to that of the information made available in the brain through the neural processes of awareness.

Similarly, our experiences of color have an intrinsic three-dimensional structure that is mirrored in the structure of information processes in the brain's visual cortex. This structure is illustrated in the color wheels and charts used by artists. Colors are arranged in a systematic pattern – red to green on one axis, blue to yellow on another, and black to white on a third. Colors that are close to one another on a color wheel are experienced as similar. It is extremely likely that they also correspond to similar perceptual representations in the brain, as part of a system of complex three-dimensional coding among neurons that is not yet fully understood. We can recast the underlying concept as a principle of structural coherence: the structure of conscious experience is mirrored by the structure of information in awareness, and vice versa.

Another candidate for a psychophysical law is a principle of organizational invariance. It holds that physical systems with the same abstract organization will give rise to the same kind of conscious experience, no matter what they are made of. For example, if the precise interactions between our neurons could be duplicated with silicon chips, the same conscious experience would arise. The idea is somewhat controversial, but I believe it is strongly supported by thought experiments describing the gradual replacement of neurons by silicon. The remarkable implication is that consciousness might someday be achieved in machines.

Information: Physical and Experiential

The ultimate goal of a theory of consciousness is a simple and elegant set of fundamental laws, analogous to the fundamental laws of physics. The principles described above are unlikely to be fundamental, however. Rather they seem to be high-level psychophysical laws, analogous to macroscopic principles in physics such as those of thermodynamics or kinematics. What might the underlying fundamental laws be? No one knows, but I don't mind speculating.

I suggest that the primary psychophysical laws may centrally involve the concept of information. The abstract notion of information, as put forward in the 1940s by Claude E. Shannon of the Massachusetts Institute of Technology, is that of a set of separate states with a basic structure of similarities and differences between them. We can think of a 10-bit binary code as an information state, for example. Such information states can be embodied in the physical world. This happens whenever they correspond to physical states (voltages, say); the differences between them can be transmitted along some pathway, such as a telephone line.

We can also find information embodied in conscious experience. The pattern of color patches in a visual field, for example, can be seen as analogous to that of the pixels covering a display screen. Intriguingly, it turns out that we find the same information states embedded in conscious experience and in underlying physical processes in the brain. The three-dimensional encoding of color spaces, for example, suggests that the information state in a color experience corresponds directly to an information state in the brain. We might even regard the two states as distinct aspects of a single information state, which is simultaneously embodied in both physical processing and conscious experience.

A natural hypothesis ensues. Perhaps information, or at least some information, has two basic aspects: a physical one and an experiential one. This hypothesis has the status of a fundamental principle that might underlie the relation between physical processes and experience. Wherever we find conscious experience, it exists as one aspect of an information state, the other aspect of which is embedded in a physical process in the brain. This proposal needs to be fleshed out to make a satisfying theory. But it fits nicely with the principles mentioned earlier – systems with the same organization will embody the same information, for example – and it could explain numerous features of our conscious experience.

The idea is at least compatible with several others, such as physicist John A. Wheeler's suggestion that information is fundamental to the physics of the universe. The laws of physics might ultimately be cast in informational terms, in which case we would have a satisfying congruence between the constructs in both physical and psychophysical laws. It may even be that a theory of physics and a theory of consciousness could eventually be consolidated into a single grander theory of information.

A potential problem is posed by the ubiquity of information. Even a thermostat embodies some information, for example, but is it conscious? There are at least two possible responses. First, we could constrain the fundamental laws so that only some information has an experiential aspect, perhaps depending on how it is physically processed. Secondly, we might bite the bullet and allow that all information has an experiential aspect – where there is complex information processing, there is complex experience, and where there is simple information processing, there is simple experience. If this is so, then even a thermostat might have experiences, although they would be much simpler than even a basic color experience, and there would certainly be no accompanying emotions or thoughts. This seems odd at first, but if experience is truly fundamental, we might expect it to be widespread. In any case, the choice between these alternatives should depend on which can be integrated into the most powerful theory.

Of course, such ideas may be all wrong. On the other hand, they might evolve into a more powerful proposal that predicts the precise structure of our conscious experience from physical processes in our brains. If this project succeeds, we will have good reason to accept the theory. If it fails, other avenues will be pursued, and alternative fundamental theories may be developed. In this way, we may one day resolve the greatest mystery of the mind.

Appendix: Dancing Qualia in a Synthetic Brain

Whether consciousness could arise in a complex, synthetic system is a question many people find intrinsically fascinating. Although it may be decades or even centuries before such a system is built, a simple thought experiment offers strong evidence that an artificial brain, if organized appropriately, would indeed have precisely the same kind of conscious experiences as a human being.

Consider a silicon-based system in which the chips are organized and function in the same way as the neurons in your brain. That is, each chip in the silicon system does exactly what its natural analogue does and is interconnected to surrounding elements in precisely the same way. Thus, the behavior exhibited by the artificial system will be exactly the same as yours. The crucial question is: Will it be conscious in the same way that you are?

Let us assume, for the purpose of argument, that it would not be. (Here we use a reasoning technique known as *reductio ad absurdum*, in which the opposite hypothesis is assumed and then shown to lead to an untenable conclusion.) That is, it either has different experiences – an experience of blue, say, when you are seeing red – or no experience at all. We will consider the first case; the reasoning proceeds similarly in both cases.

Because chips and neurons have the same function, they are interchangeable, with the proper interfacing. Chips therefore can replace neurons, producing a continuum of cases in which a successively larger proportion of neurons are replaced by chips. Along this continuum, the conscious experience of the system will also change. For example, we might replace all the neurons in your visual cortex with an identically organized version made of silicon. The resulting brain, with an artificial visual cortex, will have a different conscious experience from the original: where you had previously seen red, you may now experience purple (or perhaps a faded pink, in the case where the wholly silicon system has no experience at all).

Both visual cortices are then attached to your brain, through a two-position switch. With the switch in one mode, you use the natural visual cortex; in the other, the artificial cortex is activated. When the switch is flipped, your experience changes from red to purple, or vice versa. When the switch is flipped repeatedly, your experiences “dance” between the two different conscious states (red and purple), known as qualia.

Because your brain’s organization has not changed, however, there can be no behavioral change when the switch is thrown. Therefore, when asked about what you are seeing, you will say that nothing has changed. You will hold that you are seeing red and have seen nothing but red – even though the two colors are dancing before your eyes. This conclusion is so unreasonable that it is best taken as a *reductio ad absurdum* of the original assumption – that an artificial system with identical organization and functioning has a different conscious experience from that of a neural brain. Retraction of the assumption establishes the opposite: that systems with the same organization have the same conscious experience.

Is it Possible for Us to Act Freely?

39 Free Will as Involving Determination and Inconceivable without It*

R. E. Hobart

The thesis of this article is that there has never been any ground for the controversy between the doctrine of free will and determinism, that it is based upon a misapprehension, that the two assertions are entirely consistent, that one of them strictly implies the other, that they have been opposed only because of our natural want of the analytical imagination. In so saying I do not tamper with the meaning of either phrase. That would be unpardonable. I mean free will in the natural and usual sense, in the fullest, the most absolute sense in which for the purposes of the personal and moral life the term is ever employed. I mean it as implying responsibility, merit and demerit, guilt and desert. I mean it as implying, after an act has been performed, that one 'could have done otherwise' than one did. I mean it as conveying these things also, not in any subtly modified sense but in exactly the sense in which we conceive them in life and in law and in ethics. These two doctrines have been opposed because we have not realised that free will can be analysed without being destroyed, and that determinism is merely a feature of the analysis of it. And if we are tempted to take refuge in the thought of an 'ultimate', an 'innermost' liberty that eludes the analysis, then we have implied a deterministic basis and constitution for this liberty as well. For such a basis and constitution lie in the idea of liberty. . . .

I am not maintaining that determinism is true; only that it is true in so far as we have free will. That we are free in willing is, broadly speaking, a fact of experience. That broad fact is more assured than any philosophical analysis. It is therefore surer than the deterministic analysis of it, entirely adequate as that in the end appears to be. But it is not here affirmed that there are no small exceptions, no slight undetermined swervings, no ingredient of absolute chance. All that is here said is that such absence of determination, if and so far as it exists, is no gain to freedom, but sheer loss of it; no advantage to the moral life, but blank subtraction from it. – When I speak below of 'the indeterminist' I mean the libertarian indeterminist, that is, him who believes in free will and holds that it involves indetermination.

By the analytical imagination is meant, of course, the power we have, not by

* From R. E. Hobart, 'Free Will as Involving Determination and Inconceivable without It,' *Mind*, 63 (1934), pp. 1–27. Reprinted by permission of Oxford University Press and *Mind* Association.

nature but by training, of realising that the component parts of a thing or process, taken together, each in its place, with their relations, are identical with the thing or process itself. If it is ‘more than its parts’, then this ‘more’ will appear in the analysis. It is not true, of course, that all facts are susceptible of analysis, but so far as they are, there is occasion for the analytical imagination. We have been accustomed to think of a thing or a person as a whole, not as a combination of parts. We have been accustomed to think of its activities as the way in which, as a whole, it naturally and obviously behaves. It is a new, an unfamiliar and an awkward act on the mind’s part to consider it, not as one thing acting in its natural manner, but as a system of parts that work together in a complicated process. Analysis often seems at first to have taken away the individuality of the thing, its unity, the impression of the familiar identity. For a simple mind this is strikingly true of the analysis of a complicated machine. The reader may recall Paulsen’s ever significant story about the introduction of a railway into Germany. When it reached the village of a certain enlightened pastor, he took his people to where a locomotive engine was standing, and in the clearest words explained of what parts it consisted and how it worked. He was much pleased by their eager nods of intelligence as he proceeded. But on his finishing they said: ‘Yes, yes, Herr Pastor, but there’s a horse inside, isn’t there?’ They could not *realise* the analysis. They were wanting in the analytical imagination. Why not? They had never been trained to it. It is in the first instance a great effort to think of all the parts working together to produce the simple result that the engine glides down the track. It is easy to think of a horse inside doing all the work. A horse is a familiar totality that does familiar things. They could no better have grasped the physiological analysis of a horse’s movements had it been set forth to them.

The reason for thinking that there is no occasion for the controversy lies exclusively in the analysis of the terms employed in it. But the several analyses must all be taken together, realised jointly, before the position can be fully understood.

Self and Character

We are not concerned with the total nature of the self, but only with the aspect of it strictly involved in our question . . . It is the concrete, active self, existing through time and differing from others. The whole stress of morality arises because moral selves are not alike, because there is need of influencing some moral selves to make them refrain from certain acts or neglects, that is, in order to make them better moral selves. How do we express the difference? We call it a difference of moral qualities, traits, or character. We are having regard to the question what acts will come from these selves. By character we mean, do we not? the sum of a man’s tendencies to action, considered in their relative strength; or that sum in so far as it bears upon morals.

Now the position of the indeterminist is that a free act of will is the act of the self. The self becomes through it the author of the physical act that ensues. This

volition of the self causes the physical act but it is not in its turn caused, it is 'spontaneous'. To regard it as caused would be determinism. The causing self to which the indeterminist here refers is to be conceived as distinct from character; distinct from temperament, wishes, habits, impulses. He emphasises two things equally: the physical act springs from the self through its volition, and it does not spring merely from character, it is not simply the result of character and circumstances. If we ask, 'Was there anything that induced the self thus to act?' we are answered in effect, 'Not definitively. The self feels motives but its act is not determined by them. It can choose between them.'

The next thing to notice is that this position of the indeterminist is taken in defence of moral conceptions. There would be no fitness, he says, in our reproaching ourselves, in our feeling remorse, in our holding ourselves or anyone guilty, if the act in question were not the act of the self instead of a product of the machinery of motives.

We have here one of the most remarkable and instructive examples of something in which the history of philosophy abounds – of a persistent, an age-long deadlock due solely to the indisposition of the human mind to look closely into the meaning of its terms.

How do we reproach ourselves? We say to ourselves, 'How negligent of me!' 'How thoughtless!' 'How selfish!' 'How hasty and unrestrained!' 'That I should have been capable even for a moment of taking such a petty, irritated view!' etc. In other words, we are attributing to ourselves at the time of the act, in some respect and measure, a bad character, and regretting it. And that is the entire point of our self-reproach. . . . All the most intimate terms of the moral life imply that the act has proceeded from me, the distinctive me, from the manner of man I am or was. And this is the very thing on which the libertarian lays stress. What the indeterminist prizes with all his heart, what he stoutly affirms and insists upon, is precisely what he denies, namely, that I, the concrete and specific moral being, am the author, the source of my acts. For, of course, that is determinism. To say that they come from the self is to say that they are determined by the self – the moral self, the self with a moral quality. He gives our preferring the bad name of the machinery of motives, but they are just what we feel in ourselves when we decide. When he maintains that the self at the moment of decision may act to some extent independently of motives, *and is good or bad according as it acts in this direction or that*, he is simply setting up one character within another, he is separating the self from what he understands by the person's character as at first mentioned, only thereupon to attribute to it a character of its own, *in that he judges it good or bad*. . . .

If in conceiving the self you detach it from all motives or tendencies, what you have is not a morally admirable or condemnable, not a morally characterisable self at all. Hence it is not subject to reproach. You cannot call a self good because of its courageous free action, and then deny that its action was determined by its character. In calling it good because of that action you have implied that the action came from its goodness (which means its good character) and was a sign thereof. By their fruits ye shall know them. The indeterminist appears to imagine that he can distinguish the moral 'I' from all its propensities, regard

its act as arising in the moment undetermined by them, and yet can then (for the first time, in his opinion, with propriety!) ascribe to this 'I' an admirable quality. At the very root of his doctrine he contradicts himself. . . .

We are told, however, that it is under determinism that we should have no right any more to praise or to blame. At least we could not do so in the old sense of the terms. We might throw words of praise to a man, or throw words of blame at him, because we know from observation that they will affect his action; but the old light of meaning in the terms has gone out. Well, all we have to do is to keep asking what this old meaning was. We praise a man by saying that he is a good friend, or a hard worker, or a competent man of business, or a trusty assistant, or a judicious minister, or a gifted poet, or one of the noblest of men – one of the noblest of characters! In other words, he is a being with such and such qualities. If it is moral praise, he is a being with such and such tendencies to bring forth good acts. If we describe a single act, saying, for instance: 'Well done!' we mean to praise the person for the act as being the author of it. It is he who has done well and proved himself capable of doing so. If the happy act is accidental we say that no praise is deserved for it. If a person is gratified by praise it is because of the estimate of him, in some respect or in general, that is conveyed. Praise . . . means description, with expressed or implied admiration. If any instance of it can be found which does not consist in these elements our analysis fails. 'Praise the Lord, O my soul, *and forget not all His benefits*,' – and the Psalm goes on to tell His loving and guarding acts toward human-kind. To praise the Lord is to tell His perfections, especially the perfections of His character. This is the old light that has always been in words of praise and there appears no reason for its going out.

Indeterminism maintains that we need not be impelled to action by our wishes, that our active will need not be determined by them. Motives 'incline without necessitating'. We choose amongst the ideas of action before us, but need not choose solely according to the attraction of desire, in however wide a sense that word is used. Our inmost self may rise up in its autonomy and moral dignity, independently of motives, and register its sovereign decree.

Now, *in so far* as this 'interposition of the self' is undetermined, the act is not *its* act, it does not issue from any concrete continuing self; it is born at the moment, of nothing, hence it expresses no quality; it bursts into being from no source. The self does not register *its* decree, for the decree is not the product of just that '*it*'. The self does not rise up in *its* moral dignity, for dignity is the quality of an enduring being, influencing its actions, and therefore expressed by them, and that would be determination. *In proportion* as an act of volition starts of itself without cause it is exactly, so far as the freedom of the individual is concerned, as if it had been thrown into his mind from without – 'suggested' to him – by a freakish demon. It is exactly like it in this respect, that in neither case does the volition arise from what the man is, cares for or feels allegiance to; it does not come out of him. *In proportion* as it is undetermined, it is just as if his legs should suddenly spring up and carry him off where he did not prefer to go. Far from constituting freedom, that would mean, in the exact measure in which it took place, the loss of freedom. It would be an interference, and an utterly

uncontrollable interference, with his power of acting as he prefers. In fine, then, *just so far* as the volition is undetermined, the self can neither be praised nor blamed for it, since it is not the act of the self.

The principle of free will says: 'I produce my volitions.' Determinism says: 'My volitions are produced by *me*.' Determinism is free will expressed in the passive voice.

After all, it is plain what the indeterminists have done. It has not occurred to them that our free will may be resolved into its component elements. (Thus far a portion only of this resolution has been considered.) When it is thus resolved they do not recognise it. The analytical imagination is considerably taxed to perceive the identity of the free power that we feel with the component parts that analysis shows us. We are gratified by their nods of intelligence and their bright, eager faces as the analysis proceeds, but at the close are a little disheartened to find them falling back on the innocent supposition of a horse inside that does all the essential work. They forget that they may be called upon to analyse the horse. They solve the problem by forgetting analysis. The solution they offer is merely: 'There is a self inside which does the deciding.' Or, let us say, it is as if the *Pfarrer* were explaining the physiology of a horse's motion. They take the whole thing to be analysed, imagine a duplicate of it reduced in size, so to speak, and place this duplicate-self inside as an explanation – making it the elusive source of the 'free decisions'. They do not see that they are merely pushing the question a little further back, since the process of deciding, with its constituent factors, must have taken place within that inner self. Either it decided in a particular way because, on the whole, it preferred to decide in that way, or the decision was an underived event, a rootless and sourceless event. It is the same story over again. In neither case is there any gain in imagining a second self inside, however wonderful and elusive. Of course, it is the first alternative that the indeterminist is really imagining. If you tacitly and obscurely conceive the self as deciding *its own way*, i.e., according to its preference, but never admit or recognise this, then you can happily remain a libertarian indeterminist; but upon no other terms. In your theory there is a heart of darkness.

Freedom

In accordance with the genius of language, free will means freedom of persons in willing, just as 'free trade' means freedom of persons (in a certain respect) in trading. The freedom of anyone surely always implies his possession of a power, and means the absence of any interference (whether taking the form of restraint or constraint) with his exercise of that power. Let us consider this in relation to freedom in willing.

'Can'

We say, 'I can will this or I can will that, whichever I choose.' Two courses of

action present themselves to my mind. I think of their consequences, I look on this picture and on that, one of them commends itself more than the other, and I will an act that brings it about. I knew that I could choose either. That means that I had the power to choose either.

What is the meaning of ‘power’? A person has a power if it is a fact that when he sets himself in the appropriate manner to produce a certain event that event will actually follow. I have the power to lift the lamp; that is, if I grasp it and exert an upward pressure with my arm, *it will rise*. I have the power to will so and so; that is, if I want, that act of will will take place. That and none other is the meaning of power, is it not? A man’s being in the proper active posture of body or of mind is the cause, and the sequel in question will be the effect. (Of course, it may be held that the sequel not only does but must follow, in a sense opposed to Hume’s doctrine of cause. Very well; the question does not here concern us.)

Thus power depends upon, or rather consists in, a law. The law in question takes the familiar form that if something happens a certain something else will ensue. If A happens then B will happen. The law in this case is that if the man definitively so desires then volition will come to pass. There is a series, wish – will – act. The act follows according to the will (that is a law – I do not mean an underived law) and the will follows according to the wish (that is another law). A man has the power (sometimes) to act as he wishes. He has the power (when-ever he is not physically bound or held) to act as he wills. He has the power always (except in certain morbid states) to will as he wishes. All this depends upon the laws of his being. Wherever there is a power there is a law. In it the power wholly consists. A man’s power to will as he wishes is simply the law that his will follows his wish.

What, again, does freedom mean? It means the absence of any interference with all this. Nothing steps in to prevent my exercising my power.¹

All turns on the meaning of ‘can’. ‘I can will either this or that’ means, I am so constituted that if I definitively incline to this, the appropriate act of will will take place, and if I definitively incline to that, the appropriate act of will will take place. The law connecting preference and will exists, and there is nothing to interfere with it. My free power, then, is not an exemption from law but in its inmost essence an embodiment of law.

Thus it is true, after the act of will, that I could have willed otherwise. It is most natural to add, ‘if I had wanted to’; but the addition is not required. The point is the meaning of ‘could’. I could have willed whichever way I pleased. I had the power to will otherwise, there was nothing to prevent my doing so, and I should have done so if I had wanted. If someone says that the wish I actually had prevented my willing otherwise, so that I could not have done it, he is merely making a slip in the use of the word ‘could’. He means, that wish could not have produced anything but this volition. But ‘could’ is asserted not of the wish (a transient fact to which power in this sense is not and should not be ascribed) but of the person. And the person *could* have produced something else than that volition. He could have produced any volition he wanted; he had the power to do so.

But the objector will say, ‘The person as he was at the moment – the person as animated by that wish – could not have produced any other volition.’ Oh, yes, he could. ‘Could’ has meaning as applied not to a momentary actual phase of a person’s life, but to the person himself of whose life that is but a phase; and it means that (even at that moment) he had the power to will just as he preferred. *The idea of power, because it is the idea of a law, is hypothetical, carries in itself hypothesis as part of its very intent and meaning – if he should prefer this, if he should prefer that, – and therefore can be truly applied to a person irrespective of what at the moment he does prefer. It remains hypothetical even when applied.*² This very peculiarity of its meaning is the whole point of the idea of power. It is just because determinism is true, because a law obtains, that one ‘could have done otherwise’.

Sidgwick set over against ‘the formidable array of cumulative evidence’ offered for determinism the ‘affirmation of consciousness’ ‘that I can now choose to do’ what is right and reasonable, ‘however strong may be my inclination to act unreasonably’.³ But it is not against determinism. It is a true affirmation (surely not of immediate consciousness but of experience), the affirmation of my power to will what I deem right, however intense and insistent my desire for the wrong. I can will anything, and can will effectively anything that my body will enact. I can will it despite an inclination to the contrary of any strength you please – strength as felt by me before decision. We all know cases where we have resisted impulses of great strength in this sense and we can imagine them still stronger. I have the power to do it, and shall do it, shall exercise that power, if I prefer. Obviously in that case (be it psychologically remarked) my solicitude to do what is right will have proved itself even stronger (as measured by ultimate tendency to prevail, though not of necessity by sensible vividness or intensity) than the inclination to the contrary, for that is what is meant by my preferring to do it. I am conscious that the field for willing is open; ‘I can will’ anything that I elect to will. Sidgwick did not analyse the meaning of ‘can’, that is all. He did not precisely catch the outlook of consciousness when it says, ‘I can.’ He did not distinguish the function of the word, which is to express the availability of the alternatives I see when, before I have willed, and perhaps before my preference is decided, I look out on the field of conceivable volition. He did not recognise that I must have a word to express my power to will as I please, quite irrespective of what I shall please, and that ‘can’ is that word. It is no proof that I cannot do something to point out that I shall not do it if I do not prefer. A man, let us say, can turn on the electric light; but he will not turn it on if he walks away from it; though it is still true that he can turn it on. When we attribute power to a man we do not mean that something will accomplish itself without his wanting it to. That would never suggest the idea of power. We mean that if he makes the requisite move the thing will be accomplished. It is part of the idea that the initiative shall rest with him. The initiative for an act of will is a precedent phase of consciousness that we call the definitive inclination, or, in case of conflict, the definitive preference for it. If someone in the throes of struggle with temptation says to himself, ‘I can put this behind me,’ he is saying truth and precisely the pertinent truth. He is bringing before his mind the act of

will, un prevented, quite open to him, that would deliver him from what he deems noxious. It may still happen that the noxiousness of the temptation does not affect him so powerfully as its allurement, and that he succumbs. It is no whit less true, according to determinism, that he could have willed otherwise. To analyse the fact expressed by 'could' is not to destroy it.

But it may be asked, 'Can I will in opposition to my strongest desire at the moment when it is strongest?' If the words 'at the moment when it is strongest' qualify 'can', the answer has already been given. If they qualify 'will', the suggestion is a contradiction in terms. Can I turn-on-the-electric-light-at-a-moment-when-I-am-not-trying-to-do-so? This means, if I try to turn on the light at a moment when I am not trying to, will it be turned on? A possible willing as I do not prefer to will is not a power on my part, hence not to be expressed by 'I can.'

Everybody knows that we often will what we do not want to will, what we do not prefer. But when we say this we are using words in another sense than that in which I have just used them. In *one* sense of the words, whenever we act we are doing what we prefer, on the whole, in view of all the circumstances. We are acting for the greatest good or the least evil or a mixture of these. In the *other* and more usual sense of the words, we are very often doing what we do not wish to do, i.e., doing some particular thing we do not wish because we are afraid of the consequences or disapprove of the moral complexion of the particular thing we do wish. We do the thing that we do not like because the other thing has aspects that we dislike yet more. We are still doing what we like best on the whole. It is again a question of the meaning of words.

If the initiative for volition is not a wish, what is it? Indeterminism says that a moral agent sometimes decides against the more tempting course. He does so, let us say, because it is wrong, the other course is the right one. In other words, the desire to do right is at the critical moment stronger within him than the temptation. No, no, replies indeterminism, it is not that; he sometimes decides against the stronger desire. Very well; 'can' meaning what it does, tell us what is the leaning or favourable disposition on the part of the ego, in a case of undetermined willing, toward the volition it adopts; what is that which constitutes the ego's initiative in that direction, — since it is not a wish? Shall we say it is an approval or conscientious acceptance? Does this approval or acceptance arise from the agent's distinctive moral being? That is determinism, quite as much as if you called the initiative a wish. But the indeterminist has already answered in effect that there is no such initiative, or no effectual initiative. The act of will causes the physical act but is not itself caused. This is to deny the presence of power, according to its definition. How has it a meaning to say in advance that 'I can' will this way or that? The self, considering the alternatives beforehand, is not in a position to say, 'If I feel thus about it, this volition will take place, or if I feel otherwise the contrary will take place; I know very well how I shall feel, so I know how I shall will.' The self now existing has not control over the future 'free' volition, since that may be undetermined, nor will the self's future feelings, whatever they may be, control it. Hence the sense expressed by 'I can', the sense of power inhering in one's continuous self to sway the volition as it feels

disposed, is denied to it. All it is in a position to mean by 'I can' is, 'I do not know which will happen', which is not 'I can' at all. Nay, even looking backward, it is unable to say: 'I could have willed otherwise', for that clearly implies, 'Had I been so disposed the other volition would have taken place', which is just what cannot, according to indeterminism, be said. Surely, to paraphrase a historic remark, our 'liberty' does not seem to be of very much use to us. The indeterminist is in a peculiarly hapless position. The two things that he is most deeply moved to aver, that the free volition is the act of the self, and that the self can will one way or the other – these two things on his own theory fall utterly to pieces, and can only be maintained on the view that he opposes.

Compulsion

The indeterminist conceives that according to determinism the self is carried along by wishes to acts which it is thus necessitated to perform. This mode of speaking distinguishes the self from the wishes and represents it as under their dominion. This is the initial error. This is what leads the indeterminist wrong on all the topics of his problem. And the error persists in the most recent writings. In fact, the moral self is the wishing self. The wishes are its own. It cannot be described as under their dominion, for it has no separate predilections to be overborne by them; they themselves are its predilections. To fancy that because the person acts according to them he is compelled, a slave, the victim of a power from whose clutches he cannot extricate himself, is a confusion of ideas, a mere slip of the mind. The answer that has ordinarily been given is surely correct; all compulsion is causation, but not all causation is compulsion. Seize a man and violently force him to do something, and he is compelled – also caused – to do it. But induce him to do it by giving him reasons and his doing it is caused but not compelled.

Passivity

We have to be on our guard even against conceiving the inducement as a cause acting like the impact of a billiard ball, by which the self is precipitated into action like a second billiard ball, as an effect. The case is not so simple. Your reasons have shown him that his own preferences require the action. He does it of his own choice; he acts from his own motives in the light of your reasons. The sequence of cause and effect goes on within the self, with contributory information from without.

It is not clarifying to ask, 'Is a volition free or determined?' It is the person who is free, and his particular volition that is determined. Freedom is something that we can attribute only to a continuing being, and he can have it only so far as the particular transient volitions within him are determined. (According to the strict proprieties of language, it is surely events that are caused, not things or persons; a person or thing can be caused or determined only in the

sense that its beginning to be, or changes in it, are caused or determined.)

It is fancied that, owing to the ‘necessity’ with which an effect follows upon its cause, if my acts of will are caused I am not free in thus acting. Consider an analogous matter. When I move I use ligaments. ‘Ligament’ means that which binds, and a ligament does bind bones together. But *I* am not bound. *I* (so far as my organism is concerned) am rendered possible by the fact that my bones are bound one to another; that is part of the secret of my being able to act, to move about and work my will. If my bones ceased to be bound one to another I should be undone indeed. The human organism is detached, but it is distinctly important that its component parts shall not be detached. Just so my free power of willing is built up of tight cause-and-effect connections. The point is that when I employ the power thus constituted nothing determines the particular employment of it but *me*. Each particular act of mine is determined from outside itself, i.e., by a cause, a prior event. But not from outside me. I, the possessor of the power, am not in my acts passively played upon by causes outside me, but am enacting my own wishes in virtue of a chain of causation within me. What is needed is to distinguish broadly between a particular effect, on the one hand, and, on the other, the detached, continuous life of a mental individual and his organism; a life reactive, but reacting according to its own nature. . . .

Prediction

If we knew a man’s character thoroughly and the circumstances that he would encounter, determinism (which we are not here completely asserting) says that we could foretell his conduct. This is a thought that repels many libertarians. Yet to predict a person’s conduct need not be repellent. If you are to be alone in a room with £1000 belonging to another on the table and can pocket it without anyone knowing the fact, and if I predict that you will surely *not* pocket it, that is not an insult. I say, I know you, I know your character; you will not do it. But if I say that you are ‘a free being’ and that I really do not know whether you will pocket it or not, that is rather an insult. On the other hand, there are cases where prediction is really disparaging. If I say when you make a remark, ‘I knew you were going to say that’, the impression is not agreeable. My exclamation seems to say that your mind is so small and simple that one can predict its ideas. That is the real reason why people resent in such cases our predicting their conduct; that if present human knowledge, which is known to be so limited, can foresee their conduct, it must be more naive and stereotyped than they like to think it. It is no reflection upon the human mind or its freedom to say that one who knew it through and through (a human impossibility) could foreknow its preferences and its spontaneous choice. It is of the very best of men that even we human beings say, ‘I am sure of him.’ It has perhaps in this controversy hardly been observed how much at this point is involved, how far the question of prediction reaches. The word ‘reliable’ or ‘trustworthy’ is a prediction of behaviour. Indeed, all judgement of persons whatever, in the measure of its definitude, is such a prediction.

Material Fate

The philosopher in the old story, gazing at the stars, falls into a pit. We have to notice the pitfall in our subject to which, similarly occupied, Professor Eddington has succumbed.

What significance is there in my mental struggle to-night whether I shall or shall not give up smoking, if the laws which govern the matter of the physical universe already pre-ordain for the morrow a configuration of matter consisting of pipe, tobacco, and smoke connected with my lips?⁴

No laws, according to determinism, pre-ordain such a configuration, unless I give up the struggle. Let us put matter aside for the moment, to return to it. Fatalism says that my morrow is determined no matter how I struggle. This is of course a superstition. Determinism says that my morrow is determined through my struggle. There is this significance in my mental effort, that it is deciding the event. The stream of causation runs through my deliberations and decision, and, if it did not run as it does run, the event would be different. The past cannot determine the event except through the present. And no past moment determined it any more truly than does the present moment. In other words, each of the links in the causal chain must be in its place. Determinism (which, the reader will remember, we have not here taken for necessarily true in all detail) says that the coming result is 'pre-ordained' (literally, caused) at each stage, and therefore the whole following series for to-morrow may be described as already determined; so that did we know all about the struggler, how strong of purpose he was and how he was influenced (which is humanly impossible), we could tell what he would do. But for the struggler this fact (supposing it to be such) is not pertinent. If, believing it, he ceases to struggle, he is merely revealing that the forces within him have brought about that cessation. If on the other hand he struggles manfully he will reveal the fact that they have brought about his success. Since the causation of the outcome works through his struggle in either case equally, it cannot become for him a moving consideration in the struggle. In it the question is, 'Shall I do this or that?' It must be answered in the light of what there is to recommend to me this or that. To this question the scientific truth (according to determinism) that the deliberation itself is a play of causation is completely irrelevant; it merely draws the mind delusively away from the only considerations that concern it.

As regards the rôle of matter in the affair, if, as Professor Eddington on behalf of the determinists is here supposing, the behaviour of all matter, including the human organism, takes place according to a deterministic scheme of physical law, then we must conceive, according to the familiar formula, that the mental process is paralleled in the brain by a physical process. The whole psycho-physical occurrence would then be the cause of what followed, and the psychic side of it, the mental struggle proper, a concuse or side of the cause. To-morrow's configuration of matter will have been brought about by a material process with which the mental process was inseparably conjoined. I make this supposition

merely to show that supposing the existence of a physically complete mechanism through which all human action is caused and carried out has no tendency to turn determinism into fatalism. For the mental struggle must in that case be paralleled by a physical struggle which, so to speak, represents it and is in a manner its agent in the physical world; and upon this struggle the physical outcome will depend. (The determinist need not, but may of course, hold this doctrine of automatism, of a physically complete mechanism in human action.) . . .

Responsibility

Again, it is said that determinism takes from man all responsibility. As regards the origin of the term, a man is responsible when he is the person to respond to the question why the act was performed, how it is to be explained or justified. That is what he must answer; he is answerable for the act. It is the subject of which he must give an account; he is accountable for the act. The act proceeded from him. He is to say whether it proceeded consciously. He is to give evidence that he did or did not know the moral nature of the act and that he did or did not intend the result. He is to say how he justifies it or if he can justify it. If the act proceeded from him by pure accident, if he can show that he did the damage (if damage it was) by brushing against something by inadvertence, for example, then he has not to respond to the question what he did it for – he is not consciously responsible – nor how it is justified – he is not morally responsible, though of course he may have been responsible in these respects for a habit of carelessness.

But why does the peculiar moral stain of guilt or ennoblement of merit belong to responsibility? If an act proceeds from a man and not merely from his accidental motion but from his mind and moral nature, we judge at once that like acts may be expected from him in the future. The colour of the act for good or bad is reflected on the man. We see him now as a living source of possible acts of the same kind. If we must be on our guard against such acts we must be on our guard against such men. If we must take steps to defend ourselves against such acts we must take steps to defend ourselves against such men. If we detest such acts, we must detest that tendency in such men which produced them. He is guilty in that he knowingly did evil, in that the intentional authorship of evil is in him. Because the act proceeded in every sense from him, for that reason he is (so far) to be accounted bad or good according as the act is bad or good, and he is the one to be punished if punishment is required. And that is moral responsibility.

But how, it is asked, can I be responsible for what I will if a long train of past causes has made me will it. . . . Is it not these causes that are ‘responsible’ for my act – to use the word in the only sense, says the objector, that seems to remain for it?

The parent past produced the man, none the less the man is responsible for his acts. We can truly say that the earth bears apples, but quite as truly that trees

bear apples. The earth bears the apples by bearing trees. It does not resent the claim of the trees to bear the apples, or try to take the business out of the trees' hands. Nor need the trees feel their claim nullified by the earth's part in the matter. There is no rivalry between them. A man is a being with free will and responsibility; where this being came from, I repeat, is another story. The past finished its functions in the business when it generated him as he is. So far from interfering with him and coercing him the past does not even exist. If we could imagine it as lingering on into the present, standing over against him and stretching out a ghostly hand to stay his arm, then indeed the past would be interfering with his liberty and responsibility. But so long as it and he are never on the scene together they cannot wrestle; the past cannot overpower him. The whole alarm is an evil dream, a nightmare due to the indigestion of words. The past has created, and left extant, a free-willed being.

Notes

- 1 A word as to the relation of power and freedom. Strictly power cannot exist without freedom, since the result does not follow without it. Freedom on the other hand is a negative term, meaning the absence of something, and implies a power only because that whose absence it signifies is interference, which implies something to be interfered with. Apart from this peculiarity of the term itself, there might be freedom without any power. Absence of interference (of what would be interference if there were a power) might exist in the absence of a power; a man might be free to do something because there was nothing to interfere with his doing it, but might have no power to do it. Similarly and conveniently we may speak of a power as existing though interfered with; that is, the law may exist that would constitute a power if the interference were away.
- 2 I am encouraged by finding in effect the same remark in Prof. G. E. Moore's *Ethics* (Oxford: Oxford University Press, 1912), ch. vi, at least as regards what he terms one sense of the word 'could'. I should hazard saying, the only sense in this context.
- 3 Sidgwick, *Methods of Ethics*, 7th edn (London: Macmillan, 1907), p. 65.
- 4 *Philosophy*, Jan. 1933, p. 41.

40 Human Freedom and the Self*

Roderick M. Chisholm

A staff moves a stone, and is moved by a hand, which is moved by a man.

Aristotle, *Physics*, 256a

1. The metaphysical problem of human freedom might be summarized in the following way: Human beings are responsible agents; but this fact appears to conflict with a deterministic view of human action (the view that every event that is involved in an act is caused by some other event); and it *also* appears to conflict with an indeterministic view of human action (the view that the act, or some event that is essential to the act, is not caused at all). To solve the problem, I believe, we must make somewhat far-reaching assumptions about the self or the agent – about the man who performs the act.

Perhaps it is needless to remark that, in all likelihood, it is impossible to say anything significant about this ancient problem that has not been said before.¹

2. Let us consider some deed, or misdeed, that may be attributed to a responsible agent: one man, say, shot another. If the man *was* responsible for what he did, then, I would urge, what was to happen at the time of the shooting was something that was entirely up to the man himself. There was a moment at which it was true, both that he could have fired the shot and also that he could have refrained from firing it. And if this is so, then, even though he did fire it, he could have done something else instead. (He didn't find himself firing the shot "against his will," as we say.) I think we can say, more generally, then, that if a man is responsible for a certain event or a certain state of affairs (in our example, the shooting of another man), then that event or state of affairs was brought about by some act of his, and the act was something that was in his power either to perform or not to perform.

But now if the act which he *did* perform was an act that was also in his power *not* to perform, then it could not have been caused or determined by any event that was not itself within his power either to bring about or not to bring about. For example, if what we say he did was really something that was brought about by a second man, one who forced his hand upon the trigger, say, or who, by means of hypnosis, compelled him to perform the act, then since the act was caused by the *second* man it was nothing that was within the power of the *first* man to prevent. And precisely the same thing is true, I think, if instead of referring to a second man who compelled the first one, we speak instead of the

* From the Lindley Lecture, 1964. Copyright © 1964 by the Department of Philosophy, University of Kansas. Reprinted by permission of the author and the Department of Philosophy, University of Kansas, Lawrence, Kansas.

desires and *beliefs* which the first man happens to have had. For it what we say he did was really something that was brought about by his own beliefs and desires, if these beliefs and desires in the particular situation in which he happened to have found himself caused him to do just what it was that we say he did do, then since *they* caused it, *he* was unable to do anything other than just what it was that he did do. It makes no difference whether the cause of the deed was internal or external; if the cause was some state or event for which the man himself was not responsible, then he was not responsible for what we have been mistakenly calling his act. If a flood caused the poorly constructed dam to break, then, given the flood and the constitution of the dam, the break, we may say, *had* to occur and nothing could have happened in its place. And if the flood of desire caused the weak-willed man to give in, then he, too, had to do just what it was that he did do and he was no more responsible than was the dam for the results that followed. (It is true, of course, that if the man is responsible for the beliefs and desires that he happens to have, then he may also be responsible for the things they lead him to do. But the question now becomes: *is* he responsible for the beliefs and desires he happens to have? If he is, then there was a time when they were within his power either to acquire or not to acquire, and we are left, therefore, with our general point.)

One may object: But surely if there were such a thing as a man who is really *good*, then he would be responsible for things that he would do; yet, he would be unable to do anything other than just what it is that he does do, since, being good, he will always choose to do what is best. The answer, I think, is suggested by a comment that Thomas Reid makes on an ancient author. The author had said of Cato, "He was good because he could not be otherwise," and Reid observes: "But this saying, if understood literally and strictly, is not the praise of Cato, but of his constitution, which was no more the work of Cato, than his existence."² If Cato was himself responsible for the good things that he did, then Cato, as Reid suggests, was such that, although he had the power to do what was not good, he exercised his power only for that which was good.

All of this, if it is true, may give a certain amount of comfort to those who are tender-minded. But we should remind them that it also conflicts with a familiar view about the nature of God – with the view that St Thomas Aquinas expresses by saying that "every movement both of the will and of nature proceeds from God as the Prime Mover."³ If the act of the sinner *did* proceed from God as the Prime Mover, then God was in the position of the second agent we just discussed – the man who forced the trigger finger, or the hypnotist – and the sinner, so-called, was *not* responsible for what he did. (This may be a bold assertion, in view of the history of western theology, but I must say that I have never encountered a single good reason for denying it.)

There is one standard objection to all of this and we should consider it briefly.

3. The objection takes the form of a stratagem – one designed to show that determinism (and divine providence) is consistent with human responsibility. The stratagem is one that was used by Jonathan Edwards and by many philoso-

phers in the present century, most notably, G. E. Moore.⁴

One proceeds as follows: The expression

- (a) He could have done otherwise,

it is argued, means no more nor less than

- (b) If he had chosen to do otherwise, then he would have done otherwise.

(In place of “chosen,” one might say “tried,” “set out,” “decided,” “undertaken,” or “willed.”) The truth of statement (b), it is then pointed out, is consistent with determinism (and with divine providence); for even if all of the man’s actions were causally determined, the man could still be such that, *if* he had chosen otherwise, then he would have done otherwise. What the murderer saw, let us suppose, along with his beliefs and desires, *caused* him to fire the shot; yet he was such that *if*, just then, he had chosen or decided *not* to fire the shot, then he would not have fired it. All of this is certainly possible. Similarly, we could say, of the dam, that the flood caused it to break and also that the dam was such that, *if* there had been no flood or any similar pressure, then the dam would have remained intact. And therefore, the argument proceeds, if (b) is consistent with determinism, and if (a) and (b) say the same thing, then (a) is also consistent with determinism; hence we can say that the agent *could* have done otherwise even though he was caused to do what he did do; and therefore determinism and moral responsibility are compatible.

Is the argument sound? The conclusion follows from the premises, but the catch, I think, lies in the first premise – the one saying that statement (a) tells us no more nor less than what statement (b) tells us. For (b), it would seem, could be true while (a) is false. That is to say, our man might be such that, if he had chosen to do otherwise, then he would have done otherwise, and yet *also* such that he could not have done otherwise. Suppose, after all, that our murderer could not have *chosen*, or could not have *decided*, to do otherwise. Then the fact that he happens also to be a man such that, if he had chosen not to shoot he would not have shot, would make no difference. For if he could *not* have chosen *not* to shoot, then he could not have done anything other than just what it was that he did do. In a word: from our statement (b) above (“If he had chosen to do otherwise, then he would have done otherwise”), we cannot make an inference to (a) above (“He could have done otherwise”) unless we can *also* assert:

- (c) He could have chosen to do otherwise.

And therefore, if we must reject this third statement (c), then, even though we may be justified in asserting (b), we are not justified in asserting (a). If the man could not have chosen to do otherwise, then he would not have done otherwise – *even if* he was such that, if he *had* chosen to do otherwise, then he would have done otherwise.

The stratagem in question, then, seems to me not to work, and I would say, therefore, that the ascription of responsibility conflicts with a deterministic view of action.

4. Perhaps there is less need to argue that the ascription of responsibility also conflicts with an indeterministic view of action – with the view that the act, or some event that is essential to the act, is not caused at all. If the act – the firing of the shot – was not caused at all, if it was fortuitous or capricious, happening so to speak out of the blue, then, presumably, no one – and nothing – was responsible for the act. Our conception of action, therefore, should be neither deterministic nor indeterministic. Is there any other possibility?

5. We must not say that every event involved in the act is caused by some other event; and we must not say that the act is something that is not caused at all. The possibility that remains, therefore, is this: We should say that at least one of the events that are involved in the act is caused, not by any other events, but by something else instead. And this something else can only be the agent – the man. If there is an event that is caused, not by other events, but by the man, then there are some events involved in the act that are not caused by other events. But if the event in question is caused by the man then it is caused and we are not committed to saying that there is something involved in the act that is not caused at all.

But this, of course, is a large consequence, implying something of considerable importance about the nature of the agent or the man.

6. If we consider only inanimate natural objects, we may say that causation, if it occurs, is a relation between *events* or *states of affairs*. The dam's breaking was an event that was caused by a set of other events – the dam being weak, the flood being strong, and so on. But if a man is responsible for a particular deed, then, if what I have said is true, there is some event, or set of events, that is caused, *not* by other events or states of affairs, but by the agent, whatever he may be.

I shall borrow a pair of medieval terms, using them, perhaps, in a way that is slightly different from that for which they were originally intended. I shall say that when one event or state of affairs (or set of events or states of affairs) causes some other event or state of affairs, then we have an instance of *transeunt* causation. And I shall say that when an *agent*, as distinguished from an event, causes an event or state of affairs, then we have an instance of *immanent* causation.

The nature of what is intended by the expression "immanent causation" may be illustrated by this sentence from Aristotle's *Physics*. "Thus, a staff moves a stone, and is moved by a hand, which is moved by a man" (Book VII, Chap. 5, 256a, 6–8). If the man was responsible, then we have in this illustration a number of instances of causation – most of them *transeunt* but at least one of them *immanent*. What the staff did to the stone was an instance of *transeunt* causation, and thus we may describe it as a relation between events: "the motion of

the staff caused the motion of the stone.” And similarly for what the hand did to the staff: “the motion of the hand caused the motion of the staff.” And, as we know from physiology, there are still other events which caused the motion of the hand. Hence we need not introduce the agent at this particular point, as Aristotle does – we *need* not, though we *may*. We *may* say that the hand was moved by the man, but we may *also* say that the motion of the hand was caused by the motion of certain muscles; and we may say that the motion of the muscles was caused by certain events that took place within the brain. But some event, and presumably one of those that took place within the brain, was caused by the agent and not by any other events.

There are, of course, objections to this way of putting the matter; I shall consider the two that seem to me to be most important.

7. One may object, firstly: “If the *man* does anything, then, as Aristotle’s remark suggests, what he does is to move the *hand*. But he certainly does not *do* anything to his brain – he may not even know that he *has* a brain. And if he doesn’t do anything to the brain, and if the motion of the hand was caused by something that happened within the brain, then there is no point in appealing to ‘immanent causation’ as being something incompatible with ‘transeunt causation’ – for the whole thing, after all, is a matter of causal relations among events or states of affairs.”

The answer to this objection, I think, is this: It is true that the agent does not *do* anything with his brain, or to his brain, in the sense in which he *does* something with his hand and does something to the staff. But from this it does not follow that the agent was not the immanent cause of something that happened within his brain.

We should note a useful distinction that has been proposed by Professor A. I. Melden – namely, the distinction between “making something A happen” and “doing A.”⁵ If I reach for the staff and pick it up, then one of the things that I *do* is just that – reach for the staff and pick it up. And if it is something that I do, then there is a very clear sense in which it may be said to be something that I know that I do. If you ask me, “Are you doing something, or trying to do something, with the staff?”, I will have no difficulty in finding an answer. But in doing something with the staff, I also make various things happen which are not in this same sense things that I do: I will make various air-particles move; I will free a number of blades of grass from the pressure that had been upon them; and I may cause a shadow to move from one place to another. If these are merely things that I make happen, as distinguished from things that I do, then I may know nothing whatever about them; I may not have the slightest idea that, in moving the staff, I am bringing about any such thing as the motion of air-particles, shadows, and blades of grass.

We may say, in answer to the first objection, therefore, that it is true that our agent does nothing to his brain or with his brain; but from this it does not follow that the agent is not the immanent cause of some event within his brain; for the brain event may be something which, like the motion of the air-particles, he made happen in picking up the staff. The only difference between the

two cases is this: in each case, he made something happen when he picked up the staff; but in the one case – the motion of the air-particles or of the shadows – it was the motion of the staff that caused the event to happen; and in the other case – the event that took place in the brain – it was this event that caused the motion of the staff.

The point is, in a word, that whenever a man does something A, then (by “immanent causation”) he makes a certain cerebral event happen, and this cerebral event (by “transeunt causation”) makes A happen.

8. The second objection is more difficult and concerns the very concept of “immanent causation,” or causation by an agent, as this concept is to be interpreted here. The concept is subject to a difficulty which has long been associated with that of the prime mover unmoved. We have said that there must be some event A, presumably some cerebral event, which is caused not by any other event, but by the agent. Since A was not caused by any other event, then the agent himself cannot be said to have undergone any change or produced any other event (such as “an act of will” or the like) which brought A about. But if, when the agent made A happen, there was no event involved other than A itself, no event which could be described as *making* A happen, what did the agent’s causation consist of? What, for example, is the difference between A’s just happening, and the agent’s *causing* A to happen? We cannot attribute the difference to any event that took place within the agent. And so far as the event A itself is concerned, there would seem to be no discernible difference. Thus Aristotle said that the activity of the prime mover is nothing in addition to the motion that it produces, and Suarez said that “the action is in reality nothing but the effect as it flows from the agent.”⁶ Must we conclude, then, that there is no more to the man’s action in causing event A than there is to the event A’s happening by itself? Here we would seem to have a distinction without a difference – in which case we have failed to find a *via media* between a deterministic and an indeterministic view of action.

The only answer, I think, can be this: that the difference between the man’s causing A, on the one hand, and the event A just happening, on the other, lies in the fact that, in the first case but not the second, the event A *was* caused and was caused by the man. There was a brain event A; the agent did, in fact, cause the brain event; but there was nothing that he did to cause it.

This answer may not entirely satisfy and it will be likely to provoke the following question: “But what are you really *adding* to the assertion that A happened when you utter the words ‘The agent *caused* A to happen?’” As soon as we have put the question this way, we see, I think, that whatever difficulty we may have encountered is one that may be traced to the concept of causation generally – whether “immanent” or “transeunt.” The problem, in other words, is not a problem that is peculiar to our conception of human action. It is a problem that must be faced by anyone who makes use of the concept of causation at all; and therefore, I would say, it is a problem for everyone but the complete indeterminist.

For the problem, as we put it, referring just to “immanent causation,” or

causation by an agent, was this: "What is the difference between saying, of an event A, that A just happened and saying that someone caused A to happen?" The analogous problem, which holds for "transeunt causation," or causation by an event, is this: "What is the difference between saying, of two events A and B, that B happened and then A happened, and saying that B's happening was the *cause* of A's happening?" And the only answer that one can give is this – that in the one case the agent was the cause of A's happening and in the other case event B was the cause of A's happening. The nature of transeunt causation is no more clear than is that of immanent causation.

9. But we may plausibly say – and there is a respectable philosophical tradition to which we may appeal – that the notion of immanent causation, or causation by an agent, is in fact more clear than that of transeunt causation, or causation by an event; and that it is only by understanding our own causal efficacy, as agents, that we can grasp the concept of *cause* at all. Hume may be said to have shown that we do not derive the concept of *cause* from what we perceive of external things. How, then, do we derive it? The most plausible suggestion, it seems to me, is that of Reid, once again: namely that "the conception of an efficient cause may very probably be derived from the experience we have had . . . of our own power to produce certain effects."⁷ If we did not understand the concept of immanent causation, we would not understand that of transeunt causation.

10. It may have been noted that I have avoided the term "free will" in all of this. For even if there is such a faculty as "the will," which somehow sets our acts agoing, the question of freedom, as John Locke said, is not the question "*whether the will be free*"; it is the question "*whether a man be free*.⁸" For if there is a "will," as a moving faculty, the question is whether the man is free to will to do these things that he does will to do – and also whether he is free *not* to will any of those things that he does will to do, and, again, whether he is free to will any of those things that he does not will to do. Jonathan Edwards tried to restrict himself to the question – "Is the man free to do what it is that he wills?" – but the answer to the question will not tell us whether the man is responsible for what it is that he *does* will to do. Using still another pair of medieval terms, we may say that the metaphysical problem of freedom does not concern the *actus imperatus*; it does not concern the question whether we are free to accomplish whatever it is that we will or set out to do; it concerns the *actus elicitus*, the question whether we are free to will or to set out to do those things that we do will or set out to do.

11. If we are responsible, and if what I have been trying to say is true, then we have a prerogative which some would attribute only to God: each of us, when we act, is a prime mover unmoved. In doing what we do, we cause certain events to happen, and nothing – or no one – causes us to cause those events to happen.

12. If we are thus prime movers unmoved and if our actions, or those for which we are responsible, are not causally determined, then they are not causally determined by our *desires*. And this means that the relation between what we want or what we desire, on the one hand, and what it is that we do, on the other, is not as simple as most philosophers would have it.

We may distinguish between what we might call the “Hobbist approach” and what we might call the “Kantian approach” to this question. The Hobbist approach is the one that is generally accepted at the present time, but the Kantian approach, I believe, is the one that is true. According to Hobbism, if we *know*, of some man, what his beliefs and desires happen to be and how strong they are, if we know what he feels certain of, what he desires more than anything else, and if we know the state of his body and what stimuli he is being subjected to, then we may *deduce*, logically, just what it is that he will do – or, more accurately, just what it is that he will try, set out, or undertake to do. Thus Professor Melden has said that “the connection between wanting and doing is logical.”⁹ But according to the Kantian approach to our problem, and this is the one that I would take, there is no such logical connection between wanting and doing, nor need there even be a causal connection. No set of statements about a man’s desires, beliefs, and stimulus situation at any time implies any statement telling us what the man will try, set out, or undertake to do at that time. As Reid put it, though we may “reason from men’s motives to their actions and, in many cases, with great probability,” we can never do so “with absolute certainty.”¹⁰

This means that, in one very strict sense of the terms, there can be no science of man. If we think of science as a matter of finding out what laws happen to hold, and if the statement of a law tells us what kinds of events are caused by what other kinds of events, then there will be human actions which we cannot explain by subsuming them under any laws. We cannot say, “It is causally necessary that, given such and such desires and beliefs, and being subject to such and such stimuli, the agent will do so and so.” For at times the agent, if he chooses, may rise above his desires and do something else instead.

But all of this is consistent with saying that, perhaps more often than not, our desires do exist under conditions such that those conditions necessitate us to act. And we may also say, with Leibniz, that at other times our desires may “incline without necessitating.”

13. Leibniz’s phrase presents us with our final philosophical problem. What does it mean to say that a desire, or a motive, might “incline without necessitating”? There is a temptation, certainly, to say that “to incline” means to cause and that “not to necessitate” means not to cause, but obviously we cannot have it both ways. . . .

Let us consider a public official who has some moral scruples but who also, as one says, could be had. Because of the scruples that he does have, he would never take any positive steps to receive a bribe – he would not actively solicit one. But his morality has its limits and he is also such that, if we were to confront him with a *fait accompli* or to let him see what is about to happen (\$10,000 in cash is being deposited behind the garage), then he would succumb and be

unable to resist. The general situation is a familiar one and this is one reason that people pray to be delivered from temptation. (It also justifies Kant's remark: "And how many there are who may have led a long blameless life, who are only fortunate in having escaped so many temptations."¹¹) Our relation to the misdeed that we contemplate may not be a matter simply of being able to bring it about or not to bring it about. As St Anselm noted, there are at least four possibilities. We may illustrate them by reference to our public official and the event which is his receiving the bribe, in the following way: (i) he may be able to bring the event about himself (*facere esse*), in which case he would actively cause himself to receive the bribe; (ii) he may be able to refrain from bringing it about himself (*non facere esse*), in which case he would not himself do anything to insure that he receive the bribe; (iii) he may be able to do something to prevent the event from occurring (*facere non esse*), in which case he would make sure that the \$10,000 was *not* left behind the garage; or (iv) he may be unable to do anything to prevent the event from occurring (*non facere non esse*), in which case, though he may not solicit the bribe, he would allow himself to keep it.¹² We have envisaged our official as a man who can resist the temptation to (i) but cannot resist the temptation to (iv): he can refrain from bringing the event about himself, but he cannot bring himself to do anything to prevent it.

Let us think of "inclination without necessitation," then, in such terms as these. First we may contrast the two propositions:

- (1) He can resist the temptation to do something in order to make A happen;
- (2) He can resist the temptation to allow A to happen (i.e. to do nothing to prevent A from happening).

We may suppose that the man has some desire to have A happen and thus has a motive for making A happen. His motive for making A happen, I suggest, is one that *necessitates* provided that, because of the motive, (1) is false; he cannot resist the temptation to do something in order to make A happen. His motive for making A happen is one that *inclines* provided that, because of the motive, (2) is false; like our public official, he cannot bring himself to do anything to prevent A from happening. And therefore we can say that this motive for making A happen is one that *inclines but does not necessitate* provided that, because of the motive, (1) is true and (2) is false; he can resist the temptation to make it happen but he cannot resist the temptation to allow it to happen.

Notes

- 1 The general position to be presented here is suggested in the following writings, among others: Aristotle, *Eudemian Ethics*, book II, ch. 6; *Nicomachean Ethics*, book III, chs 1–5; Thomas Reid, *Essays on the Active Powers of Man*; C. A. Campbell, "Is 'Free Will' a Pseudo-Problem?" *Mind*, n.s. 60 (1951), pp. 441–65; Roderick M. Chisholm, "Responsibility and Avoidability," and Richard Taylor, "Determination and the Theory of Agency," in Sidney Hook, ed., *Determinism and Freedom in the*

- Age of Modern Science* (New York: New York University Press, 1958).
- 2 Thomas Reid, *Essays on the Active Powers of the Human Mind* (Cambridge, Mass.: MIT Press, 1969; first published 1788), p. 261.
 - 3 *Summa Theologiae*, First Part of the Second Part, Question VI: “On the Voluntary and Involuntary.”
 - 4 Jonathan Edwards, *Freedom of the Will* (New Haven, Conn.: Yale University Press, 1957); G. E. Moore, *Ethics* (Home University Library, 1912), ch. 6.
 - 5 A. I. Melden, *Free Action* (Oxford: Blackwell, 1961), especially ch. 3. Mr Melden’s own views, however, are quite the contrary of those proposed here.
 - 6 Aristotle, *Physics*, book III, ch. 3; Suarez, *Disputationes Metaphysicae*, Disputation 18, Section 10.
 - 7 Reid, *Essays on the Active Powers*, p. 39.
 - 8 John Locke, *Essay Concerning Human Understanding*, book II, ch. 21.
 - 9 Melden, *Free Action*, p. 166.
 - 10 Reid, *Essays on the Active Powers*, p. 291.
 - 11 In the preface to the *Metaphysical Elements of Ethics*, in T. K. Abbot, ed., *Kant’s Critique of Practical Reason and Other Works on the Theory of Ethics* (London: Longman’s Green, 1959), p. 303.
 - 12 Cf. D. P. Henry, “Saint Anselm’s *De ‘Grammatico’*,” *Philosophical Quarterly*, 10 (1960); pp. 115–26.
-

41 The Mystery of Metaphysical Freedom

Peter van Inwagen

There are many kinds of freedom – or, as I prefer to say, the word “freedom” has many senses. In one sense of the word, an agent is “free” to the extent that his actions are not subject to control by the state. It is, however, obvious that an agent may be free in this sense but unfree in other senses. However little the state may interfere with my actions, I may be unfree because I am paralyzed from the waist down or because I am subject to a neurotic fear of open spaces that makes it impossible for me to venture out of doors or because I am so poor that I am unable to afford the necessary means to what I want to do. These examples suggest that freedom is a merely negative concept – that freedom is freedom from constraint, that freedom consists in the mere absence of constraint. If freedom is in this sense a negative concept, this explains why there are many kinds of freedom: there are many kinds of freedom because there are many kinds of constraint. Because there are political constraints, there is political freedom, which exists in their absence; because there are internal psychological constraints (such as neurosis), there is psychological freedom, which exists in their absence; because there are economic constraints, there is economic freedom, which exists in their absence – and so on.

When we turn from politics and psychology and economics to metaphysics, however, we encounter discussions of freedom – discussions involving words like “freedom,” “free,” and “freely” – that it is hard to account for if freedom is no more than a negative concept. Consider, for example, the following words of Holbach:

Man's life is a line that nature commands him to describe upon the surface of the earth, without his ever being able to swerve from it, even for an instant. . . . Nevertheless, in spite of the shackles by which he is bound, it is pretended he is a free agent. . . .

Or consider the ancient problem of future contingents, which would seem to depend on considerations different from those adduced by Holbach, for it has only to do with whether statements about future events must be either true or false, and has nothing to do with causation and physical law. Consider, again, the problem of divine knowledge of future human action. Consider, finally, the problem of evil and the attempts to solve that problem that appeal to the freedom of creatures and the alleged impossibility of a free creature that is certain to do no evil.

I think it is fairly evident that the concept of freedom that figures in the discussions raised by these metaphysical problems is the same concept. I think it is not easy to see how this concept could be understood as a merely negative concept, as a concept that applies to any agent just in the case that that agent's acts are not subject to some sort of constraint.

Consider, for example, the problem of free will and determinism, the problem that is raised by the above quotation from Holbach. Although my present actions may be determined by the laws of nature and the state of the world before my birth (indeed, millions of years ago), it does not follow that this state of affairs places me under any sort of constraint. A constraint on one's behavior is an impediment to the exercise of one's will. If the state places me in chains, then my will to be elsewhere, if I attempt to exercise it, will soon come into conflict with the length and solidity of my chain. If I am an extreme agoraphobe, then my will to go about the ordinary business of life will come into conflict with sensations of panic and dislocation the moment I step out of doors. If I am very poor, my will to own a warm overcoat will come into conflict with my lack of the price of the coat. It is things of these sorts that are meant by “constraint.” And it is evident that determinism places me under no constraints. It is true that in a deterministic world, *what my will is on a given occasion* will be a consequence of the way the world was millions of years ago and the laws of nature. It is true that in a deterministic world, *whether my will happens to encounter an obstacle on a given occasion* will be a consequence of the way the world was millions of years ago and the laws of nature. But it is certainly not inevitable that my will encounter an obstacle on any given occasion in a deterministic world, and even in an indeterministic world, my will must encounter obstacles on many occasions. Indeed, there is no reason to suppose that my will will encounter obstacles more frequently in a deterministic world than in an indeterministic

world. Anyone who believes that freedom is a negative concept will therefore conclude that the so-called problem of free will and determinism is founded on confusion. (So Hobbes, Hume, Mill, and many other philosophers have concluded.)

The situation is similar with the problem of divine knowledge of future human actions. We are often told that there really is no problem about this, since the fact that God knows that one is going to tell a lie (for example) in no way forces one to lie. Since God's knowledge does not interfere with the exercise of one's will, since the false words that issue from one's mouth are the words that it was one's will to speak, God's knowledge that one was going to lie is consistent with the lie's being a free act.

All this can sound very sensible. And yet one is left with the feeling that the freedom this leaves us with is, in Kant's words, a "wretched subterfuge." This feeling can be embodied in an argument. The argument is, to my mind, a rather powerful one. If the argument is correct, then freedom is not a merely negative concept. Or, at any rate, there is a concept of freedom that is not a merely negative concept, and this concept is a very important one. It is this concept, I believe, that figures in the metaphysical problems I have cited. I will call it metaphysical freedom. In calling it metaphysical freedom, however, I do not mean to imply that it is of interest only to the metaphysician. I believe that this concept is also of importance in everyday life, and that the concept that metaphysicians employ is just this everyday concept, or perhaps a refinement of it. (I should be willing to argue that all concepts that we employ in philosophy or science or any other area of inquiry are either everyday concepts or explicable in terms of everyday concepts.)

In ordinary English, the concept of metaphysical freedom finds its primary expression in simple, common words and phrases, and not in the grand, abstract terms of philosophical art that one is apt to associate with metaphysics. (The situation is similar in French, German, and Latin. I should be surprised to learn of a language in which the concept I am calling "metaphysical freedom" could not be expressed in simple, common words and phrases.) It is true that philosophical analysis is needed to distinguish those uses of these simple words and phrases on which they express this concept from other uses on which they express other concepts. Nevertheless, in particular concrete contexts, these simple words express that very concept of freedom (not, as we shall see, a negative concept) that figures in metaphysical problems like the problem of freedom and determinism. But perhaps the meaning of these abstract remarks will not be clear without an example.

One of the simple words that expresses the concept of metaphysical freedom in English is "can." What are we asking when we ask whether I am free to tell the truth tomorrow if it has been determined by events in the remote past and the laws of nature that when, tomorrow, I confront a choice between lying and telling the truth, I shall lie? Only this: "I am free to tell the truth" means "I *can* tell the truth," and "I am not free to tell the truth" means "I *cannot* tell the truth." Metaphysical freedom, therefore, is simply what is expressed by "can." If we accept this thesis, however, we must take care to understand it properly.

world. Anyone who believes that freedom is a negative concept will therefore conclude that the so-called problem of free will and determinism is founded on confusion. (So Hobbes, Hume, Mill, and many other philosophers have concluded.)

The situation is similar with the problem of divine knowledge of future human actions. We are often told that there really is no problem about this, since the fact that God knows that one is going to tell a lie (for example) in no way forces one to lie. Since God's knowledge does not interfere with the exercise of one's will, since the false words that issue from one's mouth are the words that it was one's will to speak, God's knowledge that one was going to lie is consistent with the lie's being a free act.

All this can sound very sensible. And yet one is left with the feeling that the freedom this leaves us with is, in Kant's words, a "wretched subterfuge." This feeling can be embodied in an argument. The argument is, to my mind, a rather powerful one. If the argument is correct, then freedom is not a merely negative concept. Or, at any rate, there is *a* concept of freedom that is not a merely negative concept, and this concept is a very important one. It is this concept, I believe, that figures in the metaphysical problems I have cited. I will call it metaphysical freedom. In calling it metaphysical freedom, however, I do not mean to imply that it is of interest only to the metaphysician. I believe that this concept is also of importance in everyday life, and that the concept that metaphysicians employ is just this everyday concept, or perhaps a refinement of it. (I should be willing to argue that all concepts that we employ in philosophy or science or any other area of inquiry are either everyday concepts or explicable in terms of everyday concepts.)

In ordinary English, the concept of metaphysical freedom finds its primary expression in simple, common words and phrases, and not in the grand, abstract terms of philosophical art that one is apt to associate with metaphysics. (The situation is similar in French, German, and Latin. I should be surprised to learn of a language in which the concept I am calling "metaphysical freedom" could not be expressed in simple, common words and phrases.) It is true that philosophical analysis is needed to distinguish those uses of these simple words and phrases on which they express this concept from other uses on which they express other concepts. Nevertheless, in particular concrete contexts, these simple words express that very concept of freedom (not, as we shall see, a negative concept) that figures in metaphysical problems like the problem of freedom and determinism. But perhaps the meaning of these abstract remarks will not be clear without an example.

One of the simple words that expresses the concept of metaphysical freedom in English is "can." What are we asking when we ask whether I am free to tell the truth tomorrow if it has been determined by events in the remote past and the laws of nature that when, tomorrow, I confront a choice between lying and telling the truth, I shall lie? Only this: "I am free to tell the truth" means "I *can* tell the truth," and "I am not free to tell the truth" means "I *cannot* tell the truth." Metaphysical freedom, therefore, is simply what is expressed by "can." If we accept this thesis, however, we must take care to understand it properly.

We must take care to avoid two possible sources of confusion: the ambiguity of the word “can” and false philosophical theories about what is expressed by certain sentences in which it occurs.

As to the first point, the word “can” is extremely versatile, and can be used to express many ideas other than the idea of metaphysical freedom (a fact illustrated by this sentence). One example must suffice. In negative constructions, “can” sometimes expresses an idea that might be called “moral impossibility.” One might say to a hard-hearted son, “You can’t refuse to take your own mother into your house” – even though one knows perfectly well that in the sense of “can” we have been discussing he certainly *can* refuse to take his own mother into his house because he has already done so. We must take care that if we propose to use the simple word “can” as our means to an understanding of metaphysical freedom, we do not allow our understanding of metaphysical freedom to be influenced by any of the many other concepts this simple word can be used to express. The best way to avoid such influence is not to rely on the word “can” alone in our attempt to understand metaphysical freedom, but to examine also as many as possible of the other simple, ordinary words and phrases that can be used to express the concept of metaphysical freedom (or unfreedom). To illustrate what I mean, here are three sentences in which idioms of ordinary speech that do not involve “can” are used to express the concepts of metaphysical freedom and unfreedom:

- He will *be able* to be there in time for the meeting.
- You must not blame her for missing the meeting; she *had no choice* about that.
- It was simply *not within my power* to attend the meeting.

(Oddly enough, the phrase “of his own free will” does not express the concept of metaphysical freedom, despite the fact that “free will,” as a philosophical term of art, means just exactly what I mean by “metaphysical freedom”. To say that someone attended a meeting of his own free will is simply to say that no one forced him to attend the meeting. The phrase “of his own free will” thus expresses a merely negative concept, the concept of the absence of coercion.)

False theories about the meanings of philosophically important words and phrases abound, and the philosophically important word “can” is no exception to this generalization. There are those who, recognizing the importance of idioms like “I can do X” for the metaphysical problems of freedom, have simply insisted that this word means something that supports their favorite philosophical theories. An example of such a theory would be: “I can do X” means “There exists no impediment, obstacle, or barrier to my doing X; nothing prevents my doing X.” I will not argue specifically for the conclusion that this theory is false; the argument I will later present for the incompatibility of metaphysical freedom and determinism, however, will have the consequence that this theory about the meaning of “I can” is false – since, if the theory were true, metaphysical freedom would be compatible with determinism. At this point, I wish merely to call attention to the fact that there do exist tendentious theories about the meaning of “I can do X.”

If we consider carefully the meaning of “I can do X” (“I am able to do X”; “It is within my power to do X”) do we find that the idea expressed by this form of words is a merely negative one, the idea of the absence of some constraint or barrier or obstacle to action? It would seem not. It is true that the presence of an obstacle to the performance of an action can be sufficient for one’s being unable to perform that action. But it does not follow that the absence of all obstacles to the performance of an action is sufficient for one’s being *able* to perform that action. And the idea that ability could consist in the absence of obstacles does seem, on consideration, to be a very puzzling idea indeed. To see this, let us examine carefully the relation between the concept of ability and the concept of an obstacle. We should note that not just any obstacle to one’s performance of an action is such that its presence renders one unable to perform that action – for some obstacles can be surmounted or eliminated or bypassed (in short: some obstacles can be overcome). Let us ask a simple question: *which* obstacles to the performance of an action are such that their presence renders one unable to perform that action? Why, just those obstacles that one is *unable* to overcome, of course. And it seems fairly obvious that the concept of an obstacle that one is unable to overcome cannot be analyzed or explained in terms of the concept of an obstacle *simpliciter*. (Is the concept of an obstacle that one cannot overcome the concept of an obstacle such that there is some “decisive” obstacle to one’s overcoming it? – No, not unless a “decisive” obstacle is understood as an obstacle that one is unable to overcome. . . .) These reflections suggest very strongly that the concept expressed by the words “I can do X” or “I am able to do X” cannot be a merely negative concept, the concept of the absence of some sort of obstacle or barrier or impediment to action. But let us turn now to the question of the compatibility of determinism and metaphysical freedom. I shall present an argument for the conclusion that determinism is incompatible with metaphysical freedom. Since, as we have seen, determinism and metaphysical freedom are compatible if metaphysical freedom (the concept expressed by “I can do X”) is a merely negative concept, this argument will be in effect an argument for the conclusion that metaphysical freedom is not a merely negative concept.

As Carl Ginet has said, our freedom can only be the freedom to add to the actual past – for the past is unalterable; it is what we *find ourselves with* in any situation in which we are contemplating some course of action. (Or to put this point in the terms I have been recommending, all we *can* do, all we are *able to do*, is add to the actual past.) And, unless we are bona fide miracle workers, we can make only such additions to the actual past as conform to the laws of nature. But the only additions to the actual past that conform to a deterministic set of laws are the additions that are actually made, the additions that collectively make up the actual present and the actual future. This is simply a statement of what is meant by determinism, which is the thesis that the laws of nature and the past together determine a unique future. Therefore, if the laws of nature are deterministic, we are free to do only what we in fact do – that is, we are unable to act otherwise than we do and are ipso facto not free in the sense in which the term “free” is properly used in metaphysics.

This little argument has great persuasive power, and it is probably no more than an articulation of the reasons that lead, almost without exception, the undergraduates to whom I lecture to join Kant in regarding the merely negative freedom of Hobbes and Hume as a wretched subterfuge. If the argument is correct, as I have said, it refutes the idea that metaphysical freedom is a merely negative concept, for the past and the laws of nature are not impediments to the exercise of one's will. But, more generally, we may well ask what we are to say of this argument and its consequences, for these consequences go far beyond establishing that metaphysical freedom is not a negative concept. One possible reaction to the argument would be to say, with Holbach, that, because determinism is true, we therefore do not possess metaphysical freedom. (An epistemologically more modest reaction would be to say that, because we do not know whether determinism is true, we do not know whether we possess metaphysical freedom.) I shall return to the possibility that we lack freedom (or that we do not know whether we have freedom). For the moment, let us see where the argument leaves those of us who would like to say that we are free and that we know this. Many philosophers have regarded it as evident that we are free, and have accepted something like our argument for the incompatibility of determinism and metaphysical freedom. These philosophers, therefore, have denied that the world is deterministic, have denied that the laws of nature and the past together determine a unique future.

These philosophers (among whom I count myself) face a difficult problem. They assert or postulate that the laws of nature are indeterministic. One might ask how they know this, or what gives them the right to this postulate. These are good questions, but I will not consider them. I want to consider instead another question that these philosophers must answer: does postulating or asserting that the laws of nature are indeterministic provide any comfort to those who would like to believe in metaphysical freedom? If the laws are indeterministic, then more than one future is indeed consistent with those laws and the actual past and present – but how can anyone have any choice about which of these futures becomes actual? Isn't it just a matter of chance which becomes actual? If God were to "return" an indeterministic world to precisely its state at some time in the past, and then let the world go forward again, things might indeed happen differently the "second" time. But then, if the world is indeterministic, isn't it just a matter of chance how things *did* happen in the one, actual course of events? And if what we do is just a matter of chance – well, who would want to call that freedom?

It seems, therefore, that, in addition to our argument for the incompatibility of metaphysical freedom and determinism, we have an argument for the incompatibility of metaphysical freedom and *indeterminism*. But the world must be either deterministic or indeterministic. It follows that, unless one of the two arguments contains some logical error or proceeds from a false premise, metaphysical freedom must be a contradiction in terms, as much an impossibility as a round square or a liquid wine bottle. We may in fact *define* the problem of metaphysical freedom as the problem of discovering whether either of the two arguments is defective, and (if so) of locating the defect or defects.

The problem of metaphysical freedom, so conceived, is a very *abstract* problem. Although, for historical reasons, it is natural to think of the problem as essentially involving reference to the physical world and its supposedly intransigent laws (“man’s life is a line that nature commands him to describe on the surface of the earth . . .”), it does not. For suppose that man’s life is in fact *not* a line that nature commands him to describe on the surface of the earth. Suppose that nature presents us with two or seventeen or ten thousand lines inscribed on the surface of the earth, and says to us (in effect), “Choose whichever one of them you like.” How could it be that we really had any choice about which “line” we followed, when any deliberations we might undertake would themselves have to be segments of the lines that nature has offered us? Imagine that two of the lines that nature offers me diverge at some point – that is, imagine that the lines present the aspect of a fork in a road or a river. The common part of the two lines, the segment that immediately precedes their divergence, represents the course of my deliberations; their divergence from a common origin represents diagrammatically the fact that *either* of two futures is a possible outcome of my deliberations. My deliberations, therefore, do not determine which future I shall choose. But then what *does* determine which future I shall choose? Only chance, it would seem, and if only chance determines which of two paths into the future I follow, then how can it be that I have a choice about which of them I follow?

The problem of metaphysical freedom is so abstract, so very nearly independent of the features of the world in which agents happen to find themselves, that it could – it would; it must – arise in essentially the same form in a world inhabited only by immaterial intelligences, a world whose only inhabitants were, let us say, angels.

Let us consider such a world. It is true that if there were only angels, there would be no physical laws – or at any rate there would be nothing for the laws to apply to, so we might as well say there would be none. But if we assume the angels make choices, we have to assume that time (somehow) exists in this non-physical world, and that the agents are in different “states” at different times. And what is responsible for the way an angel changes its states with the passage of time? One possibility is that it is something structurally analogous to the laws of physics – something that stands to angels as our laws of physics stand to electrons and quarks. (I’m assuming, by the way, that these angels are metaphysical simples, that they are not composed of smaller immaterial things. If they were, we could conduct the argument in terms of the smallest immaterial things, the “elementary particles” of this imaginary immaterial world.) This “something” takes the properties of the angels at any time (and the relations they bear to one another at that time: the analogue, whatever it may be, of spatial relations in a material world) as “input,” and delivers as output a sheaf of possible futures and histories of the world. In other words, given the “state of the world” at any time, it tells you what temporal sequences of states could have preceded the world’s being in that state at that time, and it tells you what temporal sequences of states could follow the world’s being in that state at that time. Maybe it couldn’t be written as a set of differential equations (since noth-

ing I have said implies that the properties of and relations among angels are quantifiable) as the laws of our physical world presumably can, but I don't think that affects the point. And the point is: either "the sheaf of possible futures" relative to each moment has only one member or it has more than one. If it has only one, the world of angels is deterministic. And then where is their free will? (Their freedom is the freedom to add to the actual past. And they can only add to the actual past in accordance with the laws that govern the way angels change their properties and their relations to one another with time.) If it has more than one, then the fact that one possible future rather than another, equally possible, future becomes actual seems to be simply a matter of chance. And then where is their free will?

I said above that this way of looking at a postulated "world of angels" was one possibility. But are there really any others? We have to think of the angels as being temporal and as changing their properties with the passage of time if we are to think of them as making choices. And we have to think of them as bearing various relations to one another if we are to think of them as belonging to the same world. And we have to think of them as having natures if we are to think of them as being real things. Every real thing that is in time must have a nature that puts some kinds of constraints on how it can change its states with the passage of time. Or so, at any rate, it seems to me. But if we grant this much, it seems that, insofar as we can imagine a world of non-physical things (angels or any others) we must imagine the inhabitants of this world as being subject to something analogous to the laws of physics. If this "something" is deterministic, then (it seems) we can't think of the inhabitants of our imaginary world as having free will. And if this "something" is *indeterministic*, then (it seems) we can't think of the inhabitants of our imaginary world as having free will. Thus, the "problem of metaphysical freedom" is a problem so abstract and general that it arises in any imaginable world in which there are beings who make choices. The problem, in fact, arises in exactly the same way in relation to God. God, the theologians tell us, although He did in fact create a world, was free not to. (That is, He was *able* not to create a world.) But God has His own nature, which even He cannot violate and cannot change. (He cannot, for example, make Himself less than omnipotent; He cannot break a promise He has made; He cannot command immoral behavior.) And either this nature determines that He shall create a world or it does not. If it does, He was not free not to create. If it does not, then, it would seem, the fact that He *did* create a world was merely a matter of chance. For what, other than chance, could be responsible for the fact that He created a world? His choice or His will? But what determined that he should make *that* choice when the choice not to make a world was also consistent with His nature? What determined that His will should be set on making a world, when a will set on *not* making a world was also consistent with His nature? We should not be surprised that our dilemma concerning metaphysical freedom applies even to God, for the dilemma does not depend on the nature of the agent to whom the concept of metaphysical freedom is applied. The dilemma arises from the concept of metaphysical freedom itself, and its conclusion is that metaphysical freedom is a contradictory concept. And

a contradictory concept can no more apply to God than it can apply to anything else.

The concept of metaphysical freedom seems, then, to be contradictory. One way to react to the seeming contradiction in this concept would be to conclude that it was real: metaphysical freedom seems contradictory because it *is* contradictory. (This was the conclusion reached by C. D. Broad.)

But none of us really believes this. A philosopher may argue that consciousness does not exist or that knowledge is impossible or that there is no right or wrong. But no one really believes that he himself is not conscious or that no one knows whether there is such a city as Warsaw; and only interested parties believe that there is nothing morally objectionable about child brothels or slavery or the employment of poison gas against civilians. And everyone really believes in metaphysical freedom, whether or not he would call it by that name. Dr Johnson famously said, "Sir, we know our will's free, and there's an end on't." Perhaps he was wrong, but he was saying something we all believe. Whether or not we are all, as the existentialists said, condemned to freedom, we are certainly all condemned to *believe in* freedom – and, in fact, condemned to believe that we *know* that we are free. (I am not disputing the sincerity of those philosophers who, like Holbach, have denied in their writings the reality of metaphysical freedom. I am saying rather that their beliefs are contradictory. Perhaps, as they say, they believe that there is no freedom – but, being human beings, they also believe that there is. In my book on freedom, I compared them to the Japanese astronomer who was said to have believed, in the 1930s, that the sun was an astronomically distant ball of hot gas vastly larger than the earth, and also to have believed that the sun was the ancestress of the Japanese imperial dynasty.)

I would ask you to try a simple experiment. Consider some important choice that confronts you. You must, perhaps, decide whether to marry a certain person, or whether to undergo a dangerous but promising course of medical treatment, or whether to report to a superior a colleague you suspect of embezzling money. (Tailor the example to your own life.) Consider the two courses of action that confront you; since I don't know what you have chosen, I'll call them simply A and B. Do you really not believe that you are *able* to do A and *able* to do B? If you do not, then how can it be that you are trying to decide which of them to do? It seems clear to me that when *I* am trying to decide which of two things to do, I commit myself, by the very act of attempting to decide between the two, to the thesis that I am able to do each of them. If I am trying to decide whether to report my colleague, then, by the very act of trying to reach a decision about this matter, I commit myself both to the thesis that I am able to report him and to the thesis that I am able to refrain from reporting him: although I obviously cannot do *both* these things, I can (I believe) do *either*. In sum: whether we are free or not, we believe that we are – and I think we believe, too, that we *know* this. We believe that we know this even if, like Holbach, we *also* believe that we are not free, and, therefore, that we do not know that we are free.

But if we know that we are free – indeed, if we are free and do not know it –

there is some defect in one or both of our two arguments. Either there is something wrong with our argument for the conclusion that metaphysical freedom is incompatible with determinism or there is something wrong with our argument for the conclusion that metaphysical freedom is incompatible with *in*determinism – or there is something wrong with both arguments. But which argument is wrong, and why? (Or are they both wrong?) I do not know. I think no one knows. That is why my title is, “The *Mystery* of Metaphysical Freedom.” I believe I know, as surely as I know anything, that at least one of the two arguments contains a mistake. And yet, having thought very hard about the two arguments for almost thirty years, I confess myself unable to identify even a possible candidate for such a mistake. My *opinion* is that the first argument (the argument for the incompatibility of freedom and determinism) is essentially sound, and that there is, therefore, something wrong with the second argument (the argument for the incompatibility of freedom and indeterminism). But if you ask me *what* it is, I have to say that I am, as current American slang has it, absolutely clueless. Indeed the problem seems to me to be so evidently impossible of solution that I find very attractive a suggestion that has been made by Noam Chomsky (and which was developed by Colin McGinn in his recent book *The Problems of Philosophy*) that there is something about our biology, something about the ways of thinking that are “hardwired” into our brains, that renders it impossible for us human beings to dispel the mystery of metaphysical freedom. However this may be, I am certain that I cannot dispel the mystery, and I am certain that no one else has in fact done so.

42 The Agent as Cause

Timothy O'Connor

In the previous essay, Peter van Inwagen argues that “metaphysical freedom” is incompatible with a certain abstract picture of the world (commonly dubbed “determinism”), on which it evolves in strict accordance with physical laws, laws such that the state of the world at any given time ensures a unique outcome at any subsequent moment. I agree that the two are incompatible. But what, in positive terms, does the ordinary understanding of ourselves as intelligent beings who “freely” decide how we shall act require? Where do the “springs of action” lie for beings that truly enjoy “free will”? This is surprisingly difficult to answer with any confidence. A useful way of approaching this question is to consider the various ways we might modify determinism in order to accommodate free will.

The most economical change in the determinist’s basic picture is to introduce a causal “loose fit” between those factors influencing my choice (such as

my beliefs and desires) and the choice itself. We might suppose, that is, that such factors *cause* my choice in an *indeterministic* way. To say that the causation involved is “*indeterministic*” is perhaps to say that the laws governing the evolution of the world through time (including that bit of the world which is me) are fundamentally statistical: they allow that (at least at various junctures) a range of alternatives are possible, though they will specify that certain of them are far more likely than others, in accordance with some measure of probability. Applying this general idea to the case of human choices, one might suppose that a free choice requires the following features: I have reasons to act in accordance with each of a range of options. In each case, my having those reasons gives me an objective (probabilistic) tendency to act accordingly. But whatever the relative probabilities of the alternatives, each of them is possible. And whichever of them occurs, the agent’s having had a specific reason so to act will have been among the factors that caused it. Let us call this modification of the deterministic picture “causal indeterminism.”

Would this be freedom? In my judgment, it would not. It is not enough that any of a range of possible actions are *open* to me to perform. I must have the right sort of *control* over the way the decision goes in a given case. And we may ask of the causal indeterminist, how is it up to me that, on this occasion, this one among two or more causally possible choices was made? I find myself with competing motivations – in my present case, a desire to watch a basketball game, a desire to play a game with my children, and a desire to finish this article – each of a particular “strength.” On this occasion, we may suppose, the least probable outcome occurs. On other occasions, more probable outcomes occur. If I am truly acting freely, then presumably I in some way directly control or determine which outcome occurs on a given occasion. But in what does that control consist? The causal indeterminist does not have resources, it seems, to satisfactorily answer this question. Given a sufficiently large number of choices of a large number of people, the pattern of outcomes is likely to conform, more or less, to the statistical character of the underlying laws. There seems nothing more that one can say – in particular, nothing more one can say about the outcome of any particular choice. The indeterministic tendencies arising from my reasons confer a *kind* of control that is too “chancy” to ground significant responsibility. Indeed, it does not differ at all in *kind* from the control that would be had in a deterministic world; it merely introduces an element of “looseness” into its exercise. Given this added looseness, the future *is open* to alternative possibilities. But it remains unclear how I myself could be responsible (in part) for which of those alternatives is realized.

A dilemma is forming. Responsibility for our actions is inconsistent with the deterministic picture of the world. But it is also inconsistent with at least one straightforward kind of indeterministic picture, the kind that most directly carries into the sphere of human action the sort of indeterminism that many theorists believe operates at the level of fundamental physics. Indeed, a good many philosophers suppose that these two pictures (which we have labeled “causal determinism” and “causal indeterminism”) exhaust the plausible alternatives. If all this is right, then the conclusion to be drawn is that free will is simply an

inconsistent notion. It's not that we just don't happen to have free will; rather, we don't have it because it simply can't be had.

One alternative to this unpalatable conclusion is that entertained by Peter van Inwagen, in his contribution to this volume. Perhaps, van Inwagen writes, "there is something about our biology, something about the ways of thinking that are 'hardwired' into our brains, that renders it impossible for us human beings to dispel the mystery of metaphysical freedom" (see p. 374). That is, though the notion of free will isn't truly inconsistent, its nature is "cognitively closed" to us. (After all, we have no reason to be confident that we are able, even "in principle," to grasp *every* difficult notion that, say, God grasps. And the history of philosophical reflection on the idea of freedom of will suggests that it has its subtleties.)

Well, there is certainly no *arguing* against this suggestion, absent the emergence of a stable consensus of opinion on the matter – rather unlikely at this stage of the game. But one may well distrust it on the general grounds that it counsels complacency. (And why stop at the notion of free will? Philosophers disagree over the correct understanding of most significant philosophical concepts.) Furthermore, once a philosopher takes this suggestion seriously, he may well be drawn into a deeper measure of skepticism about the notion of freedom of will than initially intended. Van Inwagen, for example, tells us that he is of the opinion that free will is *incompatible* with determinism. So he supposes that it must be *compatible* with *indeterminism*, even though he fails to see *which* sort of indeterminism will clearly do the trick. But if he and the rest of us are "hardwired" in some manner that precludes our coming to understand adequately the nature of free will, is it likely that we understand it sufficiently to know even *some* of its features? At any rate, the hypothesis ought to automatically undercut one's confidence in any highly *disputed* claims, such as van Inwagen's relative confidence in the thesis that free will is incompatible with determinism. (I note that Colin McGinn, whom van Inwagen cites in this connection, supposes that free will *can* be had under determinism, even though he "can't see how".)

Rather than embrace the despair and skepticism of the "cognitive closure" hypothesis, then, let us pick up the argument where we last left it, and see whether a "positive" solution to our problem is in the offing. I argued that, if my decisions to act are simply the indeterministic effects of my beliefs and desires, then they are not up to me. What more do we *want* to say about our decisions, that causal indeterminism leaves out?

Just this, it seems: that I myself freely and directly control the outcome, where "control" here (as everywhere) is evidently a *causal* notion. And the unsatisfactoriness of causal indeterminism suggests that we have to be rather literal about the referent of "I," in this context. If I do something freely, I cannot be thought of as simply an arena in which internal and external factors work together to bring about my action (whether or not these factors are thought to operate in a strictly deterministic fashion). Instead, we want to say with Roderick Chisholm that I am the "end-of-the-line" initiator of the resulting action. What we are after, that is, is a notion of a distinctively personal form of

causality (in the parlance of philosophers, “agent causation”), as against the broadly mechanistic form of causality (“event causation”) that both the deterministic and causal indeterministic pictures represent as governing *all* forms of activity in nature without exception.

Many philosophers find this notion of “personal” or “agent” causation to be utterly mysterious, or downright incoherent. (Some of those philosophers will agree that it is natural to talk of “agent causation” when trying to articulate an understanding of free will, even though it is an incoherent idea. On their view, the term encapsulates the inconsistent strands in that notion.) Here is a simple reflection that fosters the sense of mystery. We often talk loosely of inanimate objects as causing certain things to happen. An example is the statement that Zimmerman’s car knocked down the telephone pole. But it’s clear that this does not perspicuously capture the metaphysics of the situation. It is instead simply shorthand for the assertion that the *movement* of Zimmerman’s car (a car with a certain mass) caused the pole’s falling down. It is, then, this *event* involving Zimmerman’s car that brought about the effect, and not simply the car, *qua* enduring object. (No such effects emanate from his car when Zimmerman wisely decides to keep it parked in his garage.) The problem that many see with agent causation is that it rejects any expansion of “loose” talk of agents’ causing things to happen into statements asserting that particular events *involving* those agents cause the effects in question. And that can seem mysterious: how can agents cause things to happen without its being true that they do so in virtue of certain features of themselves at the time? The agent is, after all, always an agent; yet he is not always causing some particular effect, such as deciding to complete an article on agent causation. Doesn’t this force us to acknowledge that if the agent has decided to complete that article at one particular time, there must have been something *about him* at that time in virtue of which that effect was realized? And isn’t that just to say that the *event* of the agent’s having those distinguishing features, whatever they were, is what caused the decision?

This simple reflection is perhaps the deepest basis for philosophical suspicion about the notion of agent causation. However, I have come to suspect the suspicion and its various bases. In order to have a clear view of this matter, we need to reflect further on what is involved in our ordinary understanding of causation. Unfortunately, there is precious little agreement among philosophers about these matters. But the brief remarks I will make on this score at least have the advantage of representing a fairly commonsensical view of causation.

On the theory of causation I favor, objects are inherently active or dynamic. That is, they have causal capacities, and these are not “free-floating”, but rather are linked to their intrinsic properties – those basic properties whose exact character it is the business of science to investigate.

In the more generally applicable case of *event* (or broadly mechanistic) causation, the *exercise* of such a capacity or tendency proceeds “as a matter of course”: a thing’s having, in the right circumstances, the capacity-grounding cluster of properties directly generates one of the effects within its range. (For indeterministic capacities, that effect will be but one of a range of *possible* ef-

fects; whereas in the deterministic case, there is only one possible outcome.)

The way that agent or personal causation differs from this mechanistic paradigm is in the way the relevant causal capacities are *exercised*. An agent's capacity to freely and directly control the outcome of his deliberation also requires underlying intrinsic properties which ground that capacity. (What sort of properties these might be is an interesting and in certain respects puzzling question, but it is at least partly empirical and not conceptual in nature. In any case, I shall not consider it here.) And no doubt the range of its operation is sharply circumscribed. For what is it, after all, that I directly act on, according to the agency theory? Myself – a complex system regulated by a host of stratified dynamic processes. I don't introduce events *ex nihilo*; (at best) I influence the direction of what is already going on within me. What is going on is a structured, dynamic situation open to some possibilities and not others. So the capacity is also circumscribed by physical and psychological factors at work within the agent while he deliberates. But (and here is the difference from the mechanistic paradigm) having the properties that subserve an *agent-causal* capacity does not suffice to bring about a particular effect (or even the occurrence of some effect or other within a range of possible effects); rather, it *enables* the agent to determine an effect (within the corresponding range). Whether, when, and how such a capacity will be exercised is freely determined by the agent.

That is the core metaphysical difference between the two causal paradigms. But we have yet to discuss how prior desires, intentions, and beliefs (more simply, “reasons”) may explain such agent-causal activity. I suggest that we think of the agent's immediate effect as an action-triggering state of *intention* (which endures throughout the action and guides its completion). The content of that intention, in part, is that I act here and now in a particular sort of way. But another aspect of that intention, in my view, is that an action of a specific sort be performed *for certain reasons* the agent had at the time. (After a brief deliberation, I formed the intention to continue to type these words *in order to get the editors of this volume off my back*.) And the basis of the explanatory link lies precisely in this fact that the intention refers to the guiding reason. That is, the caused intention bears its explanation on its sleeve, so to speak. Had the agent generated a different intention, it would have been done (in most cases) for a different reason, to which reason the content of the intention itself would have referred. And if the agent had *several* reasons for performing a particular action, the reason(s) that *actually* moved the agent to act, again, would be reflected in the content of the intention. (None of this is to suggest that determining this content, in retrospect, is always easy. Clearly, I can be mistaken about my own reasons for acting.)

Some say that this account of the explanatory nature of reasons cannot be right: we can simply see that any undetermined instance of agent causation would be random, since by hypothesis nothing causes it. (Even some proponents of agent causation have been worried about this, and have been led to posit infinite hierarchies of agent-causings.) But it is hard to credit this objection. Consider what is being demanded. Agent causation is a form of direct control over one's behavior *par excellence*. But this is held to be insufficient.

What is needed, it is argued, is some mechanism by virtue of which the agent controls this controlling. Put thus (though understandably it is not generally put in this way), its absurdity is evident. We needn't control our exercises of control. (For if we did, then wouldn't we also need yet another exercise of control, and so on?) On any coherent conception of human action, there is going to be a *basic* form of activity on which rests all control over less immediate effects. On the agent-causal picture, this basic activity is that of an agent's directly generating an intention to act in accordance with certain reasons.

Others have argued that the suggested account of explanations of free actions by reasons cannot be right, since the reasons to which one points in a given case won't explain why the agent acted as he did *rather than* in one of the other ways that were open to him (alternatives that by hypothesis remained open up to the very moment of choice). But while the issues involved here are subtler, this objection also fails. The objection assumes that adequately explaining an occurrence *ipso facto* involves explaining why that event occurred rather than any imaginable alternative. And this seems too strong a requirement. At bottom, explaining an occurrence involves uncovering the causal factor that generated it. In deterministic cases, where only one outcome is possible, such an explanation will also show why that event occurred rather than any other. But this should not blind us to the fact that the two targets of explanation are distinct: the simple *occurrence* itself and the *contrastive fact* that the outcome occurred rather than any other alternative. We need this distinction not just to understand human free agency, but to understand any indeterministic causal activity, including the apparently indeterministic mechanisms described by physical science. Whether (and in what circumstances) there can *also* be contrastive explanations of such indeterministic outcomes is a difficult question. But whatever we say here, there is little to recommend the claim that an occurrence that has been caused, though not uniquely determined, by some factor is thereby wholly inexplicable.

More might be said about the “nature of reasons” explanation on the picture just sketched, but I want to turn instead to the complaint that we've swung too far in the direction of freedom. In place of the diminished, freedom-less conception of human action entailed by the deterministic picture, we've substituted a rather god-like one: the agent selects from among reasons that are merely passively present before the agent as he deliberates, reasons that do not *move* the agent to act. Though rather implausible on the face of it, such a consequence is embraced by some advocates of agent causation. Chisholm, for example, compares agent causation with divine action:

If we are responsible, and if what I have been trying to say is true, then we have a prerogative which some would attribute only to God: each of us, when we act, is a prime mover unmoved. In doing what we do, we cause certain events to happen, and nothing – or no one – causes us to cause those events to happen.¹

But perhaps this is unnecessarily heroic. Though defenders of agent causation have generally insisted on a sharp divide between it and mechanistic causa-

tion, we may be able to move tentatively toward greater integration of the two. The goal is not to *reduce* agent causation, in the end, to an all-encompassing mechanistic paradigm, but rather to see how event-causal factors such as the possession of reasons to act may *shape* the distinctively agent-causal capacity. Two things, in particular, seem needed here – if not for all conceivable agents (including God and angels), then at least for human beings as we know them. First, our account should capture the way reasons (in some sense) *move* us to act as we do – and not as external pressures, but as *our* reasons, as our own internal tendencies to act to satisfy certain desires or aims. Secondly, the account should acknowledge that those reasons typically do not have “equal weight,” so to speak. It is a truism that, given the structure of my preferences, stable intentions, and so forth, and the situation with which I am faced, I am often far more likely to act in one way rather than in any other. But how might we account for this, if not in terms of a relative tendency, on the part of reasons, to *produce* our actions?

In my view, this is the biggest obstacle to a clear understanding of what free will requires. What we need is a way to modify the traditional notion of a distinctively personal kind of causal capacity and to see it, not as utterly unfettered, but as one that comes “structured”, in the sense of having built-in propensities to act (though ones that shift over time in accordance with the agent’s changing preferences). But we must do so in such a way that it remains up to me to act on these tendencies or not, so that what I do is not simply the consequence of the vagaries of “chance-like” indeterministic activity, as may be true of microphysical quantum phenomena.

So, the task of harmonizing free and responsible human agency with a world that is fundamentally mechanistic in character remains unfinished. But perhaps we’ve seen enough to dispel much of the air of profound mystery that some profess to find on considering the very idea of metaphysical freedom.

Note

- 1 “Human Freedom and the Self,” p. 362, this volume.

PART THREE

IS THERE JUST ONE WORLD?

Introduction

- 43 Speaking of Objects
W. V. O. QUINE
- 44 After Metaphysics, What?
HILARY PUTNAM
- 45 Truth and Convention
HILARY PUTNAM
- 46 Nonabsolute Existence and Conceptual Relativity: an Excerpt from “Putnam’s Pragmatic Realism”
ERNEST SOSA
- 47 Addendum to “Nonabsolute Existence and Conceptual Relativity”: Objections and Replies
ERNEST SOSA

Introduction

W. V. O. Quine, Hilary Putnam, and Ernest Sosa each suggest that the question “What is there? What exists?” does not have one answer; there is not a *unique* “complete inventory of the universe” (to use C. D. Broad’s phrase – see “The Theory of Sensa”, above). Instead there is only what exists-according-to- X , and what exists-according-to- Y , and so on – where each list generated by one of these relations may be complete, including everything there is, while nonetheless leaving out items on other lists. It is not entirely clear what “ X ” and “ Y ” should stand for, however. One possibility, suggested by the excerpts from Quine and Putnam, is a language. A stock example (but probably not a very good one) goes like this: Relative to a certain Eskimo language, there are ever so many distinct kinds of snow; but relative to English there are fewer. Sosa rejects the “linguistic turn”; if there is to be real “ontological relativity,” it cannot be mere “linguistic relativity.” His exploration of a more robust sort of relativity of what there is leads him to develop a notion of “nonabsolute existence.”

Putnam claims that ontological relativity signals the death of metaphysics. But if Sosa is right about the emptiness or implausibility of linguistic relativism, then Putnam’s eulogy is premature. Substantive, nonlinguistic varieties of relativity end up being “just more metaphysics” – for the thesis that existence is not absolute, as Sosa understands it, is surely a substantive metaphysical thesis if ever there was one. (This is in keeping with our general claim about what happens when philosophers try to show that metaphysics is impossible; see section two of the “Introduction: What is Metaphysics?”, above.)

By our lights, the relativity of what there is to our “conceptual scheme” remains a rather unattractive hypothesis. Sosa puts his finger on the most serious objection to the view: How can someone who makes existence relative to the conceptual scheme of human beings (or some extension thereof) allow for the existence of things “at present unrecognized” by that scheme, things that surpass “our present acuity and acumen”? Is it not simple hubris to suppose that the concepts we puny humans can grasp provide the measure of everything there is? – to insist that there can be no “God’s eye view” of reality (with or without a God to view things from there), no view according to which there are things we can never comprehend?

Suggestions for Further Reading

- Aune, Bruce, *Metaphysics: the Elements* (Minneapolis: University of Minnesota, 1985), pp. 126–30.
- Benardete, José, *Metaphysics: the Logical Approach* (Oxford: Oxford University Press, 1989), parts 1 and 2, and ch. 23: “Anti-Realism.”
- Carter, William R., *The Elements of Metaphysics* (Philadelphia, Penn.: Temple University Press, 1990), ch. 12: “Being Realistic.”
- Gardner, Martin, *The Whys of a Philosophical Scrivener* (New York: Quill, 1983), chs 1 and 2: “The World: Why I am not a Solipsist” and “Truth: Why I am not a Pragmatist.”

METAPHYSICS: THE BIG QUESTIONS

- Goodman, Nelson, *Ways of Worldmaking* (Indianapolis, Ind.: Hackett, 1978).
- Hales, Steven D., *Metaphysics: Contemporary Readings* (Belmont, Cal.: Wadsworth, 1998), section 3: “Realism/Anti-Realism.”
- Hamlyn, D. W., *Metaphysics* (Cambridge, UK: Cambridge University Press, 1984), ch. 3: “Ontology.”
- Hasker, William, *Metaphysics: Constructing a World View* (Downers Grove, Ill., and Leicester, UK: InterVarsity Press, 1983), ch. 1: “Introducing Metaphysics.”
- Jubien, Michael, *Contemporary Metaphysics* (Oxford: Blackwell, 1997), ch. 5: “Is Truth Relative?”
- Loux, Michael, *Metaphysics: a Contemporary Introduction* (London: Routledge, 1997), Introduction.
- Plantinga, Alvin, “How to Be an Anti-Realist,” *Proceedings and Addresses of the American Philosophical Association*, vol. 56, no. 1 (Sept. 1982), pp. 47–70.
- Post, John F., *Metaphysics: a Contemporary Introduction* (New York: Paragon House, 1991), chs 2 and 3: “Language and Reality” and “Piercing the Veil of Language.”
- Putnam, Hilary, *The Many Faces of Realism* (LaSalle, Ill.: Open Court, 1987).
- Searle, John, *The Construction of Social Reality* (New York: The Free Press, 1995), chs 7–9: “Does the Real World Exist? Part I: Attacks on Realism,” “Does the Real World Exist? Part II: Could There Be a Proof of External Realism?,” and “Truth and Correspondence.”
- van Inwagen, Peter, *Metaphysics* (Boulder, Col.: Westview Press, 1993), ch. 4: “Objectivity.”

43 Speaking of Objects*

W. V. O. Quine

I

We are prone to talk and think of objects. Physical objects are the obvious illustration when the illustrative mood is on us, but there are also all the abstract objects, or so there purport to be: the states and qualities, numbers, attributes, classes. We persist in breaking reality down somehow into a multiplicity of identifiable and discriminable objects, to be referred to by singular and general terms. We talk so inveterately of objects that to say we do so seems almost to say nothing at all; for how else is there to talk?

It is hard to say how else there is to talk, not because our objectifying pattern is an invariable trait of human nature, but because we are bound to adapt any alien pattern to our own in the very process of understanding or translating the alien sentences.

Imagine a newly discovered tribe whose language is without known affinities. The linguist has to learn the language directly by observing what the natives say under observed circumstances, encountered or contrived. He makes a first crude beginning by compiling native terms for environing objects; but here already he is really imposing his own patterns. Let me explain what I mean. I will grant that the linguist may establish inductively, beyond reasonable doubt, that a certain heathen expression is one to which natives can be prompted to assent by the presence of a rabbit, or reasonable *facsimile*, and not otherwise. The linguist is then warranted in according the native expression the cautious translation "There's a rabbit," "There we have a rabbit," "Lo! a rabbit," "Lo! rabbithood again," insofar as the differences among these English sentences are counted irrelevant. This much translation can be objective, however exotic the tribe. It recognizes the native expression as in effect a rabbit-heralding sentence. But the linguist's bold further step, in which he imposes his own object-positing pattern without special warrant, is taken when he equates the native expression or any part of it with the term "rabbit."

It is easy to show that such appeal to an object category is unwarranted even though we cannot easily, in English, herald rabbits without objectification. For we can argue from indifference. Given that a native sentence says that a so-and-so is present, and given that the sentence is true when and only when a rabbit is present, it by no means follows that the so-and-so are rabbits. They might be all the various temporal segments of rabbits. They might be all the integral or

* From W. V. O. Quine, "Speaking of Objects," *Proceedings and Addresses of the American Philosophical Association*, 31 (1957-8), pp. 5-8. Reprinted by permission of the author and the American Philosophical Association.

undetached parts of rabbits. In order to decide among these alternatives we need to be able to ask more than whether a so-and-so is present. We need to be able to ask whether this is the same so-and-so as that, and whether one so-and-so is present or two. We need something like the apparatus of identity and quantification; hence far more than we are in a position to avail ourselves of in a language in which our high point as of even date is rabbit-announcing.

And the case is yet worse: we do not even have evidence for taking the native expression as of the form "A so-and-so is present"; it could as well be construed with an abstract singular term, as meaning that rabbithood is locally manifested. Better just "Rabbiteth," like "Raineth."

But if our linguist is going to be as cagey as all this, he will never translate more than these simple-minded announcements of observable current events. A cagey linguist is a caged linguist. What we want from the linguist as a serviceable finished product, after all, is no mere list of sentence-to-sentence equivalences, like the airline throwaways of useful Spanish phrases. We want a manual of instructions for custom-building a native sentence to roughly the purpose of any newly composed English sentence, within reason, and vice versa. The linguist has to resolve the potential infinity of native sentences into a manageable limited list of grammatical constructions and constituent linguistic forms, and then show how the business of each can be approximated in English; and vice versa. Sometimes perhaps he will translate a word or construction not directly but contextually, by systematic instructions for translating its containing sentences; but still he must make do with a limited lot of contextual definitions. Now once he has carried out this necessary job of lexicography, forwards and backwards, he has read our ontological point of view into the native language. He has decided what expressions to treat as referring to objects, and, within limits, what sorts of objects to treat them as referring to. He has had to decide, however arbitrarily, how to accommodate English idioms of identity and quantification in native translation.

The word "arbitrary" needs stressing, not because those decisions are wholly arbitrary, but because they are so much more so than one tends to suppose. For, what evidence does the linguist have? He started with what we may call native observation sentences, such as the rabbit announcement. These he can say how to translate into English, provided we impute no relevance to the differences between "Here a rabbit," "Here rabbithood," and the like. Also he can record further native sentences and settle whether various persons are prepared to affirm or deny them, though he find no rabbit movements or other currently observable events to tie them to. Among these untranslated sentences he may get an occasional hint of logical connections, by finding say that just the persons who are prepared to affirm *A* are prepared to affirm *B* and deny *C*. Thereafter his data leave off and his creativity sets in.

What he does in his creativity is attribute special and distinctive functions to component words, or conspicuously recurrent fragments, of the recorded sentences. The only ways one can appraise these attributions are as follows. One can see whether they add up to representing the rabbit sentence and the like as conforming to their previously detected truth conditions. One can see also how

well they fit the available data on other sentences: sentences for which no truth conditions are known, but only the varying readiness of natives to affirm or deny them. Beyond this we can judge the attributions only on their simplicity and naturalness – to us.

Certainly the linguist will try out his theory on the natives, springing new sentences authorized by his theory, to see if they turn out right. This is a permuting of the time order: one frames the theory before all possible data are in, and then lets it guide one in the eliciting of additional data likeliest to matter. This is good scientific method, but it opens up no new kind of data. English general and singular terms, identity, quantification, and the whole bag of ontological tricks may be correlated with elements of the native language in any of various mutually incompatible ways, each compatible with all possible linguistic data, and none preferable to another save as favored by a rationalization of the native language that is simple and natural to us.

It makes no real difference that the linguist will turn bilingual and come to think as the natives do – whatever that means. For the arbitrariness of reading our objectifications into the heathen speech reflects not so much the inscrutability of the heathen mind, as that there is nothing to scrute. Even we who grew up together and learned English at the same knee, or adjacent ones, talk alike for no other reason than that society coached us alike in a pattern of verbal response to externally observable cues. We have been beaten into an outward conformity to an outward standard; and thus it is that when I correlate your sentences with mine by the simple rule of phonetic correspondence, I find that the public circumstances of your affirmations and denials agree pretty well with those of my own. If I conclude that you share my sort of conceptual scheme, I am not adding a supplementary conjecture so much as spurning unfathomable distinctions; for, what further criterion of sameness of conceptual scheme can be imagined? The case of a Frenchman, moreover, is the same except that I correlate his sentences with mine not by phonetic correspondence but according to a traditionally evolved dictionary.¹ The case of the linguist and his newly discovered heathen, finally, differs simply in that the linguist has to grope for a general sentence-to-sentence correlation that will make the public circumstances of the heathen's affirmations and denials match up tolerably with the circumstances of the linguist's own. If the linguist fails in this, or has a hard time of it, or succeeds only by dint of an ugly and complex mass of correlations, then he is entitled to say – in the only sense in which one can say it – that his heathens have a very different attitude toward reality from ours; and even so he cannot coherently suggest what their attitude is. Nor, in principle, is the natural bilingual any better off.

When we compare theories, doctrines, points of view, and cultures, on the score of what sorts of objects there are said to be, we are comparing them in a respect which itself makes sense only provincially. It makes sense only as far afield as our efforts to translate our domestic idioms of identity and quantification bring encouragement in the way of simple and natural-looking correspondences. If we attend to business we are unlikely to find a very alien culture with a predilection for a very outlandish universe of discourse, just because the outlandishness of it would detract from our sense of patness of our dictionary of

translation. There is a notion that our provincial ways of positing objects and conceiving nature may be best appreciated for what they are by standing off and seeing them against a cosmopolitan background of alien cultures; but the notion comes to nothing, for there is no *ποῦ στῶ* [Greek: “place to stand”; the expression calls to mind Archimedes’ claim that, with a long enough lever and a point to stand on, he could move the world.].² . . .

Notes

- 1 See Richard von Mises, *Positivism* (Cambridge, Mass.: Harvard University Press, 1951), pp. 46 ff.
 - 2 For a fuller development of the foregoing theme see my “Meaning and translation” in Reuben Brower’s anthology *On Translation* (Cambridge, Mass.: Harvard University Press, 1959). For criticisms that have benefited the above section of the present essay and ensuing portions I am grateful to Burton Dreben.
-

44 After Metaphysics, What?*

Hilary Putnam

The death of metaphysics is a theme that entered philosophy with Kant. In our own century, a towering figure – Ludwig Wittgenstein – sounded that note both powerfully and in a uniquely personal way; and he did not hesitate to lump epistemology together with metaphysics. (According to some of Wittgenstein’s interpreters, what is today called “analytic philosophy” was, for Wittgenstein, the most confused form of metaphysics!) At the same time, even the man on the street could see that metaphysical discussion did not abate. A simple induction from the history of thought suggests that metaphysical discussion is not going to disappear as long as reflective people remain in the world. As Gilson said at the end of a famous book, “Philosophy always buries its undertakers.”

The purpose of this lecture is not to engage in a further debate about the question “Is (or: ‘In what sense is’) metaphysics dead?” I take it as a fact of life that there is a sense in which the task of philosophy is to overcome metaphysics and a sense in which its task is to continue metaphysical discussion. In every philosopher there is a part that cries, “This enterprise is vain, frivolous, crazy – we must say, ‘Stop!',” and a part that cries, “This enterprise is simply reflection at the most general and most abstract level; to put a stop to it would be a crime against reason.” *Of course*, philosophical problems are unsolvable; but as Stanley

* From Hilary Putnam, “After Metaphysics, What?” in Kieter Henrich and Rolf-Peter Horstmann, eds, *Metaphysik nach Kant?* (Stuttgart: Hegel-Kongreß, 1987; published in 1988). Reprinted by permission of the author and Klett-Cotta.

Cavell once remarked, “There are better and worse ways of thinking about them.”

What I just said could have been said at virtually any time since the beginning of modernity. I also take it – and this too is something I am not going to argue, but take as another fact of life, although I know that there are still those who would disagree – that the enterprises of providing a *foundation* for Being and Knowledge – a successful description of the Furniture of the World or a successful description of the Canons of Justification – are enterprises that have disastrously failed, and this could not have been seen until these enterprises had been given time to prove their futility (although Kant did say something like this about the former enterprise long ago). There is a sense in which the futility of something that was called metaphysics and the futility of something that was called epistemology is a sharper, more painful, problem for *our* period – a period that hankers to be called “Post-Modern” rather than modern.

What I want to do is lay out some principles that we *should not* abandon in our despair at the failure of something that was called metaphysics and something that was called epistemology. It will soon be evident that I have been inspired to do this, in large part, by a very fruitful ongoing exchange with my friend Richard Rorty, and this paper may be viewed as yet another contribution to that exchange. For Rorty, as for the French thinkers that he admires, two ideas seem gripping: (1) the failure of our philosophical “foundations” is a failure of the whole culture, and accepting that we were wrong in wanting or thinking we could have a “foundation” requires us to be *philosophical revisionists*. By this I mean that, for Rorty or Foucault or Derrida, the failure of foundationalism makes a difference to how we are allowed to talk in ordinary life – a difference as to whether and when we are allowed to use words like “know,” and “objective,” and “fact,” and “reason.” The picture is that philosophy was not a reflection *on* the culture, a reflection some of whose ambitious projects failed, but a *basis*, a sort of pedestal, on which the culture rested, and which has been abruptly yanked out. Under the pretense that philosophy is no longer “serious” there lies hidden a gigantic seriousness. If I am right, Rorty hopes to be a doctor to the modern soul. (2) At the same time, Rorty’s analytic past shows up in this: when he rejects a philosophical controversy, as, for example, he rejects the “realism anti-realism” controversy, or the “emotive cognitive” controversy, his rejection is expressed in a Carnapian tone of voice – he scorns the controversy.

I am often asked, “Just where do you disagree with Rorty?” Apart from technical issues – of course, any two philosophers have a host of technical disagreements – I think our disagreement concerns, at bottom, these two broad attitudes. I hope that philosophical reflection may be of some real cultural value; but I do not think it has been the pedestal on which the culture rested, and I do not think our reaction to the failure of a philosophical project – even a project as central as “metaphysics” – should be to abandon ways of talking and thinking which have practical and spiritual weight. I am not, in that sense, a philosophical revisionist. And I think that what is important in philosophy is not just to say, “I reject the realist anti-realist controversy,” but to show that (and *how*)

both sides *misrepresent* the lives we live with our concepts. That a controversy is “futile” does not mean the rival pictures are unimportant. Indeed, to reject a controversy without examining the pictures involved is almost always just a way of *defending* one of those pictures (usually the one that claims to be “anti-metaphysical”). In short, I think philosophy is both more important and less important than Rorty does. It is not a pedestal on which we rest (or have rested until Rorty). The illusions that philosophy spins are illusions that belong to the nature of human life itself, and that need to be illuminated. Just saying, “That’s a pseudo-issue,” is not of itself therapeutic; it is an aggressive form of the metaphysical disease itself.

These remarks are, of course, much too general to serve as answers to the question which titles this lecture. But no one philosopher can answer that question. “After metaphysics” there can only be *philosophers* – that is, there can only be the search for those “better and worse ways of thinking” that Cavell called for. . . .

Realism with a Small “r” and with an “R”

. . . If saying what we say and doing what we do is being a “realist,” then we had better be realists – realists with a small “r.” But metaphysical versions of “realism” go beyond realism with a small “r” into certain characteristic kinds of philosophical fantasy. Here I agree with Rorty.

Here is one feature of our intellectual practice that these versions have enormous difficulty in accommodating. On the one hand, trees and chairs – the “thises and thats we can point to” – are paradigms of what we call “real,” as Wittgenstein remarked.¹ But consider now a question about which Quine, Lewis, Kripke all disagree: what is the relation between the tree or the chair and the space-time region it occupies? According to Quine the chair and the electromagnetic, etc., fields that make it up and the space-time region that contains these fields are one and the same: so the chair is a space-time region. According to Kripke, Quine is just wrong: the chair and the space-time region are two numerically distinct objects. (They have the same mass, however!) The proof is that the chair *could have occupied a different space-time region*. According to Quine, modal predicates are hopelessly vague, so this “proof” is worthless. According to Lewis, Quine is right about the chair but wrong about the modal predicates: the correct answer [according] to Lewis is that if the chair could have been in a different place, as we say, what that means is that a *counterpart* of this chair could have been in that place; not that *this very chair* (in the sense of the logical notion of identity [=]) could have been in that place.

Well, who is right? Are chairs really *identical* with their matter or does a chair somehow coexist in the same space-time region with its matter while remaining numerically distinct from it? And is their matter really identical with the fields? And are the fields really identical with the space-time regions? To me it seems clear that at least the first, and probably all three, of these questions is nonsensical. We can formalize our language in the way Kripke would and we can for-

malize our language in the way Lewis would, and (thank God!) we can leave it unformalized and not pretend the ordinary language “is” obeys the same rules as the sign “=” in systems of formal logic. Not even God could tell us if the chair is “identical” with its matter (or with the space-time region); and not because there is something He doesn’t know.

So it looks as if even something as paradigmatically “real” as a chair *has aspects that are conventional. That the chair is blue is paradigmatically a “reality,” and yet that the chair [is/is not/don’t have to decide] a space-time region is a matter of convention.*

And what of the space-time region itself? Some philosophers think of points as location *predicates*, not objects. So a space-time region is just a set of properties (if these philosophers are right) and not an object (in the sense of concrete object) at all, if this view is right. Again, it doesn’t so much seem that there is a “view” here at all, as yet another way we could reconstruct our language. But how can the existence of a concrete object (the space-time region) be a matter of *convention*? And how can the identity of A (the chair) and B (the space-time region) be a matter of *convention*? The realist with a small “r” needn’t have an answer to these questions. It is just a fact of life, he may feel, that certain alternatives are equally good while others are visibly forced. But metaphysical realism is not just the view that there are, after all, chairs, and some of them are, after all, blue, and we didn’t just *make all that up*. Metaphysical realism presents itself as a powerful transcendental picture: a picture in which there is a fixed set of “language independent” objects (and some of them are abstract and others are concrete) and a fixed “relation” between terms and their extensions. What I am saying is that the picture only partly agrees with the common sense view it purports to interpret; it has consequences which, from a common sense view, are quite absurd. There is nothing wrong at all with holding on to our realism with a small “r” and jettisoning the Big “R” Realism of the philosophers.

Although he was far from being a Big “R” realist, Hans Reichenbach had a conception of the task of philosophy² which, if it had succeeded, might well have saved Realism from the objection just raised: the task of philosophy, he wrote, is to distinguish what is fact and what is convention (“definition”) in our system of knowledge. The trouble, as Quine pointed out, is that the philosophical distinction between “fact” and “definition” on which Reichenbach depended has collapsed. As another example, not dissimilar to the one I just used, consider the conventional character of any possible answer to the question, “Is a point identical with a series of spheres that converge to it?” We know that we can take extended regions as the primitive objects, and “identify” points with sets of concentric spheres, and all geometric facts are perfectly well represented. We know that we can also take points as primitives and take spheres to be sets of points. But the very statement “we can do either” assumes a diffuse background of empirical facts. Fundamental changes in the way we do physics could change the whole picture. So “convention” does not mean *absolute convention* – truth by stipulation, free of every element of “fact.” And, on the other hand, even when we see such a “reality” as a tree, the possibility of that perception is dependent on a whole conceptual scheme, on a language in place. What is factual

and what is conventional is a matter of degree; we cannot say, “these and these elements of the world are the raw facts; the rest is convention, or a mixture of these raw facts with convention.”

What I am saying, then, is that elements of what we call “language” or “mind” *penetrate so deeply into what we call “reality” that the very project of representing ourselves as being “mappers” of something “language independent” is fatally compromised from the very start.* Like Relativism, but in a different way, Realism is *an impossible attempt to view the world from Nowhere.*

In this situation it is a temptation to say, “So we make the world,” or “our language makes up the world,” or “our culture makes up the world”; but this is just another form of the same mistake. If we succumb, once again we view the world – the only world we know – as a *product*. One kind of philosopher views it as a product from a raw material: Unconceptualized Reality. The other views it as a creation *ex nihilo*. *But the world isn't a product. It's just the world.*

Where are we then? On the one hand – this is where I hope Rorty will sympathize with what I am saying – our image of the world cannot be “justified” by anything but its success as judged by the interests and values which evolve and get modified at the same time and in interaction with our evolving image of the world itself. Just as the absolute “convention/fact” dichotomy had to be abandoned, so (as Morton White³ long ago urged) the absolute “fact/value” dichotomy has to be abandoned, and for similar reasons. On the other hand, it is part of that image itself that the world is not the product of our will – or our dispositions to talk in certain ways, either.

Notes

- 1 Lecture xxv, *Wittgenstein's Lectures on Mathematics*, ed. Cora Diamond. “Theses and thots we can point to” is from this lecture.
- 2 Hans Reichenbach's *Philosophy of Space and Time* (New York: Dover, 1957).
- 3 Morton White, *Toward Reunion in Philosophy* (Cambridge, Mass.: Harvard University Press, 1956).

45 Truth and Convention*

Hilary Putnam

The ‘internal realism’ I have defended¹ has both a positive and a negative side. Internal realism denies that there is a fact of the matter as to which of the conceptual schemes that serve us so well – the conceptual scheme of

* From Hilary Putnam, ‘Truth and Convention: On Davidson's Refutation of Conceptual Relativism,’ *Dialectica*, 41 (1987), pp. 69–77. Reprinted by permission of the author and publisher.

commonsense objects, with their vague identity conditions and their dispositional and counterfactual properties, or the scientific-philosophical scheme of fundamental particles and their ‘aggregations’ (i.e., their mereological sums), is ‘really true’. Each of these schemes contains, in its present form, bits that will turn out to be ‘wrong’ in one way or another – bits that are right and wrong *by the standards appropriate to the scheme itself* – but the question ‘which kind of “true” is really Truth’ – is one that internal realism rejects.

A simple example² will illustrate what I mean. Consider ‘a world with three individuals’ (Carnap often used examples like this when we were doing inductive logic together in the early nineteen fifties), x_1, x_2, x_3 . How many *objects* are there in this world?

Well, I *said* ‘consider a world with just three individuals’ didn’t I? So mustn’t there be three objects? Can there be non-abstract entities which are not ‘individuals’?

One possible answer is ‘no’. We can identify ‘individual’, ‘object’, ‘particular’, etc., and find no absurdity in a world with just three objects which are independent, unrelated, ‘logical atoms’. But there are perfectly good logical doctrines which lead to different results.

Suppose, for example, like some Polish logicians, I believe that for every two particulars there is an object which is their sum. (This is the basic assumption of mereology, the calculus of parts and wholes invented by Lesniewski.) If I ignore, for the moment, the so-called ‘null object’, then I will find that the world of ‘three individuals’ (as Carnap might have had it, at least when he was doing inductive logic) actually contains *seven* objects (as shown in table 2).

Table 2

<i>World 1</i>	<i>World 2</i>
x_1, x_2, x_3	$x_1, x_2, x_3, x_1 + x_2, x_1 + x_3, x_2 + x_3, x_1 + x_2 + x_3$
(A world à la Carnap)	(‘Same’ world à la Polish logician)

Some logicians (though not Lesniewski) would also say that there is a ‘null object’ which they count as a part of every object. If we accepted this suggestion, and added this individual (call it O), then we would say that Carnap’s world contains *eight* objects.

Now, the classic metaphysical realist way of dealing with such problems is well known. It is to say that there is a single world (think of this as a piece of dough) which we can slice into pieces in different ways. But this ‘cookie cutter’ metaphor founders on the question, ‘What are the “parts” of this dough?’ If the answer is that $x_1, x_2, x_3, x_1 + x_2, x_1 + x_3, x_2 + x_3, x_1 + x_2 + x_3$ are all the different ‘pieces’, then we have not a *neutral* description, but rather a *partisan* description – just the description of the Warsaw logician! And it is no accident that metaphysical realism cannot really recognize the phenomenon of conceptual

relativity – for that phenomenon turns on the fact that *the logical primitives themselves, and in particular the notions of object and existence, have a multitude of different uses rather than one absolute ‘meaning’.*

An example which is historically important, if more complex than the one just given, is the ancient dispute about the ontological status of the Euclidean plane. Imagine a Euclidean plane. Think of the points in the plane. Are these parts of the plane, as Leibniz thought? Or are they ‘mere limits’, as Kant said?

If you say, in this case, that these are ‘two ways of slicing the same dough’, then you must admit that what is a part of space, in one version of the facts, is an abstract entity (say, a set of convergent spheres – although there is not, of course, a unique way of construing points as limits) in the other version. But then you will have conceded that which entities are ‘abstract entities’ and which are ‘concrete objects’, at least, is version-relative. Metaphysical realists to this day continue to argue about whether points (spacetime points, nowadays, rather than points in the plane or in three-dimensional space) are individuals or properties, particulars or mere limits, etc. My view is that God himself, if he consented to answer the question ‘Do points really exist or are they mere limits?’, would say ‘I don’t know’; not because His omniscience is limited, but because there is a limit to how far questions make sense.

One last point before I leave these examples: *given* a version, the question ‘How many objects are there?’ has an answer, namely ‘three’ in the case of the first version (‘Carnap’s World’) and ‘seven’ in the case of the second version (‘The Polish Logician’s World’). Once we make clear how we are using ‘object’ (or ‘exist’), the question ‘How many objects exist?’ has an answer that is not at all a matter of ‘convention’. That is why I say that this sort of example does not support cultural relativism. *Of course, our concepts are culturally relative;* but it does not follow that the truth or falsity of what we say using those concepts is simply ‘determined’ by the culture. But the idea that there is an Archimedean point (or a use of ‘exist’ inherent in the world itself) from which the question ‘How many objects *really* exist?’ makes sense, is an illusion.

Nor does it help, in general, to talk about ‘meanings’ or ‘truth conditions’. Consider again the two sentences (I am referring to the same example as before):

- (1) There is an object which is partly red and partly black.
- (2) There is an object which is red and an object which is black.

Observe that (2) is a sentence which is true in both the Carnapian and the Polish logician’s version if, say, x_1 is red and x_2 is black. (1) is a sentence which is true in the Polish logician’s version. What is its status in the Carnapian version?

Let me introduce an imaginary philosopher whom I will call ‘Professor Antipode’. Professor Antipode is violently opposed to Polish mereology. He talks like this, ‘I know what you’re talking about if by an object you mean a car, or a bee, or a human being, or a book, or the Eiffel Tower. I even understand it if you refer to my nose or the hood of my car as “an object”. But when philosophers say that there is an “object” consisting of the Eiffel Tower and my nose,

that's just plain crazy. There simply is no such object. Carnap was talking just fine when he said to you "consider a world with just three objects" – I ignore Carnap's regrettable tendency to what he called "tolerance" – and it's crazy to suppose that every finite universe contains all the objects those Poles would invent, or, if you please, "postulate". You can't create objects by "postulation" any more than you can bake a cake by "postulation".'

Now, the language Carnap had in mind (we were working together on inductive logic at the time, and most often the languages we considered had only one-place predicates) probably did not contain a two-place predicate for the relation 'part of'; but even if it did, we can imagine Professor Antipode denying that there is any object of which x_1 and x_2 are both 'parts'. 'If there were such an object, it would have to be different from both of them,' he would say (and here the Polish logician would agree), 'and the only object different from both of them in the world you showed us is x_3 . But x_3 does not overlap with either x_1 or x_2 . Only in the overheated imagination of the Polish logician is there such an additional object as $x_1 + x_2$.' If we add 'Part of' to Carnap's little language, so that sentence (1) can be expressed in it, thus:

$$(3) \quad (\exists x)(\exists y)(\exists z) (y \text{ is Part Of } x \& z \text{ is Part Of } x \& \text{Red}(y) \& \text{Black}(z)).$$

then, true to his anti-Polish form, Professor Antipode will say that this sentence is false. 'Whether you say it in plain English or in fancy symbols' he growls, 'if you have a world of three non-overlapping individuals, which is what Carnap described, and each is wholly red or wholly black, which is what Carnap said, then there cannot be such a thing in that world as an "object which is partly red and partly black". Talking about the "mereological sum of x_1 and x_2 makes no more sense than talking about "the mereological sum of my nose and the Eiffel Tower".'

Professor Antipode, it will be seen, is a staunch metaphysical realist. He *knows* that only some objects are parts of other objects, and that to say that for *every* pair of objects there is an object of which they both are parts (which is an axiom of mereology) is just 'rubbish'. (In the world Carnap imagined) (1) is false and (2) is true, and there's the whole story.

Carnap himself would have taken a very different attitude. Carnap was a conceptual relativist (that is, in part, what his famous Principle of Tolerance is all about), and he would have said that we can choose to make (1) false (that is, we can choose to talk the way Professor Antipode talks) *or* we can choose to make (1) true – to talk as the Polish logician talks. There is even – and this is very important – there is even a way in which we can have the best of both worlds. We keep Carnap's version as our official version (our 'unabbreviated language'); we refrain from adding Part Of as a new primitive, as we did before, but we introduce Part Of as a *defined* expression (as 'abbreviated language', or, as Quine often puts it, as a *façon de parler*). This can be done, not by giving an *explicit* definition of Part Of, but by giving a scheme which translates the Polish logician's language into Carnap's language (and such a scheme can easily be given in a recursive way, in the case of the kind of first order language with finitely

many individuals that Carnap had in mind). Under such a scheme, (1) turns out to say no more and no less than (2).

(To verify this, assuming that ‘red’ and ‘black’ are predicates of Carnap’s language, observe that the only way a Polish logician’s object – a mereological sum – can be partly red is by containing a red atom, and the only way it can be partly black is by containing a black atom. So if (1) is true in the Polish logician’s language, then there is at least one red atom and at least one black atom – which is what (2) says in Carnap’s language. Conversely, if there is at least one black atom and at least one red atom, then their mereological sum is an ‘object’ (in the Polish logician’s sense) which is partly red and partly black.)

While the formal possibility of doing this – of ‘interpreting’ the Polish logician’s version in Carnap’s version – is easy to establish, as a result in mathematical logic, the philosophical significance of this fact, of the interpretability of the second language in the first, is more controversial. An objection – an objection to the idea that this kind of interpretability supports conceptual relativity in any way – might come from a philosopher who pursues what is called ‘meaning theory’. Such a philosopher might ask, ‘What is the point of treating (1) as an abbreviation of (2), if it doesn’t, in fact, have the same *meaning* as (2)?’ Meaning theorists who follow Donald Davidson might argue that, while (1) and (2) are ‘mathematically equivalent’ (if, like the Polish logician, and unlike Professor Antipode, we are willing to count the axioms of mereology as having the status of logical or mathematical truths), still, sentence (2) is not a sentence one would ordinarily offer as an explanation of the truth conditions of sentence (1); or at least, doing so would hardly be in accordance with what is called ‘translation practice’. And a ‘meaning theory’, it is said, must not correlate just *any* extensionally or even mathematically correct truth conditions with the sentences of the language the theory describes; the sentence used to state a truth condition for a sentence must be one that might be correlated with that sentence by ‘translation practice’. Whatever one is doing when one invents reductive definitions that enable one to explain away talk about ‘suspicious’ entities as a mere *façon de parler*, it obviously isn’t just ‘radical translation’.

One suggestion as to what one *is* doing comes from a classic article by Quine. In ‘On What There Is’ he suggested that the stance to take in a case such as the one I have been describing – in a case in which one language seems more useful than another, because it countenances entities which (although philosophically ‘suspicious’) enable us to say various things in fewer words, and in which the, at first blush, ‘richer’ language is formally interpretable in the, at first blush, ‘poorer’ language – might be to say – this is a stance Professor Antipode might adopt – ‘Sentence (1), asserting as it does the existence of mereological sums, is literally false. But if one wants to go on talking like the Polish logician while rejecting his undesirable ontological commitments, one can do that. One can responsibly take the view that the Polish logician’s story is only a useful make-believe, and yet employ its idioms, on the ground that each of the sentences in that idiom, whatever its “meaning”, *can* be regarded – by fiat, if you like – as merely a convenient abbreviation of whatever sentence in the “unabbreviated language” it is correlated with by the interpretation scheme.’

To give another example, one long familiar to students of mathematical philosophy, Frege and Russell showed that number theory is interpretable in set theory. This means that, if one wants to avoid ontological commitments to ‘unreduced numbers’ (to numbers as objects over and above sets) – and if one does not mind commitment to *sets!* – one can treat every sentence of number theory, and, indeed, every sentence in the language which uses a number word, as a mere abbreviation for another sentence, one which quantifies over sets, but not over any such entities as ‘numbers’. One need not claim that the sentence of number theory and its translation in set theory have the same ‘meaning’. If they don’t, so much the worse for our intuitive notion of a ‘number’! What this kind of interpretation – call it *reductive interpretation* – provides is evidence against the real existence of the unreduced entities, as anything over and above the entities countenanced by the language to which we are doing the reducing. The moral we should draw from the work of Frege and Russell is not that there is a conceptual *choice* to be made between using a language which countenances only sets and one which countenances sets *and* numbers, but that – unless the numbers are in fact identical with the set with which we identified them – there is no reason to believe in the existence of numbers. Talk of numbers is best treated as a mere *façon de parler*. Or so Quine maintains.

It is easy to see why Professor Antipode should like this line. In the case of the two versions we have been discussing, the reductive interpretation is syncategorematic; that is, it interprets sentence (1) (and likewise any other sentence of Carnap’s language) as a whole, but does not identify the individual words in (1) with individual words and phrases in (2); nor does it identify ‘mereological sums’ with any objects in the language to which the reducing is being done. (1) as a whole is ‘translated’ by (2) as a whole; but the noun-phrase ‘object which is partly red and partly black’ has no translation by itself. In this case the moral of the translation – the moral if Professor Antipode imitates Quine’s rhetoric – is slightly different. We cannot say *either mereological sums are identical with the entities with which we identified them or they don’t really exist* (because the ‘translation’, or relative interpretation of the Polish logician’s language in Carnap’s language, didn’t identify ‘mereological sums’ with *anything*, it just showed how to translate sentences about them syncategorematically). The moral is rather, *mereological sums don’t really exist, but it is sometimes useful to talk as if they existed*. Of course Professor Antipode would be delighted with *this* moral!

I don’t mean to give the impression that the possibility of reducing entities away by a formal translation scheme is always decisive evidence that they don’t really exist, according to Quine. Sometimes we have the choice of either doing without one batch of entities, call them the **A** entities, or doing without another batch, call them the **B** entities – the reduction may be possible in either direction. In such a case, Occam’s Razor doesn’t know who to shave! Or the reducing language may itself seem suspicious (some people think *sets* are very suspicious entities). But, when the reducing language (the *prima facie* ‘poorer’ language) is one we are happy with, and the reduction does not go both ways, it is clear that Quine regards this as very strong evidence for denying the real existence of the unreduced entities.

Carnap, on the other hand, rejected the idea that there is ‘evidence’ against the ‘existence’ of numbers (or against the existence of numbers as objects distinct from sets). He would, I am sure, have similarly rejected the idea that there is evidence against the ‘existence’ of mereological sums. I know what he would have said about this question: he would have said that the question is one of a choice of a language. On some days it may be convenient to use what I have been calling ‘Carnap’s language’ (although he would not have *objected* to the other language); on other days it may be convenient to use the Polish logician’s language. For some purposes it may be convenient to regard the Polish logician’s language of mereological sums as ‘primitive notation’; in other contexts it may be better to take Carnap’s language as the primitive notation and to regard the Polish logician’s language as ‘abbreviations’, or defined notation. And I agree with him.

It will be seen that there are a number of different stances one could take to the question of the *relation* between (1) and (2). One could say:

- (a) The two sentences are mathematically equivalent.
- (b) The two sentences are logically equivalent.
- (c) The two sentences are neither logically or mathematically equivalent.
- (d) The first sentence is false and the second true (Professor Antipode’s position).
- (e) The two sentences are alike in truth value and meaning.
- (f) The two sentences are alike in truth value and unlike in meaning.
- (g) The second sentence can be used as an abbreviation of the first, but this is really just a useful ‘make-believe’.

My own position – and my own internal realism – is that there is no fact of the matter as to which of *these* positions is correct. Taking the original dispute up into the ‘metalevel’ and reformulating it as a dispute about the properties – mathematical or logical equivalence, synonymy, or whatever – of linguistic forms doesn’t help. None of these notions is well defined enough to be a useful tool in such cases. . . .

Notes

- 1 Cf. my *Reason, Truth and History* (Cambridge: Cambridge University Press, 1981).
- 2 This example comes from my *The Many Faces of Realism* (LaSalle, Ill.: Open Court, 1987).

46 Nonabsolute Existence and Conceptual Relativity: an Excerpt from “Putnam’s Pragmatic Realism”*

Ernest Sosa

Suppose a world with just three individuals x_1 , x_2 , x_3 . Such a world is held by some “mereologists” to have in it a total of seven things or entities or objects, namely, x_1 , x_2 , x_3 , $x_1 + x_2$, $x_1 + x_3$, $x_2 + x_3$, $x_1 + x_2 + x_3$. Antimereologists by contrast prefer the more austere ontology that recognizes only the three individuals as objects that *really* exist in that world. Talk of the existence of $x_1 + x_2$ and its ilk is just convenient abbreviation of a more complex discourse that refers to nothing but the three individuals. Thus, suppose x_1 is wholly red and x_2 is wholly black. And consider

- (1) There is an object that is partly red and partly black.
- (2) There is an object that is red and an object that is black.

For the antimereologist, statement 1 is not true, if we assume that x_3 is also wholly red or wholly black. It is at best a convenient way of abbreviating the likes of 2.

Putnam has now joined Rudolf Carnap in viewing our controversy as follows:

... the question is one of the choice of language. On some days it may be convenient to use [antimereological language]; ... on other days it may be convenient to use [mereological] language.¹

Take the question

How many objects with a volume of at least 6 cubic centimeters are there in this container?

This question can have no absolute answer on the Carnap–Putnam view, even in a case where the container contains a vacuum except for three marbles each with a volume of 6 cubic centimeters. The antimereologist may say

- (3) There are three objects in the box.

But the mereologist will reply:

* From Ernest Sosa, “Putnam’s Pragmatic Realism,” *Journal of Philosophy*, 90 (1990), pp. 605–26. Reprinted by permission of the author and *Journal of Philosophy*.

- (4) There are at least seven objects in the box.

The Carnap–Putnam line is now this: *which statement we accept – 3 or 4 – is a matter of linguistic convenience*. The language of mereology has criteria of existence and identity according to which sums of individuals are objects. The language of antimereology rejects such criteria, and may even claim that by its criteria only individuals are objects.

There is a valuable insight here, I believe, but I am puzzled by the linguistic wrapping in which it is offered. After all, none of 1–4 mentions any language or any piece of language, nor does any of them say that we shall or shall not or should or should not use any language or bit of language. So I do not see how our decision actually to use or not to use any or all of the sentences 1–4 can settle the question of whether what these sentences *say* is true or false. And if the point is that these sentences do not really *say* anything, then how can they be incompatible in the first place so that a conflict or problem can arise that requires resolution? Also, it is not clear how we gain by replacing questions about atoms (or the like) with questions about *sentences* and *our* relations to some specific ones of these sentences. This is all very puzzling, and we should pause to peer more closely.

What does the proposed linguistic relativity amount to? Can it be spelled out more fully and prosaically? Here, for a start, is a possibility:

- LR1 In order to say *anything* you must adopt a language. So you must “adopt a meaning” even for so basic a term as “object”. And you might have adopted another. Thus you might adopt Carnap-language (CL) or you might adopt Polish-logician-language (PL). What you say, i.e., the utterances you make, the sentences you affirm, are not true or false absolutely, but are true or false only relative to a given language. Thus, if you say “There are three objects in this box” your utterance or sentence may be true understood as a statement of CL while it is false understood as a statement in PL.

But under this interpretation linguistic relativity seems trivially true. Who could deny that inscriptions of shapes and emissions of sounds are not true or false independently of their meaning, independently of all relativization to language or idiolect? Of course, you must “adopt a language” in order to speak (though such “adoption” need not be a conscious and voluntary act), and indeed you might have adopted another. And it seems quite uncontroversial that an utterance of yours might be true relative to one language while it is false relative to another.

Perhaps then the point is rather this:

- LR2 When we say “There are 3 objects here, not 8” we are really saying: “The following is assertible as true in our CL: ‘There are 3 objects here, not 8’.”

This is indeed in the spirit of Carnap's philosophy, whose *Logical Syntax of Language*,² published in English in 1937, defends the following theses:

- (i) Philosophy, when cognitive at all, amounts to the logical syntax of scientific language.
- (ii) But there can be alternative such languages and we are to choose between them on grounds of convenience.
- (iii) A language is completely characterized by its formation and transformation rules.

In that book Carnap also distinguishes between:

- (s1) Object sentences: e.g., "Five is a prime number," "Babylon was a big town."
- (s2) Pseudo-object sentences: e.g., "Five is not a thing but a number," "Babylon was treated of in yesterday's lecture."
- (s3) Syntactical sentences: e.g., "'Five' is not a thing-word but a number-word," "'Babylon' occurred in yesterday's lecture."

And he defends the thesis that s2 sentences seem deceptively like s1 sentences but are really s3 sentences in "material mode" disguise.

It was W. V. Quine who in 1934 suggested "material mode" to Carnap (as Quine himself reports in the section on "Semantic Ascent" in *Word and Object*³). Quine agrees that a kind of "semantic ascent" is possible, as when we shift from talk of miles to talk of "mile", but he thinks this kind of semantic ascent is *always* trivially available, not just in philosophy but in science generally and even beyond. Thus, we can paraphrase "There are wombats in Tasmania" as "'Wombat' is true of some creatures in Tasmania." Quine does grant that semantic ascent tends to be especially useful in philosophy. But he explains why as follows:

The strategy of semantic ascent is that it carries the discussion into a domain where both parties are better agreed on the objects (viz., words) and on the main terms concerning them. Words, or their inscriptions, unlike points, miles, classes, and the rest, are tangible objects of the size so popular in the marketplace, where men of unlike conceptual schemes communicate at their best. . . . No wonder it helps in philosophy.⁴

The use of this strategy, however, is clearly limited to discourse about recondite entities of controversial status. No relevant gain is to be expected from semantic ascent when the subject matter is the inventory of the marketplace itself. Tables and chairs are no more controversial than words: in fact, they seem less so, by a good margin. No general internal realism, with its conceptual or linguistic relativity, can be plausibly supported by the semantic ascent strategy offered by Quine.

In addition, questions of coherence arise concerning LR2. When we say

something of the form “The following is assertible in our CL: . . .” can we rest with a literal interpretation that does not require ascent and relativization? If not, where does ascent stop? Are we then *really* saying “The following is assertible in our CL: ‘The following is assertible in our CL: . . .’.” This way lies vicious regress. But if we *can* stop the regress with our metalinguistic reference to our sentences of CL (and to ourselves), why can we not stop it with our references to tables and chairs and other medium-sized dry goods? . . .

There is hence reason to doubt the linguistic turn taken by Carnap and now Putnam. We have found no very plausible way to conceive of the turn so that it discloses an attractive new direction in metaphysics. The only direction that seems certainly right and clearly defensible is that provided by our first interpretation above (interpretation LRI), but that also seemed trivially right, and not something anyone would deny, not even the most hard-line metaphysical realist. Nevertheless, it still seems to me that there is a valuable insight in Putnam’s now repeated appeal to the contrast between the Carnapian conceptual scheme and that of the Polish logician. But, given our recent reflections, I would like to put the insight without appeal to language or to any linguistic relativity.

The artifacts and even the natural objects that we recognize as existing at a time are normally composed of stuff or of parts in certain ways, and those which we see as enduring for an interval are normally not only thus composed of stuff or of parts at each instant of their enduring; but also the stuff or parts thus composing them right up to t , must be related in certain restricted ways to the stuff or parts that compose them right after t , for any time t within the history of such an enduring object.

Thus, the existence of a snowball at a time t and location 1 requires that there be a round quantity of snow at 1 and t sufficiently separate from other snow, etc.; and for that snowball to endure through an interval I , it is required that for every division of I into a sequence of subintervals I_1, I_2, \dots , there must be a corresponding sequence of quantities of snow Q_1, Q_2, \dots , related in certain restricted ways. By all this I mean to point to our “criteria of existence and perdurance for snowballs.”

I spoke of a snowball, its existence and perdurance, and what that requires of its sequence of constituent quantities of snow. In place of these, I might have talked of chains and constituent links, of boxes and constituent sides, or of a great variety of artifacts or natural entities such as hills or trees; or even – especially – of persons and their constituent bodies. In every case, there are criteria of existence and of perdurance for an entity of the sort in question such that necessarily an entity of the sort exists at t (perdures through I) if and only if its criteria of existence are satisfied at t (its criteria of perdurance are satisfied relative to I). Thus, necessarily a snowball exists at t if and only if at t a quantity of snow is round and separate from other snow; and a snowball perdures through I if and only if for any subdivision of I into a sequence of subintervals I_1, I_2, \dots , there must be a corresponding sequence of round, etc., quantities of snow Q_1, Q_2, \dots , such that, for all i , Q_i satisfies the conditions for being successor of Q_{i-1} in the constitution of the “life” of a snowball. And similarly for chains, boxes, hills, trees, and persons.

I am supposing a snowball to be constituted by a certain piece of snow as constituent matter and the shape of (approximate) roundness as constituent form. That particular snowball exists at that time because of the roundness of that piece of snow. More, if at that time that piece of snow were to lose its roundness, then at that time that snowball would go out of existence.

Compare now with our ordinary concept of a snowball, the concept of a snowdiscall, defined as an entity constituted by a piece of snow as matter and as form any shape between being round and being disc-shaped. At any given time, therefore, any piece of snow that constitutes a snowball constitutes a snowdiscall, but a piece of snow might at a time constitute a snowdiscall without then constituting a snowball. For every round piece of snow is also in shape between disc-shaped and round (inclusive), but a disc-shaped piece of snow is of course not round.

Any snowball SB must hence be constituted by a piece of snow PS which also then constitutes a snowdiscall SD. Now, SB is distinct (a different entity) from PS, since PS would survive squashing and SB would not. By similar reasoning, SD also is distinct from PS. And, again by similar reasoning, SB must also be distinct from SD, since enough partial flattening of PS will destroy SB but not SD. Now, there are infinitely many shapes S_1, S_2, \dots , between roundness and flatness of a piece of snow, and, for each i , having a shape between flatness and S_i would give the form of a distinctive kind of entity to be compared with snowballs and snowdiscalls. Whenever a piece of snow constitutes a snowball, therefore, it constitutes infinitely many entities all sharing its place with it.

Under a broadly Aristotelian conception, therefore, the barest flutter of the smallest leaf hence creates and destroys infinitely many things, and ordinary reality suffers a sort of "explosion."

We might perhaps resist this "explosion" of our ordinary world by embracing conceptual relativism. Constituted, supervenient entities do not just objectively supervene on their requisite, constitutive matters and forms, outside all conceptual schemes, with absolute independence from the categories recognized by any person or group. Perhaps snowballs do exist relative to all actual conceptual schemes ever, but not relative to all conceivable conceptual schemes. Just as we are not willing to countenance the existence of snowdiscalls, just so another culture might have been unwilling to countenance snowballs. We do not countenance snowdiscalls, because our conceptual scheme does not give to the snowdiscall form (being in shape between round and disc-shaped) the status required for it to be a proper constitutive form of a separate sort of entity – at least not with snow as underlying stuff.

That would block the explosion of reality, but the price is conceptual relativity. Supervenient, constituted entities do not just exist or not in themselves, free of any dependence on or relativity to a conceptual scheme. What thus exists relative to one conceptual scheme may not do so relative to another. In order for such a sort of entity to exist relative to a conceptual scheme, that conceptual scheme must recognize its constituent form as an appropriate way for a separate sort of entity to be constituted.

Must we now conceive of the existence even of the conceptual scheme itself

and of its framers and users as also relative to that conceptual scheme? And are we not then caught in a vicious circle? The framers exist only relative to the scheme and this they do in virtue of the scheme's giving their constituent form-cum-matter the required status. But to say that the scheme gives to this form-cum-matter the required status – is that not just to say that the *framers* of that scheme do so? Yet are not the framers themselves dependent on the scheme for their existence relative to it?

Answer: existence *relative* to a conceptual scheme is *not* equivalent to existence *in virtue of that conceptual scheme*. Relative to scheme *C* the framers of *C* exist *in virtue* of their constitutive matter and form, and in virtue of how these satisfy certain criteria for existence and perdurance of such subjects (among whom happen to be the framers themselves). This existence of theirs is in that way relative to *C* but not in virtue of *C*. There is hence no vicious circularity.

The picture then is roughly this. Each of us acquires and develops a view of things that includes criteria of existence and perdurance for categories of objects. When we consider whether an object of a certain sort exists, the specification of the sort will entail the relevant criteria of existence and perdurance. And when we correctly recognize that an object of that sort does exist, our claim is elliptical for “... exists relative to *this* our conceptual scheme.”

Again, this is *not* the only conceivable view of the matter. We could try to live with the explosion. And that does seem almost inevitable if we view it this way: a sort of object *O* – a constituted, supervenient sort – comes with a sort of constituent matter *M*, or sorts of constituent matters *M*₁, *M*₂, ..., and a sort of constituent form *F*. These – *M* (or *M*₁, *M*₂, ...), and *F* – we may take to be given independently of any acceptance by anyone of any criteria of existence or perdurance. For the sake of argument, then, we are accepting as given the sorts of items – *M*₁, *M*₂, ... – that will play the role of constituent matters, and also the property or relation – *F* – that will play the role of constituent form. And presumably whether or not any particular sequence of matters [*m*₁, *m*₂, ...] of sorts *M*₁, *M*₂, ..., respectively, does or does not satisfy form *F* is also generally independent of whether or not we accept any criteria of existence or perdurance, and indeed independent of whether *anyone* does so.

Suppose there is a time *t* when our conceptual scheme *C* first recognizes the appropriate criteria of existence and perdurance. According to our conceptual relativism, prior to that time *t* there were, relative to *C*, no objects of sort *O*, and in particular object *o* did not exist. But if there were no objects of sort *O*, such as *o*, relative to our scheme *C*, then why complicate our own scheme by supplementing it with criteria of existence and perdurance which do give standing to objects of sort *O*? After all, it is not as though we would fail to recognize the existence of something already in existence. By hypothesis *there are no objects of sort O*, not right up to that time *t*, anyhow.

On the other side, there is the threat of exploding reality, however. If we allow the satisfaction by any sequence *S* of any form *F* of the appropriate polyadicity and logical form to count as a criterion of existence for a corresponding sort of object, then reality right in us, before us, and all around us is unimaginably richer and more bizarre than we have ever imagined. And anyway

NONABSOLUTE EXISTENCE AND CONCEPTUAL RELATIVITY

we shall still face the problem of giving some explanation for why we focus so narrowly on the objects we do attend to, whose criteria of existence and perdurance we do recognize, to the exclusion of the plethora of other objects all around and even in the very same place.

A third option is a disappearance or elimination theory that refuses to countenance supervenient, constituted objects. But then most if not all of ordinary reality will be lost. Perhaps we shall allow ourselves to continue to use its forms of speech “. . . but only as a convenience or abbreviation.” But in using those forms of speech, in speaking of snowballs, chains, boxes, trees, hills, or even people, we shall *not* believe ourselves to be seriously representing reality and its contents. “As a convenience”: to *whom* and for what *ends*? “As an abbreviation”: of *what*?

With alternatives so grim, we are encouraged to return to our relativistic reflections. Our conceptual scheme encompasses criteria of existence and of perdurance for the sorts of objects that it recognizes. Shall we say now that a sort of object *O* exists (has existed, exists now, or will exist) relative to a scheme *C* at *t* if and only if, at *t*, *C* recognizes sort *O* by allowing the corresponding criteria? But surely there are sorts of objects that our present conceptual scheme does not recognize, such as artifacts yet uninvented and particles yet undiscovered, to take only two obvious examples. Of course, we allow there might be and probably are many such things. Not that there could be any such entities relative to our *present* conceptual scheme, however, for by hypothesis it does not recognize them. So are there sorts of objects – constituted sorts among them, as are the artifacts at least – such that they exist but not relative to our present scheme *C*? In that case we are back to our problem. What is it for there to be such objects? Is it just the in-itself satisfaction of constitutive forms by constitutive matters? That yields the explosion of reality.

Shall we say then that a constituted, supervenient sort of object *O* exists relative to our present scheme *C* if and only if *O* is recognized by *C* directly or recognized by it indirectly through being recognized by some predecessor or successor scheme? That, I fear, cannot suffice, since there might be sorts of particles that always go undiscovered by us, and sorts of artifacts in long disappeared cultures unknown to us, whose conceptual schemes are not predecessors of ours.

Shall we then say that what exists relative to our present scheme *C* is what it recognizes directly, what it recognizes indirectly through its predecessors or successors, and what it *would* recognize if we had developed appropriately or were to do so now, and had been or were to be appropriately situated? This seems the sort of answer required, but it obviously will not be easy to say what appropriateness amounts to in our formula, in its various guises.

Regardless of whatever success may await any further specification of our formula, there is the following further objection. Take a sort of object *O* recognized by our scheme *C*, with actual instances *o*1, *o*2, . . . ; for example, the sort Planet, with various particular planets as instances: Mercury, Venus, etc. Its instances, say we, exist, which amounts to saying that they exist relative to our scheme. But if we had not existed there would have been no scheme of ours for

anything to exist relative to; nor would there have been our actual scheme C either. For one thing, we may just assume the contingent existence of our actual scheme to depend on people's actually granting a certain status to certain constitutive forms. If we had not existed, therefore, the constitutive form for the sort Planet would not have had, relative to our conceptual scheme, the status required for it to be possible that there be instances of that sort, particular planets. And from this it apparently follows that if we had not existed there would have been no planets: no Mercury, no Venus, etc.

This objection conceptual relativism can rebut as follows. While existing in the actual world x we now have a conceptual scheme Cx relative to which we assert existence, when we assert it at all. Now, we suppose a possible world w in which we are not to be found, in which indeed no life of any sort is to be found. Still we may, in x : (a) consider alternative world w and recognize that our absence there would have no effect on the existence or course of a single planet or star, that Mercury, Venus, and the rest, would all still make their appointed rounds just as they do in x ; while yet (b) this recognition, which after all takes place in x , is still relativized to Cx , so that the existence in w of whatever exists in w relative to Cx need not be affected at all by the absence from w of Cx , and indeed of every conceptual scheme and of every being who could have a conceptual scheme. For when we suppose existence in w , or allow the possibility of existence in w , we do so *in x*, and we do so there still relative to Cx , to our present conceptual scheme, and what it recognizes directly or indirectly, or ideally.

If I am right we have three choices:

Eliminativism:

a disappearance view for which our ordinary talk is so much convenient abbreviation. Problem: we still need to hear: "abbreviation" of what, and "convenient" for what ends and whose ends? Most puzzling of all is how we are to take this "abbreviation" – not literally, surely.

Absolutism:

snowballs, hills, trees, planets, etc., are all constituted by the in-itself satisfaction of certain conditions by certain chunks of matter, and the like, and all this goes on independently of any thought or conceptualization on the part of anyone. Problem: this leads to the "explosion of reality."

Conceptual relativism:

we recognize potential constituted objects only relative to our implicit conceptual scheme with its criteria of existence and of perdurance. Problem: is there not much that is very small, or far away, or long ago, or yet to come, which surpasses our present acuity and acumen? How can we allow the existence of such sorts at present unrecognized by our conceptual scheme?

Right now I cannot decide which of these is least disastrous. But is there any other option?

Notes

- 1 Hilary Putnam, "Truth and Convention: On Davidson's Refutation of Conceptual Relativism," *Dialectica*, xli (1987), pp. 69–77; p. 75 [reprinted in this volume].
 - 2 Rudolf Carnap, *Logical Syntax of Language* (New York: Harcourt Brace, 1937).
 - 3 W. V. Quine, *Word and Object* (Cambridge, Mass.: MIT, 1960).
 - 4 Ibid., p. 272.
-

47 Addendum to “Nonabsolute Existence and Conceptual Relativity”: Objections and Replies

Ernest Sosa

Objection 1

What does it matter whether we “recognize” the snowdiscall form (being in shape between round and disc-shaped, inclusive)? We are anyhow “committed” to there being such a property in any case, to there being the property or condition of being shaped in that inclusive way. If a piece of snow is in shape anywhere between disc-shaped and round then it just is a snowdiscall. So there must be lots of snowdiscalls in existence and that must be nothing new. What’s the problem? Can’t we even just define a “glug” as anything that is a cat or a dog, and are there not as many glugs in existence as are in the union of the set of cats and the set of dogs? Why should anyone worry about this “explosion?” Why not just admit the obvious: that, yes, there are snowdiscalls, and glugs, even if heretofore they had not been so-called?

Not only is that obvious. If anyone is misguided enough to want to avoid admitting the obvious, it does not really help to introduce some conceptually relative notion of existence according to which the entities that so exist are only those that we are committed to through the properties and kinds that we admit in our ideology and ontology. For if we admit being a dog as an ordinary, harmless enough property, and the kind dog as well, along with being a cat, etc., then we are implicitly committed to admitting anything that is either a dog or a cat, as being “either a dog or a cat,” and that is tantamount to admitting that there are glugs – not under this description, of course, but what does that matter?

Reply

That is all quite true, of course, but not in conflict with conceptual relativism, not with the ontological conceptual relativism CR at issue here. Conceptual relativism is a thesis about ontological constitution. It presupposes that there are levels of individuals, and thus individuals on a higher level, constituted out of individuals on a lower level. The question now arises: How are the constituted entities constitutable out of the constituting entities? A partial (Aristotelian) answer is that a constituted entity must derive from the satisfaction by the constituting entity (or entities) of a condition (a property or relation, a “form”). *Any* condition whatever? That is absolutism, and leads to the “explosion.” Only conditions from a restricted set? *How*, in what way, restricted? Somehow by reference to the conceptual scheme of the speaker or thinker who attributes existence? This is conceptual relativism (of the sort at issue here).

Returning to the examples of the objection: First, yes, of course there are snowdiscalls if all one means by this is that there are pieces of snow with a shape somewhere between disc-shaped and round. And when something is so shaped and, also, more specifically, round, then it is not only such a snowdiscall but also a round piece of snow, a “snowround” let’s say. But it is just one and numerically the same thing which is then both the snowdiscall and the snowround. And this is no more puzzling than is the fact that something can be both a mother and a daughter, or both red and round, or both an apple and a piece of fruit, etc., etc. When *I* introduced the term “snowdiscall” this is not what I had in mind. In my sense, a “snowdiscall” is not just any piece of snow with a shape between round and disc-shaped. Nor is a snowball just a round piece of snow. For a round piece of snow can survive squashing, unlike the snowball that it constitutes, which is destroyed, not just changed, when it is squashed. The question is: what is special about the form of being round combined with an individual piece of snow, what is special about the ordered pair, let’s say, that makes it a suitable form/matter pair for the constitution of a constituted individual, a particular snowball? Would any other shape, between roundness and flatness, also serve as such a form, along with that individual piece of snow? Could they together yield a form/matter pair that might also serve, in its own way, for the formation, the constitution of its own individual: not a snowball, presumably, but its own different kind of individual? It is to *this* question that the absolutist would answer in the affirmative, while the conceptual relativist might well answer in the negative.

According to conceptual relativism in ontology, what then is required for a form/matter pair to serve as the form and matter for the constitution of an individual, a constituted individual? Answer: that the sort of form/matter combination in question be countenanced by the relevant conceptual scheme, a conceptual scheme determined by the context of thought and/or utterance.

Objection 2

If it is granted that things can exist prior to the development of any conceptual scheme whatever, prior to the evolution of any thinkers who could have a conceptual scheme, is that not a concession to absolutism? Is it not being conceded that things exist “out there, in themselves,” independently of conceptual schemes altogether, so that things do not exist in virtue of our conceptual choices after all? Rather things exist “in themselves.” Reality itself manages somehow to cut the cookies unaided by humans. Isn’t this just absolutism after all? What can be left of conceptual relativism after this has been granted?

Reply

Compare this. If I say “The Empire State Building is 180 miles southwest of here” my utterance is true, but the sentence I utter is true only relative to my present position. If I had uttered that sentence elsewhere then I might well have said something false. So my sentence is true relative to my spatial position, but it is not true or false just on its own, independently of such context. And, in a sense, that the Empire State Building is 180 miles southwest is something that is true relative to my present position but is false relative to many other positions. However, it is not so that the Empire State Building is 180 miles southwest of here *in virtue of* my present position. The Empire State Building would have been 180 miles southwest of here even if I had been located elsewhere. Whether I am here or not does not determine the distance and direction of the Empire State Building relative to this place here.

Conceptual relativism can be viewed as a doctrine rather like the relativism involved in the evaluation of the truth of indexical sentences or thoughts. In effect, “existence claims” can be viewed as implicitly indexical, and that is what my conceptual relativist in ontology is suggesting. So when someone says or thinks that Os exist, this is to be evaluated relative to the position of the speaker or thinker in “conceptual space” (in a special sense). Relative to the thus distinguished conceptual scheme, it might be that Os do exist, although relative to many other conceptual schemes it might rather be true to say that “Os do *not* exist.”

But what is it about a “conceptual scheme” that determines whether or not it is true to say that “Os exist”? Answer: what determines whether “there are” constituted entities of a certain sort relative to a certain conceptual scheme would be that scheme’s Criteria of Existence (or Individuation). And what are these? They are specifications of the appropriate pairings of kinds of individuals with properties or relations. Appropriate for what? For the constitution of constituted entities, *in the dispensation of that conceptual scheme*.

When one says or thinks “Os exist,” then, according to conceptual relativism this is not true or false absolutely. Its truth value must be determined relative to one’s conceptual scheme, to one’s “conceptual position,” including its criteria of existence. However, even if one’s claim that “Os exist” must be determined

relative to one's conceptual position, so that it can be very naturally said that "Os exist" relative to one's conceptual position (in that sense), it does not follow that "Os exist" only *in virtue* of one's conceptual position, in the sense that if one had not existed with some such conceptual scheme, or at least if no-one had existed with some such conceptual scheme, then there would have been "no Os in existence." This no more follows than it follows from the relativity of the truth of my statement "The Empire State Building is 180 miles southwest of here" that the Empire State Building has the distance and direction that it has from here as a result of *my* being here (even if I am the speaker or thinker). Despite the relativity of the truth of my statement, the Empire State Building *would have been* exactly where it is, 180 miles from here, even if I had not been here. Similarly, Os might have existed relative to this my (our) conceptual position, even if no-one had existed to occupy this position.

PART FOUR

WHY IS THERE A WORLD?

Introduction

Is There an Answer?

- 48 The Problem of Being: Chapter 3 of *Some Problems of Philosophy*
WILLIAM JAMES
- 49 The Puzzle of Reality: Why does the Universe Exist?
DEREK PARFIT
- 50 Response to Derek Parfit
RICHARD SWINBURNE

Does the Answer Involve a Necessary Being?

- 51 The Cosmological Argument and the Principle of Sufficient Reason
WILLIAM L. ROWE
- 52 The Ontological Argument: Chapters II–IV of the *Proslogion*
ST ANSELM
- 53 Anselm's Ontological Arguments
NORMAN MALCOLM

Introduction

Given our understanding of “world” (see the Preface), the question to be considered in this part amounts to the question of why there is something rather than nothing.

A Is There an Answer?

William James and Derek Parfit agree that the question makes sense. Parfit explores a range of forms an answer might take; Richard Swinburne, in his reply, argues that none of the alternatives Parfit suggests is as plausible or explanatorily complete as the traditional answer: God explains the existence of everything besides Himself, His own existence being taken for a “brute fact.”

B Does the Answer Involve a Necessary Being?

Not content to let the existence of God remain a brute fact, many have held that He is a necessary being – something that could not possibly have failed to exist, and so either explains itself or needs no explanation (take your pick). The cosmological and ontological arguments are supposed to show that there must be a necessary being. William Rowe finds Samuel Clarke’s version of the cosmological argument to be valid, but of little use for the purposes of convincing a modern-day atheist. Norman Malcolm detects two arguments for a necessary being in Anselm’s *Proslogion*. Malcolm argues that, although the first is fallacious, the second is valid and sound, and may even be of use in bringing unbelievers to faith in God.

Suggestions for Further Reading

Adams, Robert M., *The Virtue of Faith* (New York and Oxford: Oxford University Press, 1987), chs 13 and 14: “Has It been Proved that All Real Existence is Contingent?” and “Divine Necessity.”

Carter, William R., *The Elements of Metaphysics* (Philadelphia, Penn.: Temple University Press, 1990), ch. 11: “God.”

Edwards, Paul, “Why,” in Edwards, ed., *Encyclopedia of Philosophy* (New York: Macmillan, 1967).

Gardner, Martin, *The Whys of a Philosophical Scrivener* (New York: Quill, 1983), chs 10, 11, 12, 13, and 20: “The Gods: Why I am Not a Polytheist,” “The All: Why I am Not a Pantheist,” “The Proofs: Why I do Not Believe God’s Existence can be Demonstrated,” “Faith: Why I am Not an Atheist,” and “Surprise: Why I Cannot Take the World for Granted.”

Hasker, William, *Metaphysics: Constructing a World View* (Downers Grove, Ill., and Leicester, UK: InterVarsity Press, 1983).

Jubien, Michael, *Contemporary Metaphysics* (Oxford: Blackwell, 1997), ch. 11: “Cosmology.”

METAPHYSICS: THE BIG QUESTIONS

- Munitz, Milton K., *The Mystery of Existence* (New York: New York University Press, 1974).
- Plantinga, Alvin, *God, Freedom, and Evil* (Grand Rapids, Mich.: Eerdmans, 1977), Part 2: "Natural Theology."
- Post, John F., *Metaphysics: a Contemporary Introduction* (New York: Paragon House, 1991), chs 4 and 7: "Why Does Anything at All Exist?" and "God."
- Smith, Quentin, and L. Nathan Oaklander, *Time, Change and Freedom: an Introduction to Metaphysics* (London: Routledge, 1995), Dialogues 4 and 10: "Eternity" and "God, Time and Freedom."
- Taylor, Richard, *Metaphysics* (Englewood Cliffs, N.J.: Prentice-Hall, 1992), ch. 11: "God."
- van Inwagen, Peter, *Metaphysics* (Boulder, Col.: Westview Press, 1993) part 2: "Why the World Is."

Is There an Answer?

48 The Problem of Being: Chapter 3 of *Some Problems of Philosophy**

William James

Schopenhauer on the Origin of the Problem

How comes the world to be here at all instead of the nonentity which might be imagined in its place? Schopenhauer's remarks on this question may be considered classical. "Apart from man," he says, "no being wonders at its own existence. When man first becomes conscious, he takes himself for granted, as something needing no explanation. But not for long; for, with the rise of the first reflection, that wonder begins which is the mother of metaphysics, and which made Aristotle say that men now and always seek to philosophize because of wonder – The lower a man stands in intellectual respects the less of a riddle does existence seem to him . . . but, the clearer his consciousness becomes the more the problem grasps him in its greatness. In fact the unrest which keeps the never stopping clock of metaphysics going is the thought that the non-existence of this world is just as possible as its existence. Nay more, we soon conceive the world as something the non-existence of which not only is conceivable but would indeed be preferable to its existence; so that our wonder passes easily into a brooding over that fatality which nevertheless could call such a world into being, and mislead the immense force that could produce and preserve it into an activity so hostile to its own interests. The philosophic wonder thus becomes a sad astonishment, and like the overture to Don Giovanni, philosophy begins with a minor chord."¹

One need only shut oneself in a closet and begin to think of the fact of one's being there, of one's queer bodily shape in the darkness (a thing to make children scream at, as Stevenson says), of one's fantastic character and all, to have the wonder steal over the detail as much as over the general fact of being, and to see that it is only familiarity that blunts it. Not only that *anything* should be, but that *this* very thing should be, is mysterious! Philosophy stares, but brings no reasoned solution, for from nothing to being there is no logical bridge.

* From William James, *Some Problems of Philosophy* (New York: Longmans, Green, 1911).

Various Treatments of the Problem

Attempts are sometimes made to banish the question rather than to give it an answer. Those who ask it, we are told, extend illegitimately to the whole of being the contrast to a supposed alternative non-being which only particular beings possess. These, indeed, were not, and now are. But being in general, or in some shape, always was, and you cannot rightly bring the whole of it into relation with a primordial nonentity. Whether as God or as material atoms, it is itself primal and eternal. But if you call any being whatever eternal, some philosophers have always been ready to taunt you with the paradox inherent in the assumption. Is past eternity completed? they ask: If so, they go on, it must have had a beginning; for whether your imagination traverses it forwards or backwards, it offers an identical content or stuff to be measured; and if the amount comes to an end in one way, it ought to come to an end in the other. In other words, since we now witness its end, some past moment must have witnessed its beginning. If, however, it had a beginning, when was that, and why?

You are up against the previous nothing, and do not see how it ever passed into being. This dilemma, of having to choose between a regress which, although called infinite, has nevertheless come to a termination, and an absolute first, has played a great part in philosophy's history.

Other attempts still are made at exorcising the question. Non-being is not, said Parmenides and Zeno; only being is. Hence what is, is necessarily being – being, in short, is necessary. Others, calling the idea of nonentity no real idea, have said that on the absence of an idea can no genuine problem be founded. More curiously still, the whole ontological wonder has been called diseased, a case of *Grübelnsucht* like asking, "Why am I myself?" or "Why is a triangle a triangle?"

Rationalist and Empiricist Treatments

Rationalistic minds here and there have sought to reduce the mystery. Some forms of being have been deemed more natural, so to say, or more inevitable and necessary than others. Empiricists of the evolutionary type – Herbert Spencer seems a good example – have assumed that whatever had the least of reality, was weakest, faintest, most imperceptible, most nascent, might come easiest first, and be the earliest successor to nonentity. Little by little the fuller grades of being might have added themselves in the same gradual way until the whole universe grew up.

To others not the minimum, but the maximum of being has seemed the earliest First for the intellect to accept. "The perfection of a thing does not keep it from existing," Spinoza said, "on the contrary, it finds its existence."² It is mere prejudice to assume that it is harder for the great than for the little to be, and that easiest of all it is to be nothing. What makes things difficult in any line is the alien obstructions that are met with, and the smaller and weaker the thing the more powerful over it these become. Some things are so great and inclusive that to be is implied in their very nature. The anselmian or ontological proof of

God's existence, sometimes called the cartesian proof, criticised by Saint Thomas, rejected by Kant, re-defended by Hegel, follows this line of thought. What is conceived as imperfect may lack being among its other lacks, but if God, who is expressly defined as *Ens perfectissimum*, lacked anything whatever, he would contradict his own definition. He cannot lack being therefore: He is *Ens necessarium*, *Ens realissimum*, as well as *Ens perfectissimum*.³

Hegel in his lordly way says: "It would be strange if God were not rich enough to embrace so poor a category as Being, the poorest and most abstract of all." This is somewhat in line with Kant's saying that a real dollar does not contain one cent more than an imaginary dollar. At the beginning of his logic Hegel seeks in another way to mediate nonentity with being. Since "being" in the abstract, mere being, means nothing in particular, it is indistinguishable from "nothing"; and he seems dimly to think that this constitutes an identity between the two notions, of which some use may be made in getting from one to the other. Other still queerer attempts show well the rationalist temper. Mathematically you can deduce 1 from 0 by the following process:

$$\frac{0}{0} = \frac{1-1}{1-1} = 1.$$

Or physically if all being has (as it seems to have) a "polar" construction, so that every positive part of it has its negative, we get the simple equation: + 1 - 1 = 0, *plus* and *minus* being the signs of polarity in physics.

The Same Amount of Existence Must be Begged by All

It is not probable that the reader will be satisfied with any of these solutions, and contemporary philosophers, even rationalistically minded ones, have on the whole agreed that no one has intelligibly banished the mystery of *fact*. Whether the original nothing burst into God and vanished, as night vanishes in day, while God thereupon became the creative principle of all lesser beings; or whether all things have foisted or shaped themselves imperceptibly into existence, the same amount of existence has in the end to be assumed and begged by the philosopher. To comminute the difficulty is not to quench it. If you are a rationalist you beg a kilogram of being at once, we will say; if you are an empiricist you beg a thousand successive grams; but you beg the same amount in each case, and you are the same beggar whatever you may pretend. You leave the logical riddle untouched, of how the coming of whatever is, came it all at once, or came it piecemeal, can be intellectually understood.⁴

Conservation vs. Creation

If being gradually *grew*, its quantity was of course not always the same, and may not be the same hereafter. To most philosophers this view has seemed absurd,

neither God, nor primordial matter, nor energy being supposed to admit of increase or decrease. The orthodox opinion is that the quantity of reality must at all costs be conserved, and the waxing and waning of our phenomenal experiences must be treated as surface appearances which leave the deeps untouched.

Nevertheless, within experience, phenomena come and go. There are novelties; there are losses. The world seems, on the concrete and proximate level at least, really to grow. So the question recurs: How do our finite experiences come into being from moment to moment? By inertia? By perpetual creation? Do the new ones come at the call of the old ones? Why do not they all go out like a candle?

Who can tell off-hand? The question of being is the darkest in all philosophy. All of us are beggars here, and no school can speak disdainfully of another or give itself superior airs. For all of us alike, Fact forms a datum, gift, or *Vorgefundenes*, which we cannot burrow under, explain or get behind. It makes itself somehow, and our business is far more with its What than with its Whence or Why.

Notes

- 1 Schopenhauer, *The World as Will and Representation*: Appendix 17, “On the metaphysical need of man,” abridged.
- 2 Spinoza, *Ethics*, part i, prop. xi, scholium.
- 3 St Anselm, *Proslogion*, in *Anselm’s Basic Writings*, trs. S. N. Deane, with an introduction by Charles Harshorne, 2nd edn (LaSalle, Ill.: Open Court, 1962); Descartes, *Meditations on First Philosophy*, in *The Philosophical Writings of Descartes*, vol. II, trs. John Cottingham, Robert Stoothoff, and Dugald Murdoch (Cambridge: Cambridge University Press, 1984), Third and Fifth Meditations; Immanuel Kant, *The Critique of Pure Reason*, trs. Norman Kemp Smith (New York: St Martin’s Press, 1929), pp. 500–7.
- 4 In more technical language, one may say that fact or being is “contingent,” or matter of “chance,” so far as our intellect is concerned. The conditions of its appearance are uncertain, unforeseeable, when future, and when past, elusive.

49 The Puzzle of Reality: Why does the Universe Exist?*

Derek Parfit

It might have been true that nothing ever existed: no minds, no atoms, no space, no time. When we imagine this possibility, it can seem astonishing that

*From *Times Literary Supplement*, July 3, 1992, pp. 3–5. Reprinted by permission of the author.

anything exists. Why is there a universe? And things might have been, in countless ways, different. So why is the Universe as it is?

These facts cannot be causally explained. No law of nature could explain why there are any laws of nature, or why these laws are as they are. And, if God created the world, there cannot be a causal explanation of why God exists.

Since our questions cannot have causal answers, we may wonder whether they make sense. But there may be other kinds of answer.

Consider, first, a more particular question. Many physicists believe that, for stars, planets and life to be able to exist, the initial conditions in the Big Bang had to be precisely as they were. Why were these conditions so precisely right? Some say: 'If they had not been right, we couldn't even ask this question.' But that is no answer. It could be baffling how we survived some crash even though, if we hadn't, we could not be baffled.

Others say: 'There had to be some initial conditions, and those conditions were as likely as any others. So there is nothing to be explained.' To see what is wrong with this reply, we must distinguish two kinds of case. Suppose that, of a million people facing death, only one can be rescued. If there is a lottery to pick this one survivor, and I win, I would be very lucky. But there would be nothing to be explained. Someone had to win, and why not me? Consider next a second lottery. Unless my gaoler picks the longest of a million straws, I shall be beheaded. If I win this lottery, there *would* be something to be explained. It would not be enough to say, 'That result was as likely as any other.' In the first lottery, nothing special happened: whatever the result, someone's life would be saved. In this second lottery, the result *was* special. Of the million possible results, only one would save a life. Why was *this* what happened? Though this might be a coincidence, the chance of that is only one in a million. I could be almost certain that this lottery was rigged.

The Big Bang, it seems, was like the second lottery. 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?

On one view, this was a mere coincidence. That is conceivable, but most unlikely. On some estimates, the chance is below one in a billion billion. Others say: 'The Big Bang *was* fine-tuned. It is not surprising that God chose to make life possible.' We may be tempted to dismiss this answer, thinking it improbable that God exists. But should we put the chance as low as one in a billion billion? If not, this is a better explanation.

There is, however, a rival explanation. 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-tuning. If there were enough Big Bangs, it would be no surprise that, in a

few of these, conditions were just right for life. And it would be no surprise that our Big Bang was one of these few.

On most versions of this theory, these many worlds are not causally related, and each has its own space and time. Some object that, since our world could not be affected by such other worlds, we have no reason to believe in them. But we do have such a reason, since their existence would explain an otherwise puzzling feature of our world: the appearance of fine-tuning.

How should we choose between these explanations? The many-worlds hypothesis is more cautious, since it merely claims that there is more of the kind of reality we know. But God's existence has been claimed to be intrinsically more plausible. By 'God' we mean a being who is omnipotent, omniscient and wholly good. The existence of such a being has been claimed to be both simpler, and less arbitrary, than the existence of many complicated and specific worlds.

If such a God exists, however, why is the Universe as it is? It may not be surprising that God chose to make life possible. But the laws of nature could have been different, so there are many possible worlds that would have contained life. It is hard to understand why, with all these possibilities, God chose to create *our* world. The greatest difficulty here is the problem of evil. There appears to be suffering which any good person, knowing the truth, would have prevented if he could. If there is such suffering, there cannot be a God who is omnipotent, omniscient and wholly good.

One response to this problem is to revise our view of God. Some suggest that God is not omnipotent. But, with that revision, the hypothesis that God exists becomes less plausible. How could there be a being who, though able to create our world, cannot prevent such suffering? Others believe in a god who, whatever he is called, is not good. Though that view more easily explains the character of life on Earth, it may seem in other ways less credible.

As we shall see, there may be other answers to this problem. But we have larger questions to consider. I began by asking why things are as they are. We must also ask *how* things are. There is much about our world that we have not discovered. And, just as there may be other worlds like ours, there may be worlds that are very different.

It will help to distinguish two kinds of possibility. For each particular kind of possible world, there is the *local* possibility that such a world exists. If there is such a world, that leaves it open whether there are also other worlds. *Global* possibilities, in contrast, cover the whole Universe, or everything that ever exists. One global possibility is that *every* conceivable world exists. That is claimed by the *all-worlds hypothesis*. Another possibility, which might have obtained, is that nothing ever exists. This we can call the *Null Possibility*. In each of the remaining possibilities, the number of possible worlds that exist is between none and all. There are countless of these possibilities, since there are countless combinations of particular possible worlds.

Of these different global possibilities, one must obtain, and only one can obtain. So we have two questions. Which obtains, and why? These questions are connected. If some possibility would be less puzzling, or easier to explain, we have more reason to think that it obtains. That is why, rather than believing

that the Big Bang merely happened to be right for life, we should believe either in God or in many worlds.

Is there some global possibility whose obtaining would be in no way puzzling? That might be claimed of the Null Possibility. It might be said that, if no one had ever existed, no one would have been puzzled. But that misunderstands our question. Suppose that, in a mindless and finite Universe, an object looking like the *Times Literary Supplement* spontaneously formed. Even with no one to be puzzled, that would be, in the sense I mean, puzzling. It may next be said that, if there had never been anything, there wouldn't have been anything to be explained. But that is not so. When we imagine that nothing ever existed, what we imagine away are such things as minds and atoms, space and time. There would still have been truths. It would have been true that nothing existed, and that things might have existed. And there would have been other truths, such as the truth that 27 is divisible by 3. We can ask why these things would have been true.

These questions may have answers. We can explain why, even if nothing had ever existed, 27 would have been divisible by 3. There is no conceivable alternative. And we can explain the non-existence of such things as two-horned unicorns, or spherical cubes. Such things are logically impossible. But why would *nothing* have existed? Why would there have been no stars or atoms, no minds or bluebell woods? How could *that* be explained?

We should not claim that, if nothing had existed, there would have been nothing to be explained. But we might claim something less. Perhaps, of all the global possibilities, this would have needed the least explanation. It is much the simplest. And it seems the easiest to understand. When we imagine there never being anything, that does not seem, as our own existence can, astonishing.

Here, for example, is one natural line of thought. It may seem that, for any particular thing to exist, its existence must have been caused by other things. If that is so, what could have caused them *all* to exist? If there were an infinite series of things, the existence of each might be caused by other members of that series. But that could not explain why there was this whole series, rather than some other series, or no series. In contrast, the Null Possibility raises no such problem. If nothing had ever existed, that state of affairs would not have needed to be caused.

Even if this possibility would have been the easiest to explain, it does not obtain. Reality does not take its simplest and least puzzling form.

Consider next the all-worlds hypothesis. That may seem the next least puzzling possibility. For one thing, it avoids arbitrary distinctions. If only one world exists, we have the question: 'Out of all the possible worlds, why is *this* the one that exists?' On the many-worlds hypothesis, we have the question: 'Why are *these* the ones?' But, if *all* possible worlds exist, there is no such question. Though the all-worlds hypothesis avoids that question, it is not as simple as it seems. Is there a sharp distinction between those worlds that are and are not possible? Must all worlds be governed by natural laws? Does each kind of world exist only once? And there are further complications.

Whichever global possibility obtains, we can ask why it obtains. All that I

have claimed so far is that, with some possibilities, this question would be less puzzling. We should now ask: Could this question have an answer? Is there a theory that leaves nothing unexplained?

On one kind of view, it is logically necessary that God, or the whole Universe, exists. Though it may seem conceivable that there might never have been anything, that is not really logically possible. Some people even claim that there is only one coherent global possibility. If such a view were true, everything would be explained. But the standard objections to such views, which I shall not repeat, seem to me convincing.

Others claim that the Universe exists because its existence is good. This is the Platonic, or Axiarchic View. Even if we think this view absurd, it is worth asking whether it makes sense. That may suggest other possibilities.

The Axiarchic View can take a theistic form. It can claim that God exists because His existence is good, and that the rest of the Universe exists because God caused it to exist. But in that explanation God is redundant. If God can exist because His existence is good, so can the whole Universe.

In its simplest form, the Axiarchic View makes three claims: (1) It would be best if reality were a certain way. (2) Reality is that way. (3) (1) explains (2).

(1) is an ordinary evaluative claim, like the claim that it would be better if there was no pointless suffering. The Axiarchic View assumes, in my opinion correctly, that such claims can be true. (2) is an ordinary descriptive claim, though of a sweeping kind. What is distinctive in this view is claim (3).

Can we understand (3)? To focus on this question, we should briefly ignore the world's evils. Suppose that, as Leibniz claimed, the best possible Universe exists. Could this Universe exist *because* it is the best? That question might be confused with another. If God intentionally created the best possible world, that world would exist because it is the best. But, though God would not be part of the world that He creates, He would be part of the Universe, or the totality of what exists. And God cannot have created Himself. So an appeal to God cannot explain why the best Universe exists.

Axiarchists make a different claim. On their view, that there is a best way for reality to be explains *directly* why reality is that way. If God exists, that is because His existing is best. Truths about value are, in John Leslie's phrase, *creatively effective*.

This cannot be an ordinary causal claim. Ordinary causes are particular events, or facts about existing things. But the Axiarchic claim may have some of the meaning of an ordinary causal claim.

When we believe that X caused Y, we usually believe that, without an X, there would have been no Y. A spark caused an explosion if, without a spark, there would have been no explosion. Axiarchists might make a similar claim. They might say that, if it had not been best if reality were a certain way, reality would not have been that way. But such a claim may not help to explain the Axiarchic View, since what it asks us to imagine could not have been true. Just as pointless suffering could not have been good, the best way for reality to be could not have failed to be the best.

In defending a causal claim, we may also appeal to a generalization. Certain

conditions cause an explosion if, whenever there are such conditions, there is an explosion. It may seem that, with only one Universe, Axiarchists cannot appeal to a generalization. But that is not so. They could say that, whenever it would be better if the Universe had some particular feature, it *has* that feature.

Would that explain their claim that this is *why* the Universe has these features? That use of ‘why’ may seem utterly mysterious. But we should remember that even ordinary causation is mysterious. At the most fundamental level, we have no idea why some events cause others. And it is hard to explain what causation is.

Axiarchy can be best explained as follows. We are now assuming that, of all the countless ways that reality might be, one is both the very best, and is the way that reality is. On the Axiarchic View, *that is no coincidence*. That claim makes, I believe, some kind of sense. And, on those assumptions, it would be a reasonable conclusion.

Compared with the appeal to God, the Axiarchic View has one advantage. God cannot have settled *whether*, as part of the best Universe, He himself exists, since He can only settle anything if He does exist. But even if nothing had ever existed, it would still have been true that it would be best if the best Universe existed. So that truth might explain why this Universe exists.

The main objection to this view is the problem of evil. Our world appears to be flawed.

If we appeal to a variant of the many-worlds hypothesis, this objection can be partly met. Perhaps, in the best Universe, *all* good possible worlds exist. We would then avoid the question why things are not much better than they are. Things *are*, on the whole, much better. They are better elsewhere.

Why are they not *also* better here? One answer might be as follows. If it is best that all good worlds exist, that implies that, even in the best Universe, many worlds would not be very good. Some would be only just good enough. Perhaps our world is one of these. It would then be good that our world exists, since a good niche is thereby filled. And we might be able to explain why our world is not better than it is. The Louvre would be a worse collection if its less good paintings were turned into copies of the *Mona Lisa*. In the same way, if our world were in itself better, reality as a whole might be less good. Since every other good niche is already filled, our world would then be a mere copy of some other world, and one good niche would be left unfilled.

Even on this view, however, each world must be good enough. The existence of each world must be better, even if only slightly, than its non-existence. Can this be claimed of our world? It would be easier to make that claim on a broadly Utilitarian view. Our world’s evils might then be outweighed by what is good. But, on some principles of justice, that would not be enough. If innocent beings suffer, in lives that are not worth living, that could not be morally outweighed by the happiness of other beings. For our world to be good enough, there must be future lives in which the sufferings of each being could, in the end, be made good. Even the burnt fawn in the forest fire must live again. Or perhaps these different beings are, at some level, one.

These replies may seem too weak. We may doubt that our world could be even the least good part of the best possible Universe.

If we reject the Axiarchic View, what conclusion should we draw? Is the existence of our world a mere brute fact, with no explanation? That does not follow. If we abstract from the optimism of this view, its claims are these. One global possibility has a special feature, this is the possibility that obtains, and it obtains because it has this feature. Other views can make such claims.

Suppose that our world were part of the worst possible Universe. Its bright days may only make its tragedies worse. If reality were as bad as it could be, could we not suspect that this was no coincidence?

Suppose next, more plausibly, that all possible worlds exist. That would also be grim, since the evil of the worst worlds could hardly be outweighed. But that would be incidental. If every conceivable world exists, reality has a different distinctive feature. It is *maximal*: as full and varied as it could possibly be. If this is true, is it a coincidence? Does it merely happen to be true that, of all the countless global possibilities, the one that obtains is at this extreme? As always, that is conceivable. Coincidences can occur. But it seems hard to believe. We can reasonably assume that, if all possible worlds exist, that is *because* that makes reality as full as it could be.

Similar remarks apply to the Null Possibility. If there had never been anything, would that have been a coincidence? Would it have merely happened that, of all the possibilities, what obtained was the *only* possibility in which nothing exists? That is also hard to believe. Rather, if this possibility had obtained, that would have been because it had that feature.

Here is another special feature. Perhaps reality is as it is because that makes its fundamental laws as mathematically beautiful as they could be. That is what many physicists believe.

If some possibility obtains because it has some feature, that feature selects what reality is like. Let us call it the *Selector*. A feature is a *plausible* Selector if we can reasonably believe that, were reality to have that feature, that would not merely happen to be true.

There are countless features which are not plausible Selectors. Suppose that fifty-seven worlds exist. Like all numbers, 57 has some special features. For example, it is the smallest number that is the sum of seven primes. But that could hardly be *why* that number of worlds exist.

I have mentioned certain plausible Selectors. A possibility might obtain because it is the best, or the simplest, or the least arbitrary, or because it makes reality as full as it could be, or because its fundamental laws are as elegant as they could be. There are, I assume, other such features, some of which we have yet to discover.

For each of these features, there is the *explanatory* possibility that this feature is the Selector. That feature then explains why reality is as it is. There is one other, special explanatory possibility: that there is *no* Selector. This is like the global possibility that nothing exists. If there is no Selector, it is random that reality is as it is. Events may be in one sense random, even though they are causally inevitable. That is how it is random whether a meteorite strikes the land

or the sea. Events are random in a stronger sense if they have no cause. That is what most physicists believe about some facts at the quantum level, such as how some particles move. If it is random what reality is like, the Universe would not only have no cause. It would have no explanation of any kind. This we can call the *Brute Fact View*.

On this view, we should not expect reality to have very special features, such as being maximal, or best, or having very simple laws, or including God. In much the largest range of the global possibilities, there would exist an arbitrary set of messily complicated worlds. That is what, with a random selection, we should expect. It is unclear whether ours is one such world.

The Brute Fact View may seem hard to understand. It may seem baffling how reality could be even randomly selected. What kind of *process* could select whether time had no beginning, or whether anything ever exists? But this is not a real problem. It is logically necessary that one global possibility obtains. There is no conceivable alternative. Since it is necessary that one possibility obtains, it is necessary that it be settled which obtains. Even without any kind of process, logic ensures that a selection is made. There is no need for hidden machinery.

If reality were randomly selected, it would not be mysterious *how* the selection is made. It would be in one sense inexplicable why the Universe is as it is. But this would be no more puzzling than the random movement of a particle. If a particle can simply happen to move as it does, it could simply happen that reality is as it is. Randomness may even be *less* puzzling at the level of the whole Universe, since we know that facts at this level could not have been caused.

There would, however, be a further question. If there is no explanation why reality is as it is, why is *that* true?

Some reply that this, too, is logically necessary. On their view, the nature of the Universe must be a mere brute fact, since it could not conceivably be explained. But, as I have argued, that is not so. Though it is logically necessary that one global possibility obtain, it is not necessary that it be random which obtains. There are other explanatory possibilities.

Since it is not necessary that there be no explanation why reality is as it is, that truth might be another brute fact. There may be no explanation why there is no explanation. Perhaps both simply happen to be true. But why would *that* be true? Would it, too, simply happen to be true? And why should we accept this view? If it was randomly selected *whether* reality was selected randomly, and there are several other possibilities, why expect random selection to have been selected? Unless we can explain *why* it is random what reality is like, we may have no reason to believe that this *is* random.

Return now to the other explanatory possibilities. Each raises the same further question. Whichever possibility obtains, we can ask why it obtains. Consider first the Axiarchic View. Suppose that the best Universe exists because it is the best. Why is that true? Even if this view is true, its falsehood is at least logically conceivable. It may seem that Axiarchy could explain itself. On this view, claims about reality are true because their being true is best. It might be best if this view were true. Could that be why it *is* true? That is not possible. Even if this view is true, its being true could not be explained by its being true.

Just as God cannot have caused His own existence, the truth of the Axiarchic View cannot be what makes this view true.

Consider next the Maximalist View. Suppose that all possible worlds exist, and that this is no coincidence. Suppose these worlds all exist because that makes reality as full as it could be. If that is true, why is it true? Perhaps this truth makes reality even more maximal. But, as before, this truth could not explain itself.

A similar claim may apply to every view. As we have seen, it is not logically necessary that, of the *global* possibilities, it is random which obtains. This possibility might be selected in other ways. But it may be logically necessary that, of the *explanatory* possibilities, it is random which obtains. Perhaps nothing could select between all the possible Selectors. If that were so, it would not be mysterious that a particular explanatory claim simply happened to be true. The randomness would be fully explained, since there would be no conceivable alternative.

It may be objected that, if some claim simply happens to be true, it cannot provide an explanation. Such a claim may seem to add nothing. To illustrate this objection, return to the Maximalist View. Consider first two global possibilities: (1) Only our world exists. (2) Every conceivable world exists. These possibilities are very different. Suppose next that (2) is true. There are then two explanatory possibilities. On the Brute Fact View, (2) simply happens to be true. On the Maximalist View, (2) is true because that makes reality as full as it could be. Here again, these seem to be different possibilities. But we are now supposing that, even if the Maximalist View is true, its truth is a brute fact, with no explanation. We may think that, if that is so, the Maximalist View could not *explain* (2). If this view simply happens to be true, it may seem not to differ from the Brute Fact View.

That reaction is a mistake. On the Brute Fact View, (2) would involve an extreme coincidence. There are countless global possibilities, and most of these, unlike (2), have no very special feature. It is hard to believe that, of this vast range of possibilities, it simply happens to be true that every conceivable world exists. That is implausible because, at this level, there is an alternative. If the Maximalist View is true, the existence of all these worlds is no coincidence. At the next level, things are different. Of the plausible explanatory possibilities, all have special features. There is no possibility whose obtaining would be a coincidence. And, as we have seen, it may be logically necessary that, of *these* possibilities, one simply happens to obtain. At this level, there may be no alternative. It would then be in no way puzzling if the Maximalist View simply happens to be true.

We should not claim that, if an explanation rests on a brute fact, it is not an explanation. Scientific explanations all take this form. But we might claim something less. Any such explanation may, in the end, be merely a better description.

If that is true, there is a different answer. Even to discover how things are, we need explanations. And we may need explanations on the grandest scale. Our world may seem to have some feature that would be unlikely to be a coincidence. We might reasonably suspect that our world exists, not as a brute fact,

but because it has this feature. That hypothesis might lead us to confirm that, as it seemed, our world does have this feature. We might then reasonably conclude either that ours is the only world, or that there are many other worlds, with the same or related features. We might reach truths about the whole Universe.

Even if all explanations must end with a brute fact, we should go on trying to explain why the Universe exists, and is as it is. The brute fact may not enter at the lowest level. If the Universe exists because it has some feature, to know *what* reality is like, we must ask *why*.

Acknowledgements

Of the many recent discussions of these questions, I owe most to John Leslie's *Value and Existence* (1979) and *Universes* (1989), and to Robert Nozick's *Philosophical Explanations* (1981); then to *The Existence of God* by Richard Swinburne (1979), *The Miracle of Theism* by John Mackie (1982), Peter Unger's article in *Midwest Studies in Philosophy*, volume 9 (1989), and some unpublished work by Stephen Grover.

50 Response to Derek Parfit

Richard Swinburne

Derek Parfit is right to suppose that, on (what I take to be) his understanding of 'causal explanation' and of 'the Universe', there cannot be a causal explanation of the existence of the Universe. He apparently understands by 'the Universe' all the substances there are (that is, all the material things – stars and atoms and whatever these are made of – and all the immaterial things, such as souls or God – if these exist). He apparently understands by 'causal explanation', the causing of some event (including the coming-into-existence and continuing-in-existence of substances) by some substance. Since nothing can cause itself to exist, no substance could cause all-the-substances (including the former) to exist.

What, however, is possible is that one substance causes all the others to come into existence and continue in existence. I believe that the basic principles of inductive inference, which we use in science, historical inquiry, detective work and all other rational inquiry, have the consequence that on the evidence of observed events E, it is probable that C (where C is some substance or law or anything else) in so far as: (1) C (if it existed) would make E likely to occur; (2) if C did not exist, E would be less likely to occur; and (3) C is a simple entity (or law). I believe, and have argued at length elsewhere,¹ that where E is the observed universe (including its life-producing features, to which Parfit draws attention) and C is God, postulated as the cause of the Universe (one substance,

with zero limits to his power, knowledge and freedom), E makes the existence of C probable. (As Parfit emphasizes, someone who gives this answer needs to explain why God allows suffering to occur.) To postulate one God as cause is immensely simpler than to postulate infinitely many worlds (most of which are not life-producing) in order to explain the occurrence of our life-producing universe. A simple explanation postulates no more entities than are needed to explain the phenomena. Of course postulating God as the cause of the Universe does not explain why God exists; but then, as Parfit acknowledges, in the end there must be some ultimate brute fact (whether law or substance), and I would argue that the existence of God is the existence of the simplest substance there could be.

Parfit has, however, floated the interesting suggestion that there might be an explanation of the existence of the Universe which is not a causal explanation – some ultimate principle or law which might somehow produce a Universe, without the action of a substance. The trouble is that there are no plausible cases of real-life principles which produce effects within the universe without doing so by operating via substances. If some law of nature, say Newton's law of gravity, produces some effect (say that a stone falls to Earth), it always does so by determining how some substance will cause that effect – say, determining that the Earth will attract the stone in a certain way. Indeed, I suggest that all talk about laws of nature is reducible to talk about the powers which substances have, and the liabilities which they have to exercise them.

It is sometimes suggested that some law of Quantum Theory has the consequence that vacua will produce substances from time to time. But on investigation it turns out that 'vacua' are not nothing, but themselves rather special sorts of substance. Parfit suggests that there might be axiomatic principles, which produce events because it is good to do so. But there are no plausible examples of such principles at work in the world. When food appears on the tables of the hungry, it does not appear there because it is good that it should, but because some person (i.e. a substance) caused it to be there because he thought that it was good that it should. Nor is there operative any principle of simplicity which makes things occur because they are in some way simple – e.g. makes the laws of nature what they are because they are the simplest laws there could be. For it is easy enough to conceive of laws of nature a lot simpler than our actual laws, which are perhaps the laws of Grand Unified Field Theory, or some laws even more complicated. Certainly, as mentioned earlier, we judge that the simplest theory *compatible with observed events* is more probably the true theory than is any other one. But that is a criterion for assessing the force of evidence, not for producing what exists. If simplicity dictated what was to exist, there would be nothing, or at any rate a lot fewer things behaving in a lot simpler ways than there are. So Parfit's suggestion that there might be some non-causal explanation of the existence of the Universe involves his claiming that there is some kind of principle at work in producing the Universe, which is never operative in producing more limited effects within the Universe. But then we have absolutely no reason for supposing that that kind of principle is ever at work, or that such a principle explains anything at all.² By contrast, the theist who postulates

God as the cause of (the rest of) the Universe postulates a substance who acts intentionally – i.e. brings about some effect because he believes it good to do so. And the universe is full of many other substances including humans who bring about many different effects intentionally. In this respect explanation by God's intentional actions is like explanations by the intentional actions of humans. Of course God is supposed to be very different indeed in the extent of his power, knowledge and freedom from other substances with which we are familiar. But they are also different from each other in these respects. And God is not supposed to be *totally* different from humans. (In the traditional view, humans are made in God's 'image'.) But to postulate axiomatic or similar principles bringing something out of nothing is to postulate a totally different kind of explanation which we have no reason at all to suppose ever to operate.

Notes

- 1 See my *The Existence of God* (Oxford: Clarendon Press, revised edition, 1990); or the simpler *Is There a God?* (Oxford: Oxford University Press, 1996).
- 2 In the terms used earlier our observed E adds no probability to the claim that there is a C of this kind at work, because if such a principle operated in producing E and so such principles were among the explanations of things, one might expect E (which includes things producing other things) to include things produced by the operation of more limited such principles.

Does the Answer Involve a Necessary Being?

51 The Cosmological Argument and the Principle of Sufficient Reason*

William L. Rowe

The Cosmological Argument began with Plato, flourished in the writings of Aquinas, Leibniz, and Samuel Clarke, and was laid to rest by Hume and Kant. Although I think its death premature, if not unjustified, I shall not here attempt its resurrection. What I have in mind is more in the nature of an autopsy. I wish to uncover, clarify, and examine some of the philosophical concepts and theses essential to the reasoning exhibited in the Cosmological Argument. . . .

The Cosmological Argument is an argument for the existence of *God*. As such, the argument has two distinct parts. The first part is an argument to establish the existence of a necessary being, a being that carries the reason of its existence within itself. The second part is an argument to establish that this necessary being is God. A good deal of philosophical criticism has been directed against the first part of the argument. Much less has been directed against the second part. Indeed, some philosophers seem not to have realized that the argument has a second part. For example, in Part IX of his *Dialogues Concerning Natural Religion* Hume has Demea present a summary of only the first part of the Cosmological Argument. Demea appears to *assume* that a necessary being would be God. Thus, after concluding that there exists a necessary being, he simply remarks, “There is consequently such a Being, that is, there is a Deity.” But, of course, it is not at all obvious that the necessary being is a Deity. Indeed, Cleanthes quickly asks, “Why may not the material universe be the necessarily existent Being?” Hence, as an argument for the existence of God, the Cosmological Argument not only does but must contain a second part in which it is argued that the necessary being possesses the properties – omnipotence, infinite goodness, infinite wisdom, etc. – that God, and only God, possesses.

Using the expression “dependent being” to mean “a being that has the reason of its existence in the causal efficacy of some other being,” and the expression “independent being” to mean “a being that has the reason of its existence

* From “The Cosmological Argument and the Principle of Sufficient Reason”, *Man and World*, 1 (1968), pp. 278–92. Reprinted with kind permission of the author and Kluwer Academic Publishers.

within its own nature," we may state the argument for the existence of a necessary being (i.e., the first part of the Cosmological Argument) as follows:

1. Whatever exists is either a dependent being or an independent being; therefore,
2. Either there exists an independent being or every being is dependent;
3. It is false that every being is dependent; therefore,
4. There exists an independent being; therefore,
5. There exists a necessary being.¹

This argument consists of two premises – propositions (1) and (3) – and three inferences. The first inference is from (1) to (2), the second from (2) and (3) to (4), and the third inference is from (4) to (5). Of the premises neither is obviously true, and of the inferences only the first and second are above suspicion. Before discussing the main subject of this paper – namely, proposition (1) and its connection with the Principle of Sufficient Reason – I want to describe the argument in support of premise (3) and the main criticisms of that argument.

Why is it false that every being is dependent? Well, if every being that exists (or ever existed) is dependent then the whole of existing things, it would seem, consists of a collection of dependent beings, that is, a collection of beings each member of which exists by reason of the causal efficacy of some other being. Now this collection would have to contain an infinite number of members. For suppose it contained a *finite* number, let us say three, *a*, *b*, and *c*. Now if in Scotus's phrase "a circle of causes is inadmissible" then if *c* is caused by *b* and *b* by *a*, *a* would exist without a cause, there being no other member of the collection that could be its cause. But in that case *a* would not be what by supposition it is, namely a *dependent* being. Hence, if we grant that a circle of causes is inadmissible it is impossible that the whole of existing things should consist of a collection of dependent beings *finite* in number.

Let us suppose, then, that the dependent beings making up the collection are *infinite* in number. Why is it impossible that the whole of existing things should consist of such a collection? The proponent of the Cosmological Argument answers as follows.² The infinite collection *itself*, he argues, requires an explanation for its existence. For since it is true of each member of the collection that it might not have existed, it is true of the whole infinite collection that it might not have existed. But if the entire infinite collection might not have existed there must be some explanation for why it exists rather than not. The explanation cannot lie in the causal efficacy of some being outside of the collection since by supposition the collection includes every being that is or ever was. Nor can the explanation for why there is an infinite collection be found within the collection itself, for since no member of the collection is independent, has the reason of its existence within itself, the collection as a whole cannot have the reason of its existence within itself. Thus the conception of an infinite collection

of dependent beings is the conception of something whose existence has no explanation whatever. But since premise (1) tells us that whatever exists has an explanation for its existence, either within itself or in the causal efficacy of some other being, it cannot be that the whole of existing things consists of an infinite collection of dependent beings.

Two major criticisms have been advanced against this line of reasoning, criticisms which have achieved some degree of acceptance. According to the first criticism it *makes no sense* to apply the notion of cause or explanation to the totality of things, and the arguments used to show that the whole of existing things must have a cause or explanation are fallacious. Thus in his B.B.C. debate with Father Copleston, Bertrand Russell took the view that the concept of cause is inapplicable to the universe conceived as the total collection of things. When pressed by Copleston as to how he could rule out "the legitimacy of asking the question how the total, or anything at all comes to be there," Russell responded: "I can illustrate what seems to me your fallacy. Every man who exists has a mother, and it seems to me your argument is that therefore the human race must have a mother, but obviously the human race hasn't a mother – that's a different logical sphere."³ According to the second major criticism it is intelligible to ask for an explanation of the existence of the infinite collection of dependent beings. But the answer to this question, so the criticism goes, is provided once we learn that each member of the infinite collection has an explanation of its existence. Thus Hume remarks: "Did I show you the particular causes of each individual in a collection of twenty particles of matter, I should think it very unreasonable, should you afterwards ask me, what was the cause of the whole twenty. This is sufficiently explained in explaining the cause of the parts."⁴

Although neither criticism is, I think, decisive against the argument given in support of proposition (3), they do draw attention to two crucial steps in the Cosmological Argument. First, it seems that the infinite collection is itself viewed as an existing thing. For only if it is so viewed will it follow from premise (1) that it (the infinite collection) must have a cause or explanation of its existence. Second, the question why each member of the infinite collection exists is felt to be different from the question why the infinite collection exists. For the proponent of the argument admits that each member of the collection has an explanation of its existence – namely, in the causal efficacy of some other member – and yet denies that this explains the existence of the entire infinite collection.

Perhaps neither of these steps in the argument for proposition (3) is correct. But even if both steps are correct – that is, even if the infinite collection itself may be viewed as an object or thing, and even if to explain each member is not sufficient to explain the collection – it is important to note that it is premise (1) from which it is then inferred that there must be an explanation for the existence of the infinite collection. Thus proposition (1) plays a crucial role not only as a premise in the main argument but also as a premise in the argument for proposition (3). Having seen the crucial role that proposition (1) plays in the Cosmological Argument, we may now examine that proposition in some detail.

Proposition (1) tells us that *whatever exists* must have an explanation for its

existence. The explanation may lie either within the nature of the thing itself or in the causal efficacy of some other being. The claim that whatever exists must have an explanation of its existence I shall call the *strong form* of the Principle of Sufficient Reason. This is to be distinguished from the claim that *whatever comes into existence* must have an explanation of its existence. The latter claim I shall call the *weak form* of the Principle of Sufficient Reason. If we imagine a star that has existed from eternity, a star that never came into existence but has always existed, the strong form of the Principle of Sufficient Reason requires, whereas the weak form does not, that there be an explanation for the existence of that star. The Cosmological Argument, as we have seen, employs the strong form of the Principle of Sufficient Reason.

Can the Principle of Sufficient Reason be proved or otherwise known to be true? Some philosophers, it seems, thought that the Principle could be proved. Hume attributes the following argument to Locke.

- (1) If something exists without a cause, it is caused by nothing;
- (2) Nothing cannot be the cause of something;
therefore,
- (3) Whatever exists must have a cause.

About this “proof” Hume remarks:

It is sufficient only to observe that when we exclude all causes we really do exclude them, and neither suppose nothing nor the object itself to be the causes of the existence, and consequently can draw no argument from the absurdity of that exclusion. If everything must have a cause, it follows that upon the exclusion of other causes we must accept of the object itself or of nothing as causes. But it is the very point in question, whether everything must have a cause or not, and therefore, according to all just reasoning, it ought never to be taken for granted.⁵

It is clear from Hume’s comment that he rejects premise (1). For he takes the proponent of the argument to mean by premise (1) that if something exists without a cause it, nevertheless, has a cause – although in this case its cause will not be some other thing, it will be *nothing*. But there is a subtlety in this argument that Hume overlooks. In the natural sense of the expression “caused by nothing” it is *true* that if something exists without a cause it is caused by nothing – to be caused by nothing is simply not to be caused by any thing whatever. Taken in this way, premise (1) is true. Moreover, premise (2) is true as well. For to say that nothing cannot be the cause of something is simply to say that if something has a cause then there must be some *thing* which is its cause. But so interpreted, the premises, although true, do not yield the conclusion that everything has a cause. For from (1) if something exists without a cause then there is no thing which caused it, and (2) if something has a cause then there is a thing which caused it, it in no way follows that everything has a cause. Therefore, if the premises are interpreted so as to be clearly true, the argument is invalid; whereas, if the argument is to appear valid its first premise, as Hume points out,

is false or, at the very least, begs the question at issue. In either case the argument fails as a demonstration of the Principle of Sufficient Reason.

Of course, if, as seems likely, the Principle of Sufficient Reason cannot be – at least, has not been – demonstrated, it does not follow that it cannot be *known* to be true. Clearly, if we know any propositions to be true there must be some propositions which we can know to be true without having to *prove* them, without having to derive them from other propositions we know to be true. If this were not so, we would have to know an infinite number of propositions in order to know any proposition whatever. Hence, the fact, if it is a fact, that the Principle of Sufficient Reason cannot be demonstrated does not invalidate the view other philosophers seem to take; namely, that the Principle is a necessary truth, known *a priori*.⁶

If the Principle in its strong form is analytically true then the view of these philosophers – namely, that the Principle is a necessary truth, known *a priori*, is probably correct. For every analytically true proposition is necessary and, if known at all, presumably can be known by simply reflecting on it, without relying on empirical evidence. But is the Principle of Sufficient Reason analytically true? Clearly the Principle is not logically true. Nor, it would seem, does the mere notion of the existence of a thing *definitionally* contain the notion of a thing being caused. Kant argued – correctly, I think – that although the proposition “Every effect has a cause” is analytically true, “Every event has a cause” is not. The idea of an event, of something happening – a leaf falling, a chair collapsing, etc. – does not seem to contain the idea of something *causing* that event. If this is so then the Principle of Sufficient Reason is certainly not analytically true.

But if the Principle is not analytically true how can it be necessary? Indeed, can any proposition be necessary if it is not analytically true? Many philosophers have held that only analytically true propositions are necessary. But it is, I think, reasonable to argue, as some philosophers have, that, for example, the proposition “Whatever is red is colored” is necessary but not analytically true.⁷ For (i) we do not seem to have a *definition* of “red” or “colored” in terms of which the sentence “Whatever is red is colored” can be reduced to a sentence expressing a logical truth, and yet (ii) it certainly is *impossible* that something be red and not colored. Thus the proposition “Whatever is red is colored” may well be a synthetic, necessary proposition. Moreover, as Chisholm has argued, there seem to be reasons for the view that the proposition “Necessarily, whatever is red is colored” is known *a priori*. But even if this is correct, as I am inclined to think it is, it is far from clear that the Principle of Sufficient Reason is a synthetic, necessary proposition known *a priori*.

The difficulty with the view that the Principle, in either its strong or weak form, is *necessary* is that we do seem able to conceive of things existing, or even of things coming into existence, without having to conceive of those things as having an explanation or cause. Unlike the proposition “Some red things are not colored,” it does seem conceptually possible that something should exist and yet have no cause or explanation of its existence. As Hume remarks, “The separation, therefore, of the idea of a cause from that of a beginning of existence is plainly possible for the imagination, and consequently the actual separa-

tion of those objects is so far possible that it implies no contradiction nor absurdity . . ."⁸ Indeed, not only does the denial of the Principle seem to be possible, philosophers have held that the denial of the Principle is *true*.

. . . many philosophers have maintained that it is not true that everything that exists, or even that everything that has a beginning, has a cause, that is to say, is an effect. The world, they say, contains "spontaneous", free, or uncaused and unoriginated events. In any case they assert very positively that there is no way of proving that such uncaused events do not occur.⁹

In view of this and other difficulties, some contemporary defenders of the Cosmological Argument have retreated from the view that the Principle of Sufficient Reason is a synthetic, necessary proposition known *a priori*. Instead, they have adopted the somewhat more modest view that the Principle is a *metaphysical assumption*, a presupposition we are forced to make in order to make sense of our world. Thus, for example, Father Copleston, in his B.B.C. debate with Russell, argued that something like the Principle of Sufficient Reason is presupposed by science. "I cannot see how science could be conducted on any other assumption than that of order and intelligibility in nature."¹⁰ Another contemporary philosopher, Richard Taylor, has expressed this view as follows:

The principle of sufficient reason can be illustrated in various ways, as we have done, and if one thinks about it, he is apt to find that he presupposes it in his thinking about reality, but it cannot be proved. It does not appear to be itself a necessary truth, and at the same time it would be most odd to say it is contingent. If one were to try proving it, he would sooner or later have to appeal to considerations that are less plausible than the principle itself. Indeed, it is hard to see how one could even make an argument for it, without already assuming it. For this reason it might properly be called a presupposition of reason itself. One can deny that it is true, without embarrassment or fear of refutation, but one is apt to find that what he is denying is not really what the principle asserts. We shall, then, treat it here as a datum – not something that is provably true, but as something which all men, whether they ever reflect upon it or not, seem more or less to presuppose.¹¹

What are we to make of this view? It must be admitted, I think, that this view is a good deal more plausible than the view that the Principle is a necessary truth, known *a priori*. For the proponent of this more modest view is not contending – or, at least, need not contend – that the Principle states a necessary truth about reality. All he contends is that the Principle is presupposed by us in our dealings with the world. To this he may add that without this presupposition we cannot make any sense of the world. However, there are several critical points pertinent to this view that need discussion.

First, does the scientist in his work really assume that everything that happens has a cause? In the debate between Russell and Copleston, Russell took the view that Physicists need not and do not assume that every event has a cause. "As for things not having a cause, the physicists assure us that individual quantum transitions in atoms have no cause."¹² Again, he remarks:

... a physicist looks for causes; that does not necessarily imply that there are causes everywhere. A man may look for gold without assuming that there is gold everywhere; if he finds gold, well and good, if he doesn't he's had bad luck. The same is true when the physicist looks for causes.¹³

How are we to settle this matter? Philosophers who hold that the causal principle is a fundamental assumption reply that the Heisenberg uncertainty principle "tells us something about the success (or the lack of it) of the present atomic theory in correlating observations, but not about nature in itself. . ."¹⁴ Moreover, it is observed that the failure to find causes does not lead anyone to abandon the causal principle. Indeed, it is sometimes argued that it is *impossible* to obtain empirical evidence against the principle.¹⁵ If we don't find gold in a hill after a careful search, we conclude that there's no gold there to be found. But if we don't find the cause of a certain event, we don't conclude that the event has no cause, only that it is extremely difficult to discover. Perhaps, then, there is some reason to think that we do assume that whatever happens has an explanation or cause.

But even if it is granted that in our dealings with the world we presuppose that whatever happens has a cause, there seems to be a serious difficulty confronting the recent defenders of the Cosmological Argument. For what the Cosmological Argument requires – or, more exactly, what the versions argued by Samuel Clarke, Copleston, and Taylor require – is what I have called the strong form of the Principle of Sufficient Reason. That is, their arguments require as a premise the principle that whatever exists – even an eternal being – has a cause or explanation of its existence. But what we have just granted to be presupposed by us in our dealing with the world is the principle that whatever *happens* has a cause. This latter principle implies that whatever begins to exist has a cause, since the coming into existence of a thing is an event, a happening. Thus the principle we have granted to be presupposed in science and commonsense implies what I have called the weak form of the Principle of Sufficient Reason. But it does not imply the strong form of the Principle; it does not imply that whatever exists has a cause. If something comes into existence, its coming into existence is something that happens. But if something exists from eternity, its eternal existence is not one of the things that happen. Hence, even if it be granted that we presuppose a cause for whatever happens, it does not follow that we presuppose a cause or explanation for whatever exists.

Can it reasonably be argued that the strong form of the Principle of Sufficient Reason is, as Taylor suggests, a presupposition that all men make, a presupposition of reason itself? We have granted as a presupposition of reason that there must be a cause or explanation for any thing that comes into existence.¹⁶ Thus if we imagine a star to have come into existence, say, a thousand years ago, it is presupposed that there must be an explanation for its having come into existence. That is, it is assumed by us that there must be a set of prior events that was sufficient to cause the birth of that star. To say, "Nothing caused the birth of the star, it just popped into existence and there is no reason why it came into existence" is, we have granted, to deny a fundamental presupposition of reason

itself. But imagine that there is a star in the heavens that never came into existence, a star that has always existed, that has existed from eternity. Do we presuppose that there must be an explanation for the eternal existence of this star? I am doubtful that we do. But short of a metaphysical investigation of mind and its relation to nature, it seems quite impossible to answer this question. Perhaps, then, our most fruitful course here is simply to note the consequences for the Cosmological Argument if the Principle of Sufficient Reason in its strong form is, as Copleston and Taylor maintain, a presupposition all men make.

However, before considering this last question it is, I think, important to clarify the nature of the question concerning a thing's existence to which the Principle of Sufficient Reason demands there be an answer. Of the star that came into existence a thousand years ago, we may ask "Where did it come from?", "What brought it into existence?", or "Why did it come into existence?" Clearly none of these questions can be asked properly of a star that has existed from eternity. Once we learn that it has always existed we realize that it never came into existence. But there is a simpler question that can be asked both about the eternally existing star and about the star that came into existence a thousand years ago; namely, "Why does this thing exist?" Although we may answer – or, at least, show to be improper – the question "Why did this thing come into existence?" by pointing out that it has always existed, the question "Why does this thing exist rather than not?" cannot be answered or even turned aside by pointing out that it has always existed. As Taylor has noted:

... it is no answer to the question, why a thing exists, to state *how long* it has existed. A geologist does not suppose that he has explained why there should be rivers and mountains merely by pointing out that they are old. Similarly, if one were to ask, concerning the ball of which we have spoken, for some sufficient reason for its being, he would not receive any answer upon being told that it had been there since yesterday. Nor would it be any better answer to say that it had existed since before anyone could remember, or even that it had always existed; for the question was not one concerning its age but its existence.¹⁷

The question, then, to which the Principle of Sufficient Reason requires that there be an answer is: "Why does this thing exist?" This question, I am claiming, may be sensibly asked about a star that has existed from eternity, or one that has existed for only a thousand years.

It should be clear that it is one thing to argue, as I have done, that the question "Why does this thing exist?" makes sense when asked of something that has always existed, and another thing to argue, as I have not done, that all men presuppose that there must be an adequate answer to that question, even when it is asked about something that has existed from eternity. We have granted as a presupposition of reason that there must be an adequate answer to the question when the being of which it is asked has come into existence. But, as I have indicated, it seems at least doubtful that the strong form of the Principle of Sufficient Reason is a presupposition of reason itself.

Suppose, as Taylor, Copleston, and others have claimed, that the Principle of Sufficient Reason in its strong form is a metaphysical assumption that all men

make, whether or not they reflect sufficiently to become aware of the assumption. What bearing would this have on the Cosmological Argument? It would not, of course, show that it is a good argument. For (1) the argument could be invalid, (2) some premise other than the premise expressing the Principle could be false, and (3) even the premise expressing the Principle could be false. The fact, if it is a fact, that all of us presuppose that whatever exists has an explanation of its existence does not imply that nothing exists without a reason for its existence. Nature is not bound to satisfy our presuppositions. As James has remarked in another connection, "In the great boarding-house of nature, the cakes and the butter and the syrup seldom come out so even and leave the plates so clean." However, if we do make such a presupposition we could not *consistently* reject the Cosmological Argument solely because it contains as a premise the Principle of Sufficient Reason. That is, if we reject the argument it must be for some reason other than its appeal to the Principle of Sufficient Reason.

If, as seems likely, the strong form of the Principle is not a presupposition of reason itself, and if, as I have argued, the Principle is neither analytically true nor a synthetic, necessary truth, known *a priori*, the Cosmological Argument – in so far as it requires the strong form of the Principle as a premise – cannot, I think, reasonably be maintained to be a *proof* of the existence of God. For unless there is a way of knowing the Principle to be true other than those we have explored, it follows that we do not know the Principle to be true. But if we do not know that one of the essential premises of an argument is true then we do not know that it is a good argument for its conclusion. It may, of course, be a perfectly good argument. But if to claim of an argument that it is a *proof* of its conclusion is to imply that its premises are *known* to be true, then we are not entitled to claim that the Cosmological Argument is a proof of the existence of God.

Notes

- 1 This argument is an adaptation of Samuel Clarke's discussion in his Boyle lectures of 1704, published under the title *A Demonstration of the Being and Attributes of God*. This work consists of twelve propositions, and arguments in support of these propositions. The first three propositions and their arguments constitute the first part of the Cosmological Argument. That is, the arguments for the first three propositions are designed to establish the existence of a necessary being. The substance of these arguments, I believe, is contained in the argument I have presented. There is also some resemblance between the argument I have presented and the argument Demea states in Part IX of the *Dialogues*. This is to be expected since Demea's argument is a brief restatement by Hume of the argument formulated by Clarke.
- 2 See, for example, Samuel Clarke's discussion of Propositions II and III in his *Demonstration*.
- 3 "The Existence of God: a Debate between Bertrand Russell and Father F.C. Copleston," John Hick (ed.), *The Existence of God* (New York: Macmillan, 1964), p. 175. The debate was originally broadcast by the British Broadcasting Corporation in 1948. References are to the debate as reprinted in *The Existence of God*.
- 4 Hume, *Dialogues*, Part IX.
- 5 *A Treatise of Human Nature*, book I, part III, section III.

- 6 Samuel Clarke, for example, makes the following remark in correspondence with a critic:

Nothing can be more absurd, than to suppose that anything (or any circumstance of any thing) is; and yet that there be absolutely no reason why it is, rather than not. Tis easy to conceive, that we may indeed be utterly ignorant of the reasons, or grounds, or causes of many things. But, that anything is; and that there is a real reason in nature why it is, rather than not; these two are as necessarily and essentially connected, as any two correlates whatever, as height and depth, etc.

The letter from which this passage comes is included in the 9th edition of the work from which our quotations from the *Demonstration* have been taken, p. 490.

- 7 See R. M. Chisholm, *Theory of Knowledge*, pp. 87–90.

- 8 *Treatise*, book 1, part III, section III.

- 9 John Laird, *Theism and Cosmology* (New York: Philosophical Library, 1942), p. 95. . . .

- 10 “A Debate,” p. 176.

- 11 Richard Taylor, *Metaphysics* (Englewood Cliffs, N.J.: Prentice-Hall, 1963), pp. 86–7.

- 12 “A Debate,” p. 176.

- 13 Ibid., p. 177.

- 14 Father Copleston, “A Debate,” p. 176.

- 15 G. J. Warnock has argued this in “Every Event Has a Cause,” *Logic and Language*, II, edited by Antony Flew (London: Blackwell, 1953). . . .

- 16 Clarke, perhaps for reasons of simplicity, usually speaks of requiring a cause only for the existence of a thing. But, of course, the Principle of Sufficient Reason is not meant to require an explanation only for the existence of a thing. Thus if a table is made by a carpenter and subsequently painted red, sawed in half, or even destroyed, Clarke’s view – and the view of others who have appealed to the Principle of Sufficient Reason – is that there *must be* an explanation not only for the fact that the table came into existence but also for any change that occurs to it. Thus Clarke remarks (in a passage quoted earlier), “Nothing can be more absurd, than to suppose that any thing (or any circumstance of any thing) is; and yet that there be absolutely *no reason why it is* rather than *not*.”

- 17 Taylor, *Metaphysics*, p. 88.

52 The Ontological Argument: Chapters II–IV of the *Proslogion** ---

St Anselm

Chapter II

Truly there is a God, although the fool hath said in his heart, There is no God.

And so, Lord, do thou, who dost give understanding to faith, give me, so far as thou knowest it to be profitable, to understand that thou art as we believe; and that thou art that which we believe. And, indeed, we believe that thou art a being than which nothing greater can be conceived. Or is there no such nature, since the fool hath said in his heart, there is no God? (Psalm xiv. 1). But, at any rate, this very fool, when he hears of this being of which I speak – a being than which nothing greater can be conceived – understands what he hears, and what he understands is in his understanding; although he does not understand it to exist.

For, it is one thing for an object to be in the understanding, and another to understand that the object exists. When a painter first conceives of what he will afterwards perform, he has it in his understanding, but he does not yet understand it to be, because he has not yet performed it. But after he has made the painting, he both has it in his understanding, and he understands that it exists, because he has made it.

Hence, even the fool is convinced that something exists in the understanding, at least, than which nothing greater can be conceived. For, when he hears of this, he understands it. And whatever is understood, exists in the understanding. And assuredly that, than which nothing greater can be conceived, cannot exist in the understanding alone. For, suppose it exists in the understanding alone: then it can be conceived to exist in reality; which is greater.

Therefore, if that, than which nothing greater can be conceived, exists in the understanding alone, the very being, than which nothing greater can be conceived, is one, than which a greater can be conceived. But obviously this is impossible. Hence, there is no doubt that there exists a being, than which nothing greater can be conceived, and it exists both in the understanding and in reality.

* From S. N. Deane, trans., *St Anselm: Basic Writings*, 2nd edition (LaSalle, Ill.: Open Court Publishing, 1968).

Chapter III

God cannot be conceived not to exist. – God is that, than which nothing greater can be conceived. – That which can be conceived not to exist is not God.

And it assuredly exists so truly, that it cannot be conceived not to exist. For, it is possible to conceive of a being which cannot be conceived not to exist; and this is greater than one which can be conceived not to exist. Hence, if that, than which nothing greater can be conceived, can be conceived not to exist, it is not that, than which nothing greater can be conceived. But this is an irreconcilable contradiction. There is, then, so truly a being than which nothing greater can be conceived to exist, that it cannot even be conceived not to exist; and this being thou art, O Lord, our God.

So truly, therefore, dost thou exist, O Lord, my God, that thou canst not be conceived not to exist; and rightly. For, if a mind could conceive of a being better than thee, the creature would rise above the Creator; and this is most absurd. And, indeed, whatever else there is, except thee alone, can be conceived not to exist. To thee alone, therefore, it belongs to exist more truly than all other beings, and hence in a higher degree than all others. For, whatever else exists does not exist so truly, and hence in a less degree it belongs to it to exist. Why, then, has the fool said in his heart, there is no God (Psalm xiv. 1), since it is so evident, to a rational mind, that thou dost exist in the highest degree of all? Why, except that he is dull and a fool?

Chapter IV

How the fool has said in his heart what cannot be conceived. – A thing may be conceived in two ways: (1) when the word signifying it is conceived; (2) when the thing itself is understood. As far as the word goes, God can be conceived not to exist; in reality he cannot.

But how has the fool said in his heart what he could not conceive; or how is it that he could not conceive what he said in his heart? since it is the same to say in the heart, and to conceive.

But, if really, nay, since really, he both conceived, because he said in his heart; and did not say in his heart, because he could not conceive; there is more than one way in which a thing is said in the heart or conceived. For, in one sense, an object is conceived, when the word signifying it is conceived; and in another, when the very entity, which the object is, is understood.

In the former sense, then, God can be conceived not to exist; but in the latter, not at all. For no one who understands what fire and water are can conceive fire to be water, in accordance with the nature of the facts themselves, although this is possible according to the words. So, then, no one who understands what God is can conceive that God does not exist, although he says these words in his heart, either without any, or with some foreign, signification. For,

God is that than which a greater cannot be conceived. And he who thoroughly understands this, assuredly understands that this being so truly exists, that not even in concept can it be non-existent. Therefore, he who understands that God so exists, cannot conceive that he does not exist.

I thank thee, gracious Lord, I thank thee; because what I formerly believed by thy bounty, I now so understand by thine illumination, that if I were unwilling to believe that thou dost exist, I should not be able not to understand this to be true.

53 Anselm's Ontological Arguments*

Norman Malcolm

I

I believe that in Anselm's *Proslogion* and *Responsio editoris* there are two different pieces of reasoning which he did not distinguish from one another, and that a good deal of light may be shed on the philosophical problem of "the ontological argument" if we do distinguish them. In Chapter 2 of the *Proslogion*¹ Anselm says that we believe that God is *something a greater than which cannot be conceived*. (The Latin is *aliquid quo nihil maius cogitari possit*. Anselm sometimes uses the alternative expressions *aliquid quo maius nihil cogitari potest*, *id quo maius cogitari nequit*, *aliquid quo maius cogitari non valet*.) Even the fool of the Psalm who says in his heart there is no God, when he hears this very thing that Anselm says, namely, "something a greater than which cannot be conceived," understands what he hears, and what he understands is in his understanding though he does not understand that it exists.

Apparently Anselm regards it as tautological to say that whatever is understood is in the understanding (*quidquid intelligitur in intellectu est*): he uses *intelligitur* and *in intellectu est* as interchangeable locutions. The same holds for another formula of his: whatever is thought is in thought (*quidquid cogitatur in cogitatione est*).²

Of course many things may exist in the understanding that do not exist in reality; for example, elves. Now, says Anselm, something a greater than which cannot be conceived exists in the understanding. But it cannot exist *only* in the understanding, for to exist in reality is greater. Therefore that thing a greater than which cannot be conceived cannot exist only in the understanding, for then a greater thing could be conceived: namely, one that exists both in the understanding and in reality.³

* From Norman Malcolm, "Anselm's Ontological Arguments," *Philosophical Review*, 69 (1960), pp. 41-62. Copyright © 1960 by Cornell University. Reprinted by permission of the publisher.

Here I have a question. It is not clear to me whether Anselm means that (a) existence in reality by itself is greater than existence in the understanding, or that (b) existence in reality and existence in the understanding together are greater than existence in the understanding alone. Certainly he accepts (b). But he might also accept (a), as Descartes apparently does in *Meditation III* when he suggests that the mode of being by which a thing is “objectively in the understanding” is *imperfect*.⁴ Of course Anselm might accept both (a) and (b). He might hold that in general something is greater if it has both of these “modes of existence” than if it has either one alone, but also that existence in reality is a more perfect mode of existence than existence in the understanding.

In any case, Anselm holds that something is greater if it exists both in the understanding and in reality than if it exists merely in the understanding. An equivalent way of putting this interesting proposition, in a more current terminology, is: something is greater if it is both conceived of and exists than if it is merely conceived of. Anselm’s reasoning can be expressed as follows: *id quo maius cogitari nequit* cannot be merely conceived of and not exist, for then it would not be *id quo maius cogitari nequit*. The doctrine that something is greater if it exists in addition to being conceived of, than if it is only conceived of, could be called the doctrine that *existence is a perfection*. Descartes maintained, in so many words, that existence is a perfection,⁵ and presumably he was holding Anselm’s doctrine, although he does not, in *Meditation V* or elsewhere, argue in the way that Anselm does in *Proslogion 2*.

When Anselm says “And certainly, that than which nothing greater can be conceived cannot exist merely in the understanding. For suppose it exists merely in the understanding, then it can be conceived to exist in reality, which is greater,”⁶ he is claiming that if I conceive of a being of great excellence, that being would be *greater* (more excellent, more perfect) if it existed than if it did not exist. His supposition that “it exists merely in the understanding” is the supposition that it is conceived of but does not exist. Anselm repeated this claim in his reply to the criticism of the monk Gaunilo. Speaking of the being a greater than which cannot be conceived, he says:

I have said that if it exists merely in the understanding it can be conceived to exist in reality, which is greater. Therefore, if it exists merely in the understanding obviously the very being a greater than which cannot be conceived, is one a greater than which can be conceived. What, I ask, can follow better than that? For if it exists merely in the understanding, can it not be conceived to exist in reality? And if it can be so conceived does not he who conceives of this conceive of a thing greater than it, if it does exist merely in the understanding? Can anything follow better than this: that if a being a greater than which cannot be conceived exists merely in the understanding, it is something a greater than which can be conceived? What could be plainer?⁷

He is implying, in the first sentence, that if I conceive of something which does not exist then it is possible for it to exist, and *it will be greater if it exists than if it does not exist*.

The doctrine that existence is a perfection is remarkably queer. It makes sense

and is true to say that my future house will be a better one if it is insulated than if it is not insulated; but what could it mean to say that it will be a better house if it exists than if it does not? My future child will be a better man if he is honest than if he is not; but who would understand the saying that he will be a better man if he exists than if he does not? Or who understands the saying that if God exists He is more perfect than if He does not exist? One might say, with some intelligibility, that it would be better (for oneself or for mankind) if God exists than if He does not – but that is a different matter.

A king might desire that his next chancellor should have knowledge, wit, and resolution; but it is ludicrous to add that the king's desire is to have a chancellor who exists. Suppose that two royal councillors, A and B, were asked to draw up separately descriptions of the most perfect chancellor they could conceive, and that the descriptions they produced were identical except that A included existence in his list of attributes of a perfect chancellor and B did not. (I do not mean that B put nonexistence in his list.) One and the same person could satisfy both descriptions. More to the point, any person who satisfied A's description would *necessarily* satisfy B's description and *vice versa!* This is to say that A and B did not produce descriptions that differed in any way but rather one and the same description of necessary and desirable qualities in a chancellor. A only made a show of putting down a desirable quality that B had failed to include.

I believe I am merely restating an observation that Kant made in attacking the notion that “existence” or “being” is a “real predicate.” He says:

By whatever and by however many predicates we may think a thing – even if we completely determine it – we do not make the least addition to the thing when we further declare that this thing *is*. Otherwise, it would not be exactly the same thing that exists, but something more than we had thought in the concept; and we could not, therefore, say that the exact object of my concept exists.⁸

Anselm's ontological proof of *Proslogion* 2 is fallacious because it rests on the false doctrine that existence is a perfection (and therefore that “existence” is a “real predicate”). It would be desirable to have a rigorous refutation of the doctrine but I have not been able to provide one. I am compelled to leave the matter at the more or less intuitive level of Kant's observation. In any case, I believe that the doctrine does not belong to Anselm's other formulation of the ontological argument. It is worth noting that Gassendi anticipated Kant's criticism when he said, against Descartes:

Existence is a perfection neither in God nor in anything else; it is rather that in the absence of which there is no perfection. . . . Hence neither is existence held to exist in a thing in the way that perfections do, nor if the thing lacks existence is it said to be imperfect (or deprived of a perfection), so much as to be nothing.⁹

II

I take up now the consideration of the second ontological proof, which Anselm presents in the very next chapter of the *Proslogion*. (There is no evidence that he thought of himself as offering two different proofs.) Speaking of the being a greater than which cannot be conceived, he says:

And it so truly exists that it cannot be conceived not to exist. For it is possible to conceive of a being which cannot be conceived not to exist; and this is greater than one which can be conceived not to exist. Hence, if that, than which nothing greater can be conceived, can be conceived not to exist, it is not that than which nothing greater can be conceived. But this is a contradiction. So truly, therefore, is there something than which nothing greater can be conceived, that it cannot even be conceived not to exist.

And this being thou art, O Lord, our God.¹⁰

Anselm is saying two things: first, that a being whose nonexistence is logically impossible is “greater” than a being whose nonexistence is logically possible (and therefore that a being a greater than which cannot be conceived must be one whose nonexistence is logically impossible); second, that *God* is a being than which a greater cannot be conceived.

In regard to the second of these assertions, there certainly is a use of the word “God,” and I think far the more common use, in accordance with which the statements “God is the greatest of all beings,” “God is the most perfect being,” “God is the supreme being,” are *logically* necessary truths, in the same sense that the statement “A square has four sides” is a logically necessary truth. If there is a man named “Jones” who is the tallest man in the world, the statement “Jones is the tallest man in the world” is merely true and is not a logically necessary truth. It is a virtue of Anselm’s unusual phrase, “a being a greater than which cannot be conceived,”¹¹ to make it explicit that the sentence “God is the greatest of all beings” expresses a logically necessary truth and not a mere matter of fact such as the one we imagined about Jones.

With regard to Anselm’s first assertion (namely, that a being whose nonexistence is logically impossible is greater than a being whose nonexistence is logically possible) perhaps the most puzzling thing about it is the use of the word “greater.” It appears to mean exactly the same as “superior,” “more excellent,” “more perfect.” This equivalence by itself is of no help to us, however, since the latter expressions would be equally puzzling here. What is required is some explanation of their use.

We do think of *knowledge*, say, as an excellence, a good thing. If A has more knowledge of algebra than B we express this in common language by saying that A has a *better* knowledge of algebra than B, or that A’s knowledge of algebra is *superior* to B’s, whereas we should not say that B has a better or superior *ignorance* of algebra than A. We do say “greater ignorance,” but here the word “greater” is used purely quantitatively.

Previously I rejected *existence* as a perfection. Anselm is maintaining in the

remarks last quoted, not that existence is a perfection, but that *the logical impossibility of nonexistence is a perfection*. In other words, *necessary existence is a perfection*. His first ontological proof uses the principle that a thing is greater if it exists than if it does not exist. His second proof employs the different principle that a thing is greater if it necessarily exists than if it does not necessarily exist.

Some remarks about the notion of *dependence* may help to make this latter principle intelligible. Many things depend for their existence on other things and events. My house was built by a carpenter: its coming into existence was dependent on a certain creative activity. Its continued existence is dependent on many things: that a tree does not crush it, that it is not consumed by fire, and so on. If we reflect on the common meaning of the word "God" (no matter how vague and confused this is), we realize that it is incompatible with this meaning that God's existence should *depend* on anything. Whether we believe in Him or not we must admit that the "almighty and everlasting God" (as several ancient prayers begin), the "Maker of heaven and earth, and of all things visible and invisible" (as is said in the Nicene Creed), cannot be thought of as being brought into existence by anything or as depending for His continued existence on anything. To conceive of anything as dependent upon something else for its existence is to conceive of it as a lesser being than God.

If a housewife has a set of extremely fragile dishes, then as dishes they are *inferior* to those of another set like them in all respects except that they are *not* fragile. Those of the first set are *dependent* for their continued existence on gentle handling; those of the second set are not. There is a definite connection in common language between the notions of dependency and inferiority, and independence and superiority. To say that something which was dependent on nothing whatever was superior to ("greater than") anything that was dependent in any way upon anything is quite in keeping with the everyday use of the terms "superior" and "greater." Correlative with the notions of dependence and independence are the notions of *limited* and *unlimited*. An engine requires fuel and this is a limitation. It is the same thing to say that an engine's operation is *dependent* on as that it is *limited* by its fuel supply. An engine that could accomplish the same work in the same time and was in other respects satisfactory, but did not require fuel, would be a *superior* engine.

God is usually conceived of as an *unlimited* being. He is conceived of as a being who *could not* be limited, that is, as an absolutely unlimited being. This is no less than to conceive of Him as *something a greater than which cannot be conceived*. If God is conceived to be an absolutely unlimited being He must be conceived to be unlimited in regard to His existence as well as His operation. In this conception it will not make sense to say that He depends on anything for coming into or continuing in existence. Nor, as Spinoza observed, will it make sense to say that something could *prevent* Him from existing.¹² Lack of moisture can prevent trees from existing in a certain region of the earth. But it would be contrary to the concept of God as an unlimited being to suppose that anything other than God Himself could prevent Him from existing, and it would be self-contradictory to suppose that He Himself could do it.

Some may be inclined to object that although nothing could prevent God's

existence, still it might just *happen* that He did not exist. And if He did exist that too would be by chance. I think, however, that from the supposition that it could happen that God did not exist it would follow that, if He existed, He would have mere duration and not eternity. It would make sense to ask, "How long has He existed?", "Will He still exist next week?", "He was in existence yesterday but how about today?", and so on. It seems absurd to make God the subject of such questions. According to our ordinary conception of Him, He is an eternal being. And eternity does not mean endless duration, as Spinoza noted. To ascribe eternity to something is to exclude as senseless all sentences that imply that it has duration. If a thing has duration then it would be merely a *contingent* fact, if it was a fact, that its duration was endless. The moon could have endless duration but not eternity. If something has endless duration it will *make sense* (although it will be false) to say that it will cease to exist, and it will make sense (although it will be false) to say that something will *cause* it to cease to exist. A being with endless duration is not, therefore, an absolutely unlimited being. That God is conceived to be eternal follows from the fact that He is conceived to be an absolutely unlimited being.

I have been trying to expand the argument of *Proslogion* 3. In *Responsio* 1 Anselm adds the following acute point: if you can conceive of a certain thing and this thing does not exist then if it *were* to exist its nonexistence would be *possible*. It follows, I believe, that if the thing were to exist it would depend on other things both for coming into and continuing in existence, and also that it would have duration and not eternity. Therefore it would not be, either in reality or in conception, an unlimited being, *aliquid quo nihil maius cogitari possit*.

Anselm states his argument as follows:

If it [the thing a greater than which cannot be conceived] can be conceived at all it must exist. For no one who denies or doubts the existence of a being a greater than which is inconceivable, denies or doubts that if it did exist its non-existence, either in reality or in the understanding, would be impossible. For otherwise it would not be a being a greater than which cannot be conceived. But as to whatever can be conceived but does not exist: if it were to exist its non-existence either in reality or in the understanding would be possible. Therefore, if a being a greater than which cannot be conceived, can even be conceived, it must exist.¹³

What Anselm has proved is that the notion of contingent existence or of contingent nonexistence cannot have any application to God. His existence must either be logically necessary or logically impossible. The only intelligible way of rejecting Anselm's claim that God's existence is necessary is to maintain that the concept of God, as a being a greater than which cannot be conceived, is self-contradictory or nonsensical.¹⁴ Supposing that this is false, Anselm is right to deduce God's necessary existence from his characterization of Him as a being a greater than which cannot be conceived.

Let me summarize the proof. If God, a being a greater than which cannot be conceived, does not exist then He cannot *come* into existence. For if He did He

would either have been *caused* to come into existence or have *happened* to come into existence, and in either case He would be a limited being, which by our conception of Him He is not. Since He cannot come into existence, if He does not exist His existence is impossible. If He does exist He cannot have come into existence (for the reasons given), nor can He cease to exist, for nothing could cause Him to cease to exist nor could it just happen that He ceased to exist. So if God exists His existence is necessary. Thus God's existence is either impossible or necessary. It can be the former only if the concept of such a being is self-contradictory or in some way logically absurd. Assuming that this is not so, it follows that He necessarily exists.¹⁵

It may be helpful to express ourselves in the following way: to say, not that *omnipotence* is a property of God, but rather that *necessary omnipotence* is; and to say, not that omniscience is a property of God, but rather that *necessary omniscience* is. We have criteria for determining that a man knows this and that and can do this and that, and for determining that one man has greater knowledge and abilities in a certain subject than another. We could think of various tests to give them. But there is nothing we should wish to describe, seriously and literally, as "testing" God's knowledge and powers. That God is omniscient and omnipotent has not been determined by the application of criteria: rather these are requirements of our conception of Him. They are internal properties of the concept, although they are also rightly said to be properties of God. *Necessary existence* is a property of God in the *same sense* that *necessary omnipotence* and *necessary omniscience* are His properties. And we are not to think that "God necessarily exists" means that it follows necessarily from something that God exists *contingently*. The a priori proposition "God necessarily exists" entails the proposition "God exists," if and only if the latter also is understood as an a priori proposition: in which case the two propositions are equivalent. In this sense Anselm's proof is a proof of God's existence.

Descartes was somewhat hazy on the question of whether existence is a property of things that exist, but at the same time he saw clearly enough that *necessary existence* is a property of God. Both points are illustrated in his reply to Gassendi's remark, which I quoted above:

I do not see to what class of reality you wish to assign existence, nor do I see why it may not be said to be a property as well as omnipotence, taking the word property as equivalent to any attribute or anything which can be predicated of a thing, as in the present case it should be by all means regarded. Nay, necessary existence in the case of God is also a true property in the strictest sense of the word, because it belongs to Him and forms part of His essence alone.¹⁶

Elsewhere he speaks of "the necessity of existence" as being "that crown of perfections without which we cannot comprehend God."¹⁷ He is emphatic on the point that necessary existence applies solely to "an absolutely perfect Being."¹⁸ . . .

IV

I turn to the question of whether the idea of a being a greater than which cannot be conceived is self-contradictory. Here Leibniz made a contribution to the discussion of the ontological argument. He remarked that the argument of Anselm and Descartes

is not a paralogism, but it is an imperfect demonstration, which assumes something that must still be proved in order to render it mathematically evident; that is, it is tacitly assumed that this idea of the all-great or all-perfect being is possible, and implies no contradiction. And it is already something that by this remark it is proved that, assuming that God is possible, he exists, which is the privilege of divinity alone.¹⁹

Leibniz undertook to give a proof that God is possible. He defined a *perfection* as a simple, positive quality in the highest degree.²⁰ He argued that since perfections are *simple* qualities they must be compatible with one another. Therefore the concept of a being possessing all perfections is consistent.

I will not review his argument because I do not find his definition of a perfection intelligible. For one thing, it assumes that certain qualities or attributes are “positive” in their intrinsic nature, and others “negative” or “privative,” and I have not been able to clearly understand that. For another thing, it assumes that some qualities are intrinsically simple. I believe that Wittgenstein has shown in the *Investigations* that nothing is *intrinsically* simple, but that whatever has the status of a simple, an indefinable, in one system of concepts, may have the status of a complex thing, a definable thing, in another system of concepts.

I do not know how to demonstrate that the concept of God – that is, of a being a greater than which cannot be conceived – is not self-contradictory. But I do not think that it is legitimate to demand such a demonstration. I also do not know how to demonstrate that either the concept of a material thing or the concept of *seeing* a material thing is not self-contradictory, and philosophers have argued that both of them are. With respect to any particular reasoning that is offered for holding that the concept of seeing a material thing, for example, is self-contradictory, one may try to show the invalidity of the reasoning and thus free the concept from the charge of being self-contradictory *on that ground*. But I do not understand what it would mean to demonstrate *in general*, and not in respect to any particular reasoning, that the concept is not self-contradictory. So it is with the concept of God. I should think there is no more of a presumption that it is self-contradictory than is the concept of seeing a material thing. Both concepts have a place in the thinking and the lives of human beings.

But even if one allows that Anselm’s phrase may be free of self-contradiction, one wants to know how it can have any *meaning* for anyone. Why is it that human beings have even *formed* the concept of an infinite being, a being a greater than which cannot be conceived? This is a legitimate and important question. I am sure there cannot be a deep understanding of that concept without an understanding of the phenomena of human life that give rise to it. To

give an account of the latter is beyond my ability. I wish, however, to make one suggestion (which should not be understood as autobiographical).

There is the phenomenon of feeling guilt for something that one has done or thought or felt or for a disposition that one has. One wants to be free of this guilt. But sometimes the guilt is felt to be so great that one is sure that nothing one could do oneself, nor any forgiveness by another human being, would remove it. One feels a guilt that is beyond all measure, a guilt "a greater than which cannot be conceived." Paradoxically, it would seem, one nevertheless has an intense desire to have this incomparable guilt removed. One requires a forgiveness that is beyond all measure, a forgiveness "a greater than which cannot be conceived." Out of such a storm in the soul, I am suggesting, there arises the conception of a forgiving mercy that is limitless, beyond all measure.²¹ This is one important feature of the Jewish and Christian conception of God.

I wish to relate this thought to a remark made by Kierkegaard, who was speaking about belief in Christianity but whose remark may have a wider application. He says:

There is only one proof of the truth of Christianity and that, quite rightly, is from the emotions, when the dread of sin and a heavy conscience torture a man into crossing the narrow line between despair bordering upon madness – and Christendom.²²

One may think it absurd for a human being to feel a guilt of such magnitude, and even more absurd that, if he feels it, he should *desire* its removal. I have nothing to say about that. It may also be absurd for people to fall in love, but they do it. I wish only to say that there *is* that human phenomenon of an unbearably heavy conscience and that it is importantly connected with the genesis of the concept of God, that is, with the formation of the "grammar" of the word "God." I am sure that this concept is related to human experience in other ways. If one had the acuteness and depth to perceive these connections one could grasp the *sense* of the concept. When we encounter this concept as a problem in philosophy, we do not consider the human phenomena that lie behind it. It is not surprising that many philosophers believe that the idea of a necessary being is an arbitrary and absurd construction.

What is the relation of Anselm's ontological argument to religious belief? This is a difficult question. I can imagine an atheist going through the argument, becoming convinced of its validity, acutely defending it against objections, yet remaining an atheist. The only effect it could have on the fool of the Psalm would be that he stopped saying in his heart "There is no God," because he would now realize that this is something he cannot meaningfully say or think. It is hardly to be expected that a demonstrative argument should, in addition, produce in him a living faith. Surely there is a level at which one can view the argument as a piece of logic, following the deductive moves but not being touched religiously? I think so. But even at this level the argument may not be without religious value, for it may help to remove some philosophical scruples that stand in the way of faith. At a deeper level, I suspect that the

argument can be thoroughly understood only by one who has a view of that human “form of life” that gives rise to the idea of an infinitely great being, who views it from the *inside* not just from the outside and who has, therefore, at least some inclination to *partake* in that religious form of life. This inclination, in Kierkegaard’s words, is “from the emotions.” This inclination can hardly be an *effect* of Anselm’s argument, but is rather presupposed in the fullest understanding of it. It would be unreasonable to require that the recognition of Anselm’s demonstration as valid must produce a conversion.²³

Notes

- 1 I have consulted the Latin text of the *Proslogion*, of *Gaunilonis Pro Insipiente*, and of the *Responsio editoris*, in S. Anselmi, *Opera Omnia*, ed. F. C. Schmitt (Secovii, 1938), vol. 1. With numerous modifications, I have used the English translation by S. N. Deane: *St Anselm* (La Salle, Ill.: Open Court, 1948).
- 2 See *Proslogion* 1 and *Responsio* 2.
- 3 Anselm’s actual words are: “Et certe id quo maius cogitari nequit, non potest esse in solo intellectu. Si enim vel in solo intellectu est, potest cogitari esse et in re, quod maius est. Si ergo id quo maius cogitari non potest, est in solo intellectu: id ipsum quo maius cogitari non potest, est quo maius cogitari potest. Sed certe hoc esse non potest.” *Proslogion* 2.
- 4 Haldane and Ross, *The Philosophical Works of Descartes*, vol. 1 (New York: Macmillan, 1931), p. 163.
- 5 *Op. cit.*, p. 182.
- 6 *Proslogion* 2, Deane, *St Anselm*, p. 8.
- 7 *Responsio* 2; Deane, *St Anselm*, pp. 157–8.
- 8 *The Critique of Pure Reason*, tr. by Norman Kemp Smith (New York: Macmillan, 1929), p. 505.
- 9 Haldane and Ross, *The Philosophical Works of Descartes*, vol. II, p. 186.
- 10 *Proslogion* 3; Deane, *St Anselm*, pp. 8–9.
- 11 Professor Robert Calhoun has pointed out to me that a similar locution had been used by Augustine. In *De moribus Manichaeorum* (bk II, ch. 11, sec. 24), he says that God is a being *quo esse out cogitari melius nihil possit* (*Patrologiae Patrum Latinorum*, J. P. Migne, ed. [Paris, 1841–5], vol. 32; *Augustinus*, vol. 1).
- 12 *Ethics*, Part I, prop. 11.
- 13 *Responsio* 1; Deane, *St Anselm*, pp. 154–5.
- 14 Gaunilo attacked Anselm’s argument on this very point. He would not concede that a being a greater than which cannot be conceived existed in his understanding (*Gaunilonis Pro Insipiente*, secs. 4 and 5; Deane, *St Anselm*, pp. 148–50). Anselm’s reply is: “I call on your faith and conscience to attest that this is most false” (*Responsio* 1; Deane, *St Anselm*, p. 154). Gaunilo’s faith and conscience will attest that it is false that “God is not a being a greater than which is inconceivable,” and false that “He is not understood (*intelligitur*) or conceived (*cogitatur*)” (*ibid*). Descartes remarks that one would go to “strange extremes” who denied that we understand the words “*that thing which is the most perfect that we can conceive*; for that is what all men call God” (Haldane and Ross, *The Philosophical Works of Descartes*, vol. II, p. 129).
- 15 [The following elegant argument occurs in *Responsio* 1: “That than which a greater cannot be conceived cannot be conceived to begin to exist. Whatever can be conceived to exist and does not exist, can be conceived to begin to exist. Therefore,

that than which a greater cannot be conceived, cannot be conceived to exist and yet not exist. So if it can be conceived to exist it exists from necessity." (*Nam quo maius cogitari nequit non potest cogitari esse nisi sine initio. Quidquid autem potest cogitari esse et non est, per initium potest cogitari esse. Non ergo quo maius cogitari nequit cogitari potest esse et non est. Si ergo cogitari potest esse, ex necessitate est.*) (Schmitt, *Opera Omnia*, p. 131; Deane, *St Anselm*, p. 154.)]

- 16 Haldane and Ross, *The Philosophical Works of Descartes*, vol. II, p. 228.
- 17 Ibid., vol. I, p. 445.
- 18 E.g., ibid., Principle 15, p. 225.
- 19 *New Essays Concerning the Human Understanding*, Book IV, ch. 10; A. G. Langley, ed. (La Salle, Ill.: Open Court Publishing, 1949), p. 504.
- 20 See Ibid., Appendix X, p. 714.
- 21 [Psalm 116: "The sorrows of death compassed me, and the pains of hell gat hold upon me: I found trouble and sorrow. Then called I upon the name of the Lord; O Lord, I beseech thee, deliver my soul." Psalm 130: "Out of the depths have I cried unto thee, O Lord."]
- 22 *The Journals*, tr. by A. Dru (New York: Oxford University Press, 1938), sec. 926.
- 23 [Since the appearance of this essay many acute criticisms of it have been published or communicated to me in private correspondence. In *The Philosophical Review*, LXX, No. 1 (January 1961), there are the following articles: Raziel Abelson, "Not Necessarily"; R. E. Allen, "The Ontological Argument"; Paul Henle, "Uses of the Ontological Argument"; Gareth B. Matthews, "On Conceivability in Anselm and Malcolm"; Alvin Plantinga, "A Valid Ontological Argument?"; Terence Penelhum, "On the Second Ontological Argument." Some other published articles are: Jan Berg, "An Examination of the Ontological Proof," *Theoria*, xxvii, No. 3 (1961); T. P. Brown, "Professor Malcolm on 'Anselm's Ontological Arguments,'" *Analysis*, October 1961; W. J. Huggett, "The Nonexistence of Ontological Arguments," *The Philosophical Review*, LXXI, No. 3 (July 1962); Jerome Shaffer, "Existence, Prediction, and the Ontological Argument," *Mind*, LXXI, No. 283 (July 1962). It would be a major undertaking to attempt to reply to all of the criticisms, and I hope that my not doing so will not be construed as a failure to appreciate them. I do not know that it is possible to meet all of the objections; on the other hand, I do not know that it is impossible.]

PART FIVE

IS METAPHYSICS POSSIBLE?

Introduction

- 54 The Rejection of Metaphysics: Chapter 1 of *Philosophy and Logical Syntax*
RUDOLF CARNAP
- 55 Postmodernism, Feminism, and Metaphysics: an Excerpt from *Thinking Fragments*
JANE FLAX
- 56 Metaphysics and Feminist Theory: Excerpts from “Feminist Metaphysics” and “Anti-Essentialism in Feminist Theory”
CHARLOTTE WITT

Introduction

Many celebrated philosophers have been opposed to metaphysics as a matter of principle; Immanuel Kant, A. J. Ayer, Rudolf Carnap, Ludwig Wittgenstein, Hilary Putnam, Jacques Derrida, and many others have called for or attempted to bring about the “death of metaphysics.” So far, however, Etienne Gilson has been right: Metaphysics always buries its undertakers.¹

This part includes representatives of two very different philosophical movements, each heralding the end of metaphysics: logical positivism and postmodern critical theory of the “deconstructionist” sort. Carnap represents the first; and the incoherence at the heart of Carnap’s positivist critique of metaphysics is briefly described above, in the “Introduction: What is Metaphysics?”, section 2. Recent challenges to metaphysics from postmodernism and feminism are clearly and succinctly summarized and endorsed (with qualifications) by Jane Flax. Near the end of the excerpt from Flax’s book, she suggests that feminists should ally themselves with the postmodernist critics of metaphysics. Charlotte Witt, on the other hand, argues that there is no good reason for feminists to do so; that there is nothing inherently “masculine and oppressive” about the metaphysical enterprise.

It is useful to compare Flax’s description of anti-metaphysical currents in postmodern thought with the excerpts from Putnam and Quine in Part III, above. The inaccessibility of an “Archimedean point” – a neutral standpoint from which absolute truth can be ascertained – is a theme common to the anti-realism of Quine, Putnam, and Flax’s postmodern anti-metaphysicians.

Suggestions for Further Reading

- Aune, Bruce, *Metaphysics: The Elements* (Minneapolis: University of Minnesota, 1985), chs 6 and 7: “Worlds, Objects, and Structure” and “Meaning, Truth, and Metaphysics.”
- Benardete, José, *Metaphysics: The Logical Approach* (Oxford: Oxford University Press, 1989), ch. 21: “No Entity without Identity.”
- Carter, William R., *The Elements of Metaphysics* (Philadelphia, Penn.: Temple University Press, 1990), ch. 12: “Being Realistic.”
- Gardner, Martin, *The Whys of a Philosophical Scrivener* (New York: Quill, 1983), chs 1 and 2: “The World: Why I am Not a Solipsist” and “Truth: Why I am Not a Pragmatist.”
- Hamlyn, D. W., *Metaphysics* (Cambridge, UK: Cambridge University Press, 1984), chs 1 and 2: “Introduction” and “Appearance and Reality.”
- Jubien, Michael, *Contemporary Metaphysics* (Oxford: Blackwell, 1997), ch. 5: “Is Truth Relative?”
- Post, John F., *Metaphysics: A Contemporary Introduction* (New York: Paragon House, 1991), chs 2 and 3: “Language and Reality” and “Piercing the Veil of Language.”
- Richard Rorty, “The World Well Lost,” *Journal of Philosophy*, 69 (1972), pp. 649–65.
- van Inwagen, Peter, *Metaphysics* (Boulder, Col.: Westview Press, 1993), ch. 4: “Objectivity.”

METAPHYSICS: THE BIG QUESTIONS

Note

- 1 Cf. E. Gilson, *The Unity of Philosophical Experience* (New York: Scribner's Sons, 1937), pp. 306–7.

54 The Rejection of Metaphysics: Chapter 1 of *Philosophy and Logical Syntax**

Rudolf Carnap

1 Verifiability

The problems of philosophy as usually dealt with are of very different kinds. From the point of view which I am here taking we may distinguish mainly three kinds of problems and doctrines in traditional philosophy. For the sake of simplicity we shall call these parts *Metaphysics*, *Psychology*, and *Logic*. Or, rather, there are not three distinct regions, but three sorts of components which in most theses and questions are combined: a metaphysical, a psychological, and a logical component.

The considerations that follow belong to the third region: we are here carrying out *Logical Analysis*. The function of logical analysis is to analyse all knowledge, all assertions of science and of everyday life, in order to make clear the sense of each such assertion and the connections between them. One of the principal tasks of the logical analysis of a given proposition is to find out the method of verification for that proposition. The question is: What reasons can there be to assert this proposition; or: How can we become certain as to its truth or falsehood? This question is called by the philosophers the epistemological question; epistemology or the philosophical theory of knowledge is nothing other than a special part of logical analysis, usually combined with some psychological questions concerning the process of knowing.

What, then, is the method of verification of a proposition? Here we have to distinguish between two kinds of verification: direct and indirect. If the question is about a proposition which asserts something about a present perception, e.g. ‘Now I see a red square on a blue ground,’ then the proposition can be tested directly by my present perception. If at present I do see a red square on a blue ground, the proposition is directly verified by this seeing; if I do not see that, it is disproved. To be sure, there are still some serious problems in connection with direct verification. We will however not touch on them here, but give our attention to the question of *indirect* verification, which is more important for our purposes. A proposition P which is not directly verifiable can only be verified by direct verification of propositions deduced from P together with other already verified propositions.

Let us take the proposition P_1 : ‘This key is made of iron.’ There are many ways of verifying this proposition; e.g.: I place the key near a magnet; then I

* From Rudolf Carnap, *Philosophy and Logical Syntax* (London: Kegan Paul, Trench, Trubner, 1935). Reprinted by permission of Routledge.

perceive that the key is attracted. Here the deduction is made in this way:

- Premises.*
- P₁: ‘This key is made of iron’; the proposition to be examined.
 - P₂: ‘If an iron thing is placed near a magnet, it is attracted’; this is a physical law, already verified.
 - P₃: ‘This object – a bar – is a magnet’; proposition already verified.
 - P₄: ‘The key is placed near the bar’; this is now directly verified by our observation.

From these four premises we can deduce the conclusion:

- P₅: ‘The key will now be attracted by the bar.’

This proposition is a prediction which can be examined by observation. If we look, we either observe the attraction or we do not. In the first case we have found a positive instance, an instance of verification of the proposition P₁ under consideration; in the second case we have a negative instance, an instance of disproof of P₁.

In the first case the examination of the proposition P₁ is not finished. We may repeat the examination by means of a magnet, i.e. we may deduce other propositions similar to P₅ by the help of the same or similar premises as before. After that, or instead of that, we may make an examination by electrical tests, or by mechanical, chemical, or optical tests, etc. If in these further investigations all instances turn out to be positive, the certainty of the proposition P₁ gradually grows. We may soon come to a degree of certainty sufficient for all practical purposes, but *absolute* certainty we can never attain. The number of instances deducible from P₁ by the help of other propositions already verified or directly verifiable is *infinite*. Therefore there is always a possibility of finding in the future a negative instance, however small its probability may be. Thus the proposition P₁ *can never be completely verified*. For this reason it is called an *hypothesis*.

So far we have considered an individual proposition concerning one single thing. If we take a general proposition concerning all things or events at whatever time and place, a so-called natural *law*, it is still clearer that the number of examinable instances is infinite and so the proposition is an hypothesis.

Every assertion P in the wide field of science has this character, that it asserts either something about present perceptions or other experiences, and therefore is verifiable by them, or that propositions about future perceptions are deducible from P together with some other already verified propositions. If a scientist should venture to make an assertion from which no perceptive propositions could be deduced, what should we say to that? Suppose, e.g., he asserts that there is not only a gravitational field having an effect on bodies according to the known laws of gravitation, but also a *levitational field*, and on being asked what sort of effect this levitational field has, according to his theory, he answers that there is no observable effect; in other words, he confesses his inability to give rules according to which we could deduce perceptive propositions from his assertion. In that case

our reply is: your assertion is no assertion at all; it does not speak about anything; it is nothing but a series of empty words; it is simply without sense.

It is true that he may have images and even feelings connected with his words. This fact may be of psychological importance; logically, it is irrelevant. What gives theoretical meaning to a proposition is not the attendant images and thoughts, but the possibility of deducing from it perceptive propositions, in other words, the possibility of verification. To give sense to a proposition the presence of images is not sufficient; it is not even necessary. We have no actual image of the electro-magnetic field, nor even, I should say, of the gravitational field. Nevertheless the propositions which physicists assert about these fields have a perfect sense, because perceptive propositions are deducible from them. I by no means object to the proposition just mentioned about a levitational field that we do not know how to imagine or conceive such a field. My only objection to that proposition is that we are not told how to verify it.

2 Metaphysics

What we have been doing so far is *logical analysis*. Now we are going to apply these considerations not to propositions of physics as before, but to propositions of *metaphysics*. Thus our investigation belongs to *logic*, to the third of the three parts of philosophy spoken about before, but the *objects* of this investigation belong to the first part.

I will call *metaphysical* all those propositions which claim to represent knowledge about something which is over or beyond all experience, e.g. about the real Essence of things, about Things in themselves, the Absolute, and such like. I do not include in metaphysics those theories – sometimes called metaphysical – whose object is to arrange the most general propositions of the various regions of scientific knowledge in a well-ordered system; such theories belong actually to the field of empirical science, not of philosophy, however daring they may be. The sort of propositions I wish to denote as metaphysical may most easily be made clear by some examples: ‘The Essence and Principle of the world is Water’, said Thales; ‘Fire’, said Heraclitus; ‘the Infinite’, said Anaximander; ‘Number’, said Pythagoras. ‘All things are nothing but shadows of eternal ideas which themselves are in a spaceless and timeless sphere’, is a doctrine of Plato. From the Monists we learn: ‘There is only one principle on which all that is, is founded’; but the Dualists tell us: ‘There are two principles.’ The Materialists say: ‘All that is, is in its essence material’, but the Spiritualists say: ‘All that is, is spiritual.’ To metaphysics (in our sense of the word) belong the principal doctrines of Spinoza, Schelling, Hegel, and – to give at least one name of the present time – Bergson.

Now let us examine this kind of proposition from the point of view of *verifiability*. It is easy to realize that such propositions are not verifiable. From the proposition: ‘The Principle of the world is Water’ we are not able to deduce any proposition asserting any perceptions or feelings or experiences whatever which may be expected for the future. Therefore the proposition, ‘The Principle of

the world is Water', asserts nothing at all. It is perfectly analogous to the proposition in the fictive example above about the levitational field and therefore it has no more sense than that proposition. The Water-Metaphysician – as we may call him – has no doubt many images connected with his doctrine; but they cannot give sense to the proposition, any more than they could in the case of the levitational field. Metaphysicians cannot avoid making their propositions non-verifiable, because if they made them verifiable, the decision about the truth or falsehood of their doctrines would depend upon experience and therefore belong to the region of empirical science. This consequence they wish to avoid, because they pretend to teach knowledge which is of a higher level than that of empirical science. Thus they are compelled to cut all connection between their propositions and experience; and precisely by this procedure they deprive them of any sense.

3 Problems of Reality

So far I have considered only examples of such propositions as are usually called metaphysical. The judgment I have passed on these propositions, namely, that they have no empirical sense, may perhaps appear not very astonishing, and even trivial. But it is to be feared that the reader will experience somewhat more difficulty in agreement when I now proceed to apply that judgment also to philosophical doctrines of the type which is usually called epistemological. I prefer to call them also metaphysical because of their similarity, in the point under consideration, to the propositions usually so called. What I have in mind are the doctrines of Realism, Idealism, Solipsism, Positivism and the like, taken in their traditional form as asserting or denying the Reality of something. The Realist asserts the Reality of the external world, the Idealist denies it. The Realist – usually at least – asserts also the Reality of other minds, the Solipsist – an especially radical Idealist – denies it, and asserts that only his own mind or consciousness is real. Have these assertions sense?

Perhaps it may be said that assertions about the reality or unreality of something occur also in empirical science, where they are examined in an empirical way, and that therefore they have sense. This is quite true. But we have to distinguish between two concepts of reality, one occurring in empirical propositions and the other occurring in the philosophical propositions just mentioned. When a zoologist asserts the reality of kangaroos, his assertion means that there are things of a certain sort which can be found and perceived at certain times and places; in other words that there are objects of a certain sort which are elements of the space-time-system of the physical world. This assertion is of course verifiable; by empirical investigation every zoologist arrives at a positive verification, independent of whether he is a Realist or an Idealist. Between the Realist and the Idealist there is full agreement as to the question of the reality of things of such and such sort, i.e. of the possibility of locating elements of such and such sort in the system of the physical world. The disagreement begins only when the question about the Reality of the physical world as a whole is raised.

But this question has no sense, because the reality of anything is nothing else than the possibility of its being placed in a certain system, in this case, in the space-time-system of the physical world, and such a question has sense only if it concerns elements or parts, not if it concerns the system itself.

The same result is obtained by applying the criterion explained before: the possibility of deducing perceptive propositions. While from the assertion of the reality or the existence of kangaroos we *can* deduce perceptive propositions, from the assertion of the Reality of the physical world this is not possible; neither is it possible from the opposite assertion of the Unreality of the physical world. Therefore both assertions have no empirical content – no sense at all. It is to be emphasized that this criticism of having no sense applies equally to the assertion of Unreality. Sometimes the views of the *Vienna Circle* have been mistaken for a denial of the Reality of the physical world, but we make no such denial. It is true that we reject the thesis of the Reality of the physical world; but we do not reject it as false, but as having no sense, and its Idealistic *anti*-thesis is subject to exactly the same rejection. We neither assert nor deny these theses, we reject the whole question.

All the considerations which apply to the question of the Reality of the physical world apply also to the other philosophical questions of Reality, e.g. the Reality of other minds, the Reality of the given, the Reality of universals, the Reality of qualities, the Reality of relations, the Reality of numbers, etc. If any philosophical thesis answering any of these questions positively or negatively is added to the system of scientific hypotheses, this system will not in the least become more effective; we shall not be able to make any further prediction as to future experiences. Thus all these philosophical theses are deprived of empirical content, of theoretical sense; they are pseudo-theses.

If I am right in this assertion, the philosophical problems of Reality – as distinguished from the empirical problems of reality – have the same logical character as the problems (or rather, pseudo-problems) of transcendental metaphysics earlier referred to. For this reason I call those problems of Reality not epistemological problems – as they usually are called – but metaphysical.

Among the metaphysical doctrines that have no theoretical sense I have also mentioned *Positivism*, although the *Vienna Circle* is sometimes designated as Positivistic. It is doubtful whether this designation is quite suitable for us. In any case we do not assert the thesis that only the Given is Real, which is one of the principal theses of traditional Positivism. The name Logical Positivism seems more suitable, but this also can be misunderstood. At any rate it is important to realize that our doctrine is a logical one and has nothing to do with metaphysical theses of the Reality or Unreality of anything whatever. What the character of a *logical* thesis is, will be made clear in the following chapters.

4 Ethics

One division of philosophy, which by some philosophers is considered the most important, has not been mentioned at all so far, namely, the philosophy of

values, with its main branch, moral philosophy or *Ethics*. The word 'Ethics' is used in two different senses. Sometimes a certain empirical investigation is called 'Ethics', *viz.* psychological and sociological investigations about the actions of human beings, especially regarding the origin of these actions from feelings and volitions and their effects upon other people. Ethics in this sense is an empirical, scientific investigation; it belongs to empirical science rather than to philosophy. Fundamentally different from this is ethics in the second sense, as the philosophy of moral values or moral norms, which one can designate normative ethics. This is not an investigation of facts, but a pretended investigation of what is good and what is evil, what it is right to do and what it is wrong to do. Thus the purpose of this philosophical, or normative, ethics is to state norms for human action or judgments about moral values.

It is easy to see that it is merely a difference of formulation, whether we state a norm or a value judgment. A norm or rule has an imperative form, for instance: 'Do not kill!' The corresponding value judgment would be: 'Killing is evil.' This difference of formulation has become practically very important, especially for the development of philosophical thinking. The rule, 'Do not kill,' has grammatically the imperative form and will therefore not be regarded as an assertion. But the value statement, 'Killing is evil,' although, like the rule, it is merely an expression of a certain wish, has the grammatical form of an assertive proposition. Most philosophers have been deceived by this form into thinking that a value statement is really an assertive proposition, and must be either true or false. Therefore they give reasons for their own value statements and try to disprove those of their opponents. But actually a value statement is nothing else than a command in a misleading grammatical form. It may have effects upon the actions of men, and these effects may either be in accordance with our wishes or not; but it is neither true nor false. It does not assert anything and can neither be proved nor disproved.

This is revealed as soon as we apply to such statements our method of logical analysis. From the statement 'Killing is evil' we cannot deduce any proposition about future experiences. Thus this statement is not verifiable and has no theoretical sense, and the same thing is true of all other value statements.

Perhaps somebody will contend in opposition that the following proposition is deducible: 'If a person kills anybody he will have feelings of remorse.' But this proposition is in no way deducible from the proposition 'Killing is evil.' It is deducible only from psychological propositions about the character and the emotional reactions of the person. These propositions are indeed verifiable and not without sense. They belong to psychology, not to philosophy; to psychological ethics (if one wishes to use this word), not to philosophical or normative ethics. The propositions of normative ethics, whether they have the form of rules or the form of value statements, have no theoretical sense, are not scientific propositions (taking the word scientific to mean any assertive proposition).

To avoid misunderstanding it must be said that we do not at all deny the possibility and importance of a scientific investigation of value statements as well as of acts of valuation. Both of these are acts of individuals and are, like all other kinds of acts, possible objects of empirical investigation. Historians, psy-

chologists, and sociologists may give analyses and causal explanations of them, and such historical and psychological propositions about acts of valuation and about value statements are indeed meaningful scientific propositions which belong to ethics in the first sense of this word. But the value statements themselves are here only objects of investigation; they are not propositions in these theories, and have, here as elsewhere, no theoretical sense. Therefore we assign them to the realm of metaphysics.

5 Metaphysics as Expression

Now we have analysed the propositions of metaphysics in a wide sense of this word, including not only transcendental metaphysics, but also the problems of philosophical Reality and lastly normative ethics. Perhaps many will agree that the propositions of all these kinds of metaphysics are not verifiable, i.e. that their truth cannot be examined by experience. And perhaps many will even grant that for this reason they have not the character of scientific propositions. But when I say that they are without sense, assent will probably seem more difficult. Someone may object: these propositions in the metaphysical books obviously have an effect upon the reader, and sometimes a very strong effect; therefore they certainly *express* something. That is quite true, they *do* express something, but nevertheless they have no sense, no theoretical content.

We have here to distinguish two functions of language, which we may call the expressive function and the representative function. Almost all the conscious and unconscious movements of a person, including his linguistic utterances, express something of his feelings, his present mood, his temporary or permanent dispositions to reaction, and the like. Therefore we may take almost all his movements and words as symptoms from which we can infer something about his feelings or his character. That is the expressive function of movements and words. But besides that, a certain portion of linguistic utterances (e.g. 'this book is black'), as distinguished from other linguistic utterances and movements, has a second function: these utterances represent a certain state of affairs; they tell us that something is so and so; they assert something, they predicate something, they judge something.

In special cases, this asserted state may be the same as that which is inferred from a certain expressive utterance; but even in such cases we must sharply distinguish between the assertion and the expression. If, for instance, somebody is laughing, we may take this as a symptom of his merry mood; if on the other hand he tells us without laughing: 'Now I am merry,' we can learn from his words the same thing which we inferred in the first case from his laughing. Nevertheless, there is a fundamental mental difference between the laughter and the words: 'I am merry now.' This linguistic utterance *asserts* the merry mood, and therefore it is either true or false. The laughter does not assert the merry mood but *expresses* it. It is neither true nor false, because it does not assert anything, although it may be either genuine or deceptive.

Now many linguistic utterances are analogous to laughing in that they have

only an expressive function, no representative function. Examples of this are cries like 'Oh, Oh' or, on a higher level, lyrical verses. The aim of a lyrical poem in which occur the words 'sunshine' and 'clouds', is not to inform us of certain meteorological facts, but to express certain feelings of the poet and to excite similar feelings in us. A lyrical poem has no assertional sense, no theoretical sense, it does not contain knowledge.

The meaning of our anti-metaphysical thesis may now be more clearly explained. This thesis asserts that metaphysical propositions – like lyrical verses – have only an expressive function, but no representative function. Metaphysical propositions are neither true nor false, because they assert nothing, they contain neither knowledge nor error, they lie completely outside the field of knowledge, of theory, outside the discussion of truth or falsehood. But they are, like laughing, lyrics, and music, expressive. They express not so much temporary feelings as permanent emotional or volitional dispositions. Thus, for instance, a Metaphysical system of Monism may be an expression of an even and harmonious mode of life, a Dualistic system may be an expression of the emotional state of someone who takes life as an eternal struggle; an ethical system of Rigorism may be expressive of a strong sense of duty or perhaps of a desire to rule severely. Realism is often a symptom of the type of constitution called by psychologists extraverted, which is characterized by easily forming connections with men and things; Idealism, of an opposite constitution, the so-called introverted type, which has a tendency to withdraw from the unfriendly world and to live within its own thoughts and fancies.

Thus we find a great similarity between metaphysics and lyrics. But there is one decisive difference between them. Both have no representative function, no theoretical content. A metaphysical proposition, however – as distinguished from a lyrical verse – *seems* to have some, and by this not only is the reader deceived, but the metaphysician himself. He believes that in his metaphysical treatise he has asserted something, and is led by this into argument and polemics against the propositions of some other metaphysician. A poet, however, does not assert that the verses of another are wrong or erroneous; he usually contents himself with calling them bad.

The non-theoretical character of metaphysics would not be in itself a defect; all arts have this non-theoretical character without thereby losing their high value for personal as well as for social life. The danger lies in the *deceptive* character of metaphysics; it gives the illusion of knowledge without actually giving any knowledge. This is the reason why we reject it.

6 Psychology

When we have eliminated metaphysical problems and doctrines from the region of knowledge or theory, there remain still two kinds of philosophical questions: psychological and logical. Now we shall eliminate the psychological questions also, not from the region of knowledge, but from philosophy. Then, finally, philosophy will be reduced to logic alone (in a wide sense of this word). [See Figure 10.]

EXPRESSIVE FUNCTION
OF LANGUAGEREPRESENTATIVE FUNCTION
OF LANGUAGE

Arts

Science (= the System of Theoretical Knowledge)

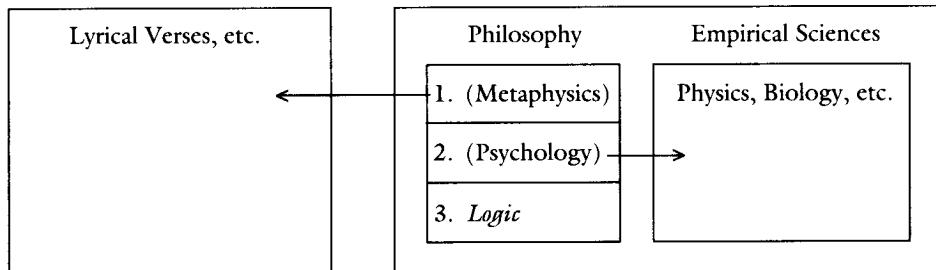


Figure 10

Psychological questions and propositions are certainly not without sense. From such propositions we can deduce other propositions about future experiences and by their help we can verify the psychological propositions. But the propositions of psychology belong to the region of empirical science in just the same way as do the propositions of chemistry, biology, history and the like. The character of psychology is by no means more philosophical than that of the other sciences mentioned. When we look at the historical development of the sciences we see that philosophy has been the mother of them all. One science after another has been detached from philosophy and has become an independent science. Only in our time has the umbilical cord between psychology and philosophy been cut. Many philosophers have not yet realized quite clearly that psychology is no longer an embryo, but an independent organism, and that psychological questions have to be left to empirical research.

Of course we have no objection to connecting psychological and logical investigations, any more than to connecting investigations of any scientific kind. We reject only the confusion of the two kinds of questions. We demand that they should be clearly distinguished even where in practice they are combined. The confusion sometimes consists in dealing with a logical question as if it were a psychological one. This mistake – called Psychologism – leads to the opinion that logic is a science concerning thinking, that is, either concerning the actual operation of thinking or the rules according to which thinking should proceed. But as a matter of fact the investigation of operations of thinking as they really occur is a task for psychology and has nothing to do with logic. And learning how to think *aright* is what we do in every other science as well as in logic. In astronomy we learn how to think aright about stars; in logic we learn how to think aright about the special objects of logic. What these special objects of logic are, will be seen in the next chapter. In any case thinking is not an object of logic, but of psychology.

Psychological questions concern all kinds of so-called psychic or mental events, all kinds of sensations, feelings, thoughts, images, etc., whether they are conscious or unconscious. These questions of psychology can be answered only by experience, not by philosophizing.

7 Logical Analysis

The only proper task of *Philosophy* is *Logical Analysis*. And now the principal question to be answered here will be: ‘*What is logical analysis?*’ In our considerations so far we have already practised logical analysis: we have tried to determine the character of physical hypotheses, of metaphysical propositions (or rather, pseudo-propositions), of psychological propositions. And now we have to apply logical analysis to logical analysis itself; we have to determine the character of the propositions of logic, of those propositions which are the results of logical analysis.

The opinion that metaphysical propositions have no sense because they do not concern any facts, has already been expressed by *Hume*. He writes in the last chapter of his ‘Enquiry Concerning Human Understanding’ (published in the year 1748) as follows: ‘It seems to me, that the only objects of the abstract sciences or of demonstration, are quantity and number. . . . All other enquiries of men regard only matter of fact and existence; and these are evidently incapable of demonstration. . . . When we run over libraries, persuaded of these principles, what havoc must we make? If we take in our hand any volume, of divinity or school metaphysics, for instance; let us ask, Does it contain any abstract reasoning concerning quantity or number? No. Does it contain any experimental reasoning concerning matter of fact and existence? No. Commit it then to the flames: for it can contain nothing but sophistry and illusion.’ We agree with this view of Hume, which says – translated into our terminology – that only the propositions of mathematics and empirical science have sense, and that all other propositions are without sense.

But now it may perhaps be objected: ‘How about your own propositions? In consequence of your view your own writings, including this book, would be without sense, for they are neither mathematical nor empirical, that is, verifiable by experience.’ What answer can be given to this objection? What is the character of my propositions and in general of the propositions of logical analysis? This question is decisive for the consistency of the view which has been explained here.

An answer to the objection is given by Wittgenstein in his book *Tractatus Logico-Philosophicus*.¹ This author has developed most radically the view that the propositions of metaphysics are shown by logical analysis to be without sense. How does he reply to the criticism that in that case his own propositions are also without sense? He replies by agreeing with it. He writes: ‘The result of philosophy is not a number of “philosophical propositions,” but to make propositions clear’ (p. 77). ‘My propositions are elucidatory in this way: he who understands me finally recognizes them as senseless, when he has climbed out through them, on them, over them. (He must so to speak throw away the ladder, after he has climbed up on it.) He must surmount these propositions; then he sees the world rightly. Whereof one cannot speak, thereof one must be silent’ (p. 189).

I, as well as my friends in the *Vienna Circle*, owe much to Wittgenstein,

especially as to the analysis of metaphysics. But on the point just mentioned I cannot agree with him. In the first place he seems to me to be inconsistent in what he does. He tells us that one cannot state philosophical propositions and that whereof one cannot speak, thereof one must be silent; and then instead of keeping silent, he writes a whole philosophical book. Secondly, I do not agree with his statement that all his propositions are quite as much without sense as metaphysical propositions are. My opinion is that a great number of his propositions (unfortunately not all of them) have in fact sense; and that the same is true for all propositions of logical analysis.

It will be the purpose of the following chapters to give reasons for this positive answer to the question about the character of philosophical propositions, to show a way of formulating the results of logical analysis, a way not exposed to the objection mentioned, and thus to exhibit an *exact method of philosophy*.

Notes

- 1 Ludwig Wittgenstein, *Tractatus Logico-Philosophicus*, trs. C. K. Ogden (London: Routledge & Kegan Paul, 1927).

55 Postmodernism, Feminism, and Metaphysics: an Excerpt from *Thinking Fragments**

Jane Flax

Postmodern philosophies of knowledge can contribute to a more accurate and self-critical understanding of our theorizing and the intentions that underlie it. . . . However, postmodernist discourses are deficient in their treatment of issues of gender and self, and there are also important absences in their discussions of power and knowledge. Like feminist theory, postmodern philosophy is not a unified and homogeneous field. The persons and discourses associated with postmodernism include Nietzsche, Foucault, Derrida, Deleuze and Guattari, Lyotard, Rorty, Cavell, Barthes, semiotics, deconstruction, psychoanalysis, archaeology, genealogy, and nihilism.¹ Postmodernists share at least one common object of attack – the Enlightenment – but they approach this object from many different points of view and attack it with various methods and for diverse purposes.

Despite their many differences, these discourses are all “deconstructive”; they

* From Jane Flax, *Thinking Fragments: Psychoanalysis, Feminism, and Postmodernism in the Contemporary West* (Berkeley, Cal.: University of California Press, 1989). Copyright © 1989 by the Regents of the University of California. Reprinted by permission of the author and publisher.

seek to distance us from and make us skeptical about the ideas concerning truth, knowledge, power, history, self, and language that are often taken for granted within and serve as legitimations for contemporary Western culture. According to postmodernists, many of these still predominant ideas are derived from the distinctive set of philosophical and political assumptions characteristic of Western thinking at least since the Enlightenment. Hence they seek to displace the metanarrative of Enlightenment through a variety of rhetorical strategies. . . . To carry out this deconstruction, postmodernists construct stories about the Enlightenment in which the disparate views of a variety of thinkers, including Descartes, Kant, and Hegel, are integrated into (and reduced to) one “master narrative.” This master narrative then serves as an adversary against which postmodernist rhetoric can be deployed.

According to postmodernists, “the Enlightenment” story has these major themes and characters:²

1. A coherent, stable self (the author). The most distinctive and valued property of this Enlightenment self is a form of reason capable of privileged insight into its own processes and into the “laws of nature.” . . .

2. A distinctive and privileged mode of story telling – philosophy (the critic and judge). The philosopher stipulates the criteria for adequate story telling, and, it turns out, only philosophy can fully satisfy these criteria. Only philosophy can provide an objective, reliable, and universalizable “foundation” for knowledge and for judging all truth claims.

3. A particular notion of “truth” (the hero). True knowledge represents something “real” and unchanging (universal) about our minds or the structure of the natural world. The “real” is that which has an existence independent of the knower; it is not merely created or transformed by the mind in the process of knowing.

4. A distinctive political philosophy (the moral) that posits complex and necessary interconnections between reason, autonomy and freedom. . . . The Enlightenment hope is that utilizing knowledge in the service of legitimate power will assure both freedom and progress. Knowledge can then be both “neutral” (e.g., grounded in universal reason, not in particular “interests”) and socially beneficial.

5. A transparent medium of expression (language). Enlightenment philosophers posit or presume a realist or correspondence theory of language in which objects are not linguistically or socially constructed; they are merely made present to consciousness by naming or by the right use of language.

6. A rationalist and teleological philosophy of history (the plot). Events in the plot do not occur randomly; they are connected by and through an underlying, meaningful, and rational structure comprehensible by reason. The pre-given

purpose of history is the progressive perfection of humans and the ever more complete realization of their capabilities and projects.

7. An optimistic and rationalist philosophy of human nature (character development). Humans are said to be intrinsically good, able to reason and to be rationally governed. Goodness will naturally unfold and be expressed as people's external circumstances become more favorable (e.g., as authority becomes enlightened, and the natural world is better controlled and utilized through science).

8. A philosophy of knowledge (an ideal form). Science serves as the exemplar of the right use of reason and the paradigm of all true knowledge. Science "progresses" (e.g., acquires ever more accurate knowledge of the "real" world) by applying and improving its own unique "logic of discovery." The objects of scientific investigation exist "out there," independent of the scientist or subject.

Postmodern philosophers try to reveal the internally contradictory nature of each of these claims. They also posit a set of ideas at least partially outside Enlightenment beliefs. Postmodern philosophers also claim their deconstructions can open up spaces in or from which different and more varied ideas and practices may begin to emerge. The partial and problematic qualities of their achievements and claims can best be seen when postmodernists enter into conversations with psychoanalysts and feminists.

"Masters of Suspicion": Postmodern Positionings

For someone accustomed to more conventional philosophies, reading the postmodernists can be a frustrating endeavor. These authors do not offer a set of logical and sustained arguments or a synthetic or even coherent viewpoint. Instead they present a series of "positions" and a heterogeneous polyphony of voices.³ This style or styles is congruent with postmodernism itself. Among the characteristic traits and purposes of postmodernist thought is the displacement of epistemology and metaphysics by rhetoric. Postmodernists intend to replace the search for and enunciation of truth, which they believe has dominated Western philosophy since Plato, with the art of conversation or persuasive speech. In conversation the philosopher's voice would be no more authoritative than any other. A problem is that this voice still tends to override or direct too many others. It also retains the privilege of defining what "game" is to be played and its rules.

Certain themes, devices, and moves recur in postmodernist rhetoric. Radical and dramatic claims are frequently put forth and tend to cluster around certain highly charged themes. One of the most important claims is that Western culture is about to experience or has already experienced, but has been denying, an interrelated series of deaths. These include the deaths of Man, History, and Metaphysics. Postmodernists' "death" announcements are dramatic proclama-

tions and partially metaphorical ways of stating a complex set of interdependent ideas. At this point I will only indicate some of the information each is meant to convey.

1 The Death of Man

Postmodernists wish to destroy all essentialist concepts of human being or nature. They consider all concepts of Man to be fictive devices that acquire a naturalistic guise both in their construction and in repeated use within a language game or set of social practices. In order to become authoritative in a culture dominated by the "will to truth," the conventional origins of all concepts of Man must be disguised. In fact Man is a social, historical, or linguistic artifact, not a noumenal or transcendental Being.

In their view Man is actually "decentered." His attempts to impose a fictive or narrative order or structure on experience or events are constantly preconstituted and undermined by desire, language, the unconscious, and the unintended effects of the violence required to impose such an order. Man is forever caught in the web of fictive meaning, in chains of signification, in which the subject is merely another position in language.

As a purely fictive character, Man has nothing that could serve as the basis for his stepping outside this web or for breaking free from it. There is no "Archimedes point," no moment of autonomy, no pure reason or constituting consciousness with independent, nonlinguistic, or nonhistorical access to the Real or Being of the World.

2 The Death of History

The idea that History has any intrinsic order or logic is another fiction of Man. Man constructs stories he calls History in order to find or justify a place for himself within time. . . . He creates "master narratives" in which History is his, the subject's, coming to Be in and through time.⁴ At the end of this story/time, Man's reason or labor will be made fully Real, and thus nothing will be alien to or estranged from him. . . .

The idea that History exists for or is his Being is more than just another precondition and justification for the fiction of Man. This idea also supports and underlies the concept of Progress, which is itself such an important part of Man's story. The notion of Progress depends on the idea that there is some pre-given goal toward which Man is steadily moving. This goal or purpose is meant for Man; it expresses or realizes him at his best. The closer he comes to it, the closer he comes to himself, to his essence.

Such an idea of Man and History privileges and presupposes the value of unity, homogeneity, totality, closure, and identity. This story requires positing one innate quality of Man that is best – reason, the capacity to labor, or the political life. All other qualities of Man should be subordinated to or serve the One. In juxtaposition to this story and to displace it, the postmodernists tell another, different one. The real is flux. History is a series of random events with

no intrinsic order and no necessary laws that produce causality or even continuity. There is thus no empirical or logical reason to privilege unity, homogeneity, closure, or identity over difference, heterogeneity, alterity, and openness.

Furthermore there may be ethical or political reasons for reversing the value placed on unity over difference or homogeneity over heterogeneity. In order to make the whole appear Rational, the contradictory stories of others must be erased, devalued, suppressed. Any appearance of unity presupposes and requires a prior act of violence. Only by forcibly suppressing elements of the flux can History acquire a structured and unitary appearance.

It follows from this view of the Real and History that conflict and violence are endemic to the story of Man in time. There is no end to History, no closed totality in which the “stages” are pre-given, cumulative, irreversible, or progressive. There can be no guarantee that after a finite amount of struggle our work will succeed and be finished forever. The always temporary victor of a particular conflict may succeed in imposing his story as the whole truth, but all such victories and stories are in principle unstable and reversible. Furthermore no combatant can justly claim, though many do, to be merely the vehicle or instrument for an extrahistorical or social Good. There is no transcendental or disinterested position from which such a Good could be identified or from which it could be said that the Truth or the Good did in fact triumph in a particular instance.

3 The Death of Metaphysics

Western philosophy has been under the spell of the “metaphysics of presence” at least since Plato.⁵ Most Western philosophers took as their task the construction of a philosophic system in which something Real would and could be represented in thought. This Real is understood to be an external or universal subject or substance, existing “out there” independent of the knower. The philosopher’s desire is to “mirror,” register, mimic, or make present the Real. Truth is understood as correspondence to it.

For postmodernists this quest for the Real conceals most Western philosophers’ desire, which is to master the world once and for all by enclosing it within an illusory but absolute system they believe represents or corresponds to a unitary Being beyond history, particularity, and change. In order to mask his idealizing desire, the philosopher must claim that this Being is not the product, artifact, or effect of a particular set of historical or linguistic practices. It can only be the thought of the Real itself.

The philosopher also obscures another aspect of his desire: to claim a special relation and access to the True or Real. He claims that, in a sense, the presence of the Real for us depends on him – the clarity of his consciousness, the purity of his intention. Only the philosopher has the capacity for Reason, the love of wisdom (philo-sophia), the grasp of method, or the capacity to construct a logic adequate to the Real. Just as the Real is the ground of Truth, so too philosophy as the privileged representative of the Real and interrogator of truth claims must play a “foundational” role in all “positive knowledge.” . . .

Postmodernists attack the “metaphysics of presence” and the Western phi-

osopher's self-understanding in a number of ways. They question the philosophies of mind, truth, language, and the Real that underlie and ground any such transcendental or foundational claims. However, from feminist and psychoanalytic perspectives it sometimes appears that the underlying purposes of this attack are unclear and ambiguous. At times it seems to me that postmodernists are engaged in the same strategic operations in relation to modern philosophies that Kant applied to older concepts of reason: to subject them to critique in order to resituate them on firmer ground over which the philosopher can then reassert the continuing legitimacy of his exclusive command. Nonetheless, postmodernist critiques of the "metaphysics of presence" in many ways complement, correct, and strengthen psychoanalytic and feminist deconstructions of mind, truth, language, reality, and philosophy. Feminist theories are much more sensitive to the "play" of gender (including its obscured presence in postmodernisms), and both feminists and psychoanalysts have clearer understandings of the complexities of subjectivity and selfhood.

Postmodernists' positions on "metaphysics" include:

1. *"Metaphysical Minds."* There is and can be no transcendental mind; on the contrary postmodernists claim that what we call the mind or reason is only an effect of discourse. There are no immediate or indubitable features of mental life. Sense data, ideas, intentions, or perceptions are already preconstituted. Such experiences only occur in and reflect a variety of linguistically and socially pre-determined practices. . . .
2. *"Metaphysical" Truth.* Truth for postmodernists is also an effect of discourse. Each discourse has its own distinctive set of rules or procedures that govern the production of what is to count as a meaningful or truthful statement. Each discourse or "discursive formation" is simultaneously enabling and limiting. The rules of a discourse enable us to make certain sorts of statements, but the same rules force us to stay within the system and to make only those statements that conform to these rules. A discourse as a whole cannot be true or false because truth is always contextual and rule-dependent. Instead discourses are local, heterogeneous, and incommensurable. No non-discourse-dependent or transcendental rules exist that could govern all discourses or a choice between them. Truth claims are in principle "undecidable."
3. *"Metaphysical Language."* Postmodernists claim that notions of language as a transparent or neutral medium are wrong. Each of us is born into an ongoing set of language games that we must learn in order to be understood by and to understand others. The meaning of our experience and our understanding of it cannot be independent of the fact that such experience and all thought about it are grasped and expressed in and through language. To the degree that thought depends on language, thought and "the mind" itself will be socially and historically constituted. . . .

4. *The “Metaphysics” of Reality.* The Real is unstable and perpetually in flux. Western metaphysics creates a false appearance of unity by reducing the flux and heterogeneity of experience into binary and supposedly natural or essentialist oppositions that include identity/difference, nature/culture, truth/rhetoric, speech/writing, and male/female. . . . The members of these binary pairs are not equal. Instead the first member of each is meant to dominate the second, which becomes defined as the “other” of the first. . . . The other has no independent or autonomous character of its own; for example, “woman” is defined as a deficient man in discourses from Aristotle through Freud. Once these oppositions are seen as fictive, asymmetric, and conditions of possibility for the philosopher’s story then a premise that underlies all variants of the metaphysics of Presence can be revealed: To be other, to be different than the defining One is bad. It is better to be defined and determined as the lesser other of the One than to be outside Being altogether.

5. *The “Metaphysics” of Philosophy.* Philosophy is necessarily a fictive, non-representational activity. As a product of the human mind, philosophy has no special relation to Truth or the Real. The philosopher merely creates stories about these concepts and about his own activities. His stories are no more true than any other. There is no way to test whether one story is closer to the truth than another because there is no transcendental standpoint or mind unenmeshed in its own story. Philosophers should seek instead an infinite “dissemination” of meanings. They should abjure any attempt to construct a closed system in which the other, “deferred,” or “excess” are “pushed to the margins” and made to disappear in the interest of coherence and unity. . . .

The Emergence of a Distinctively Feminist Question: The “Other” Says No

In 1949 Simone de Beauvoir, one of the founding mothers of contemporary feminist theory, described the constricting and constricted lives of the “second sex.” De Beauvoir delineated the many ways in which “woman” is defined and limited in her being as the (always lesser) “other” to man. In male-dominant cultures no woman escapes the consequences of such a position. Even the most “independent” woman is still mutilated and deformed by the ideas and social relations that more deeply affect her less fortunate sisters.

. . . No particularly visible or active women’s movements existed when de Beauvoir wrote her book. Perhaps even she could not have anticipated – although surely she hoped for – some of the remarkable (but far from sufficient) changes in gender relations that have occurred since the re-emergence of feminism in the late 1960s. Feminist theorists are deeply indebted to these women’s movements. For many, including myself, participating in consciousness raising sessions and other movement activities forced into awareness aspects of experience that we had too often taken for granted. Such experiences included the fear of rape and unwanted pregnancy, the absence of female professors, the masculinist bias of

many academic fields, the violence exercised against women, the restrictions on and distortions and exploitations of women's sexuality, the sexual division of labor, and our exclusion from most positions of political and economic power.

Like many other women I sought to make sense of and to contribute to these transformations in my and others' consciousness and to translate our developing ideas into social and political changes. Like many academic and intellectual women I attempted to fit what I was learning about women's experiences and histories (outside the "mainstream" of academia) into pre-existing theoretical frameworks (liberalism, Marxism, psychoanalysis, critical theory), only to find that these could not account for much of this material. In fact it gradually became evident that these frameworks also were not free from the effects of gender and hence ultimately inhibited our understanding. Emboldened and prodded by the existence of increasingly diverse and active women's movements and the unsatisfactory results of our attempts to simply "add women" and "stir" us into pre-existing ways of thinking and being, many feminist theorists have come to believe we have no choice but to go beyond the "given world."⁶

. . . That men appear to be and in many cases are the wardens, or at least the trustees, within a society should not blind us to the extent to which they too are governed by the rules of gender. However, contrary to the views of some postmodernists, this does not mean that men and women occupy a fundamentally equivalent status – as "split" signified/subjects. One of the distinguishing features of feminist theories is the claim that gender relations, at least as they have been organized so far, are (variable) forms of domination. Feminist theorists are motivated in part by an active concern with justice and a desire to contribute to the overcoming of women's subordinations. The inequalities among men matter a great deal – to individual men, to the women and children connected to them, and to those concerned about justice. Nonetheless these do not negate and should not obviate the fact that men as a group remain privileged relative to most women in most societies and that there are systematic forces that generate, maintain, and replicate gendered relations of domination. One of the purposes of the study of gender for feminist theorists is to understand these forces as they operate in specific societies with the hope that such understanding may contribute to eliminating gender domination. . . .

Inserting both men and women within contexts of the social relations of gender has had a paradoxical effect on the status and self-understanding of feminist theorists. Feminists have begun to ask one another a number of important questions about the status of our stories about gender. Both psychoanalytic and postmodernist theories can be useful to (and in fact in some cases have stimulated) the further working out of these questions. If both men and women are formed in and through gender systems, then the thinking of women (or feminists) as well as that of men (or nonfeminists) must be shaped in complex and sometimes unconscious ways by gender relations. How can such stories in any sense be more true, more accurate, less distorted, or more "objective" than others? Are the stories feminists tell about gender more privileged, more deserving of our attention or respect? Or are they just different – an-other voice or a (hopefully) welcome, dissonant strain within the "conversation of mankind [sic]?"⁷

... Feminists have been attracted to at least two very different concepts of the project of feminist theorizing. One conception derives from Enlightenment ideas about knowledge, truth, and freedom; the other derives from postmodernist critiques of these ideas. Feminist theorists have tried to maintain two different epistemological positions. The first is that the mind, the self, and knowledge are socially constituted, and what we can know depends on our social practices and contexts. The second is that feminist theorists can uncover truths about the whole as it "really is." Those who support the second position reject many postmodern ideas and must depend upon certain assumptions about truth and the knowing subject that I find increasingly problematic. To attain such a truth (e.g., the "real" explanation for gender arrangements at any time is x . . .) would require the existence of an "archimedes point" outside of current, social, and self-comprehension and beyond our embeddedness in it. From this point we could see and represent an "objective" view of the whole. What we see and report would have to be untransformed by the activities of perception and of reporting our vision in language. The object seen (social whole or gender arrangement) would have to be apprehended by a mind sufficiently empty of the biases of its society and nearly perfectly transcribed by and into a transparent language.

This sort of "truth" is the necessary ground for a "feminist standpoint" that could be more true than previous (male) ones. The notion of a feminist standpoint, equivalent both epistemologically and ethically to the status Marx and Lukàcs assign to that of the proletariat, has been very productive for and influential in the development of feminist theories, but it is highly problematic.⁸ It depends on unexamined and questionable assumptions and motivations, including an optimism that people will act rationally on their "interests" and that reality has a structure that a more perfect reason can discover more perfectly. Both these assumptions in turn depend on an uncritical appropriation of the Enlightenment ideas discussed above. Furthermore the notion of such a "standpoint" assumes that the oppressed are not in some fundamental ways damaged by their social experience. On the contrary this position assumes that the oppressed have a privileged, unitary, and not just different relation to and ability to comprehend a reality that is "out there" waiting for our representation.

This view also presupposes gendered social relations in which there is a category of beings who are or can be fundamentally like one another by virtue of their sex – that is, it assumes the uniform otherness men assign to women. Such a standpoint requires that women, unlike men, can be free from determination by their own participation in relations of domination, for example, those rooted in the social relations of race, class, or homophobia. Somehow all these barriers to objectivity will be cleared away leaving only an unmediated relation to truth and reality.

... Anyone contemplating the history of the West in the twentieth century has a right to be skeptical of its self-representation as having substituted reason and law for authority or the resolution of conflict. As Weber argues, the rule of law is not totally other than, independent of, or exempt from force or violence.⁹ Any culture that retains the possibility of nuclear annihilation as the last resort

for its “defense” seems to me trapped more within Kafka’s nightmare world than in the sunnier one of Kant’s categorical imperative. Hence we remain too much within the terms of the ruling discourse or set of illusions if we hope that truth or a search for it may set us free. Furthermore this hope could be dangerous. Under its spell we may find ourselves caught up in complicity with dangerous transcendental illusion(s): of the possibility of a “real” nonconflictual entity, a “nostalgia” for the “whole or the one,” or a belief that one can “seize reality” once and for all – illusions that can produce only a “return to terror,” of which our century has certainly had more than enough.¹⁰ . . .

Notes

- 1 Especially influential works include Friedrich Nietzsche, *Beyond Good and Evil* (New York: Vintage, 1966), and his *The Will to Power* (New York: Vintage, 1968); Michel Foucault, *Power/Knowledge: Selected Interviews and Other Writings 1972–77*, ed. Cohn Gordon (New York: Pantheon, 1980), and his *Language, Counter-Memory, Practice*, ed. Donald F. Bouchard (Ithaca, N.Y.: Cornell University Press, 1980); Jacques Derrida, *Marges de la philosophie* (Paris: Editions de Minuit, 1972), and his *Writing and Difference*, trans. Alan Bass (Chicago: University of Chicago Press, 1978); Giles Deleuze and Felix Guattari, *On the Line* (New York: Semiotext[e], 1983), and their *Anti-Oedipus: Capitalism and Schizophrenia* (Minneapolis: University of Minnesota Press, 1983); Stanley Cavell, *The Claim of Reason* (New York: Oxford University Press, 1979); Roland Barthes, *S/Z*, trans. Richard Miller (New York: Hill & Wang, 1974), and his *The Fashion System*, trans. Matthew Ward and Richard Howard (New York: Hill & Wang, 1983). Already there is a large and ever-growing literature on and in postmodernism. Among the works I have found most helpful are Terence Hawkes, *Structuralism and Semiotics* (Berkeley and Los Angeles: University of California Press, 1977); Herbert L. Dreyfus and Paul Rabinow, *Michel Foucault: Beyond Structuralism and Hermeneutics* (Chicago: University of Chicago Press, 1982); Harvey West, ed., *The Idea of the Post-Modern* (Seattle: Henry Art Gallery, University of Washington, 1981); Quentin Skinner, ed., *The Return of Grand Theory in the Human Sciences* (New York: Cambridge University Press, 1985); Michael Ryan, *Marxism and Deconstruction: A Critical Articulation* (Baltimore: Johns Hopkins University Press, 1982); Vincent Descombes, *Modern French Philosophy* (New York: Cambridge University Press, 1982); Fredric Jameson, “The Cultural Logic of Capital,” *New Left Review*, 146 (July–August 1984), pp. 53–92; Henry Louis Gates, Jr., ed., *“Race,” Writing and Difference* (Chicago: University of Chicago Press, 1986); John Rajchman and Cornel West, eds, *Post-Analytic Philosophy* (New York: Columbia University Press, 1985); Christopher Norris, *Derrida* (Cambridge, Mass.: Harvard University Press, 1987); and *Feminist Studies*, 14, no. 1 (Spring 1988).
- 2 What follows is a summary of some of the ideas of Derrida, Foucault, Lyotard, and Rorty. For more detail and differentiation see Chapter 6 [of Flax, *Thinking Fragments* (not included here)].
- 3 Cf. Jacques Derrida, “Positions,” in Jacques Derrida, *Positions*, trans. Alan Bass (Chicago: University of Chicago Press, 1981); and Foucault, “Two Lectures,” in Foucault, *Power/Knowledge*.
- 4 Jean-François Lyotard, *The Postmodern Condition: A Report on Knowledge* (Minneapolis: University of Minnesota Press, 1984), pp. 27–41.

- 5 Derrida, "Violence and Metaphysics," in Derrida, *Writing and Difference*.
- 6 A representative sample of contemporary feminist theorists would include Barbara Smith, ed., *Home Girls: A Black Feminist Anthology* (New York: Kitchen Table: Women of Color Press, 1983); Cherrie Moraga and Gloria Anzaldua, eds, *This Bridge Called My Back* (Watertown, Mass.: Persephone Press, 1981); Elizabeth Abel, Marianne Hirsch, and Elizabeth Langland, *The Voyage In: Fictions of Female Development* (Hanover, N.H., and London: University Press of New England, 1983); Zillah R. Eisenstein, ed., *Capitalist Patriarchy and the Case for Socialist Feminism* (New York: Monthly Review Press, 1979); Vivian Gornick and Barbara K. Morgan, eds, *Woman in Sexist Society* (New York: Mentor, 1971); Annette Kuhn and Ann Marie Wolpe, eds, *Feminism and Materialism* (Boston: Routledge & Kegan Paul, 1978); Hunter College Women's Studies Collective, *Women's Realities, Women's Choices* (New York: Oxford University Press, 1983); Elaine Marks and Isabelle de Courtivron, eds, *New French Feminisms* (New York: Schocken Books, 1981); Joyce Trebilcot, ed., *Mothering: Essays in Feminist Theory* (Totowa, N.J.: Rowman & Allanheld, 1984); Sherry B. Ortner and Harriet Whitehead, eds, *Sexual Meanings: The Cultural Construction of Gender and Sexuality* (New York: Cambridge University Press, 1981); Nancy C. M. Hartsock, *Money, Sex and Power* (New York: Longman, 1983); Ann Snitow, Christine Stansell, and Sharon Thompson, eds, *The Powers of Desire: The Politics of Sexuality* (New York: Monthly Review Press, 1983); Sandra Harding and Merill B. Hintikka, eds, *Discovering Reality: Feminist Perspectives on Epistemology, Metaphysics, Methodology and Philosophy of Science* (Boston: D. Reidel, 1983); Carol C. Gould, *Beyond Domination: New Perspectives on Women and Philosophy* (Totowa, N.J.: Rowman & Allanheld, 1984); Allison M. Jagger, *Feminist Politics and Human Nature* (Totowa, N.J.: Rowman & Allanheld, 1983); Martha Blaxall and Barbara Reagan, eds, *Women and the Workplace* (Chicago: University of Chicago Press, 1976); Isaac D. Balbus, *Marxism and Domination* (Princeton, N.J.: Princeton University Press, 1982); bell hooks, *Feminist Theory: From Margin to Center* (Boston: South End Press, 1984); Audre Lorde, *Sister Outsider* (Trumansberg, N.Y.: Crossing Press, 1984); Gloria T. Hull, Patricia Bell Scott, and Barbara Smith, *All the Women are White, All the Black are Men, But Some of Us are Brave: Black Women's Studies* (Old Westbury, N.Y.: Feminist Press, 1982); Sandra Harding, *The Science Question in Feminism* (Ithaca, N.Y.: Cornell University Press, 1986); and Virginia Sapiro, *The Political Integration of Women* (Urbana: University of Illinois Press, 1984). On the history of the "second wave" of feminism, see Vicky Randall, *Women and Politics: An International Perspective*, 2nd edn (Chicago: University of Chicago Press, 1987); Ethel Klein, *Gender Politics* (Cambridge, Mass.: Harvard University Press, 1984); and Sara Evans, *Personal Politics* (New York: Vintage, 1980).
- 7 This is Richard Rorty's phrase in his *Philosophy and the Mirror of Nature* (Princeton, N.J.: Princeton University Press, 1979), pp. 389–94.
- 8 For discussion of the feminist standpoint, see Hartsock, "The Feminist Standpoint," in Harding and Hintikka, *Discovering Reality*, and her *Money, Sex and Power*; and Harding, *The Science Question*, Chaps 6 and 7.
- 9 See Max Weber, "Politics as a Vocation," in *From Max Weber*, ed. H. H. Gerth and C. Wright Mills (New York: Oxford University Press, 1958); and Max Horkheimer and Theodor W. Adorno, *Dialectic of Enlightenment* (New York: Herder & Herder, 1972).
- 10 Jean-François Lyotard, *The Postmodern Condition: A Report on Knowledge* (Minneapolis: University of Minnesota Press, 1984), pp. 81–2.

56 Metaphysics and Feminist Theory: Excerpts from “Feminist Metaphysics” and “Anti-Essentialism in Feminist Theory”*

Charlotte Witt

In this essay I explore the idea that feminist theory has consequences for metaphysics. What sort of consequences? Two claims are particularly important. The first claim, which I address here, is that every acceptable feminist theory is inherently anti-metaphysical.¹ Further, because some feminists equate metaphysics with essentialism, some have argued that feminist thinking must be anti-essentialist. On this view, then, the contribution of feminist theory to metaphysics is the metaphilosopical position that one ought to stop doing metaphysics, and/or reject essentialism, in order to theorize in an appropriately feminist manner. The second major claim, which I defend elsewhere, is that feminist theory makes a distinctive contribution to metaphysics.² Here the point is the reverse of the first: Not only is feminist theory not inherently anti-metaphysical, but it contributes significantly to our understanding of at least some metaphysical issues.

What are the metaphilosopical consequences of feminist thought for metaphysics? I argue that there are no specifically *feminist* reasons for rejecting either metaphysics or essentialism. My argument has three parts. Because it would be impossible to devise an argument in principle concerning the implications for feminist theory of every possible anti-metaphysical position, I consider the anti-metaphysical claims of Richard Rorty and Jean-François Lyotard, whose views have currency in feminist circles. I show that their positions neither originate in feminist concerns nor provide an adequate basis for social change.

Second, I consider a historical argument that might be thought to provide the specifically feminist motivation for the rejection of metaphysics. I argue that the feminist claim that the history of philosophy – including metaphysics – in the Western tradition is phallocentric, or male-biased, requires showing not only that women have been excluded from traditional philosophical categories but also that their inclusion makes a conceptual difference. Feminists who are developing a gender critique of the tradition (and I am one of them) must also engage in the project of reconceiving traditional categories; they ought to do metaphysics. Rather than providing the specifically feminist reasons for adopt-

* “Metaphysics and Feminist Theory” incorporates passages from the following: “Feminist Metaphysics,” in Louise M. Antony and Charlotte Witt, eds, *A Mind of One’s Own: Feminist Essays on Reason and Objectivity* (Boulder, Col.: Westview Press, 1993); and “Anti-Essentialism in Feminist Theory,” *Philosophical Topics*, 23 (1995), pp. 321–44. Reprinted by permission of the author, *Philosophical Topics*, and Westview Press.

ing an anti-metaphysical metaphilosophy, the gendered reading of the history of philosophy itself requires metaphysical thinking.

But, is the gendered reading of the history of philosophy itself a remnant of metaphysical thinking that ought to be rejected by feminists because of its essentialist implications? I end by considering the claim, made by many postmodern feminist philosophers, that using gender to interpret the history of philosophy commits the sin of gender essentialism, and essentialism ought to be rejected by feminists. I argue that the central arguments against gender essentialism made by postmodern feminist thinkers are mistaken. Hence, there is no reason for feminists to reject readings of the history of philosophy that use categories of gender, which is a good thing since understanding the sexism of the philosophical canon requires the use of these categories.

Feminist Theory, Postmodernism, and the Rejection of Metaphysics

Let us begin with the claim that metaphysics ought to be rejected root and branch because it is an inherently masculine and oppressive enterprise. As I see this issue, the central question is whether there are any specifically feminist reasons for rejecting metaphysics. By a feminist reason, I mean either a reason that is clearly and directly related to the experience of women or a reason that is needed to explain or to ameliorate that situation. If, for example, you used the historical claim that the metaphysical tradition is male-biased to explain why women are thought to lack the power of transcendent reason in our culture, then you would be giving a feminist reason in my sense of the term. I argue that there are no feminist reasons to accept either the postmodern or the pragmatist rejection of metaphysics. Moreover, these positions lack the theoretical resources required for an adequate feminist criticism of patriarchy's ideology and institutions.

Feminist theorists Nancy Fraser and Linda Nicholson describe what contemporary pragmatism and postmodernism have in common:

Writers like Richard Rorty and Jean-François Lyotard begin by arguing that Philosophy with a capital "P" is no longer a viable or credible enterprise. From here, they go on to claim that philosophy, and by extension, theory more generally, can no longer function to *ground* politics and social criticism. With the demise of foundationalism comes the demise of the view that casts philosophy in the role of *founding* discourse vis-à-vis social criticism.³

The idea common to both Rorty and Lyotard is that philosophy must give up its self-image as engaging in a unique type of foundational inquiry, and accept a new status as one kind of inquiry among others.

Rorty and Lyotard differ with regard to their reasons for rejecting metaphysics. Rorty rejects what he calls "representationalism" because it assumes the possibility of a correct set of representations or a privileged description on the

one hand, and reality, that which is to be represented or described, on the other hand. Rorty's rejection of this foundationalist picture of philosophy is a result of his pragmatist metaphilosophy: "Anti-representationalists do not think such efforts (of representationalists like Thomas Nagel) insane, but they do think that the history of philosophy shows them to have been fruitless and undesirable."⁴ Metaphysics has outlived whatever usefulness it once had, and it should be retired like the rotary telephone.

Lyotard shares Rorty's rejection of philosophy as a foundational discourse not because he shares Rorty's pragmatist metaphilosophy, but because the idea of a "metadiscourse," like philosophy, no longer has legitimacy in our times, the postmodern period. Lyotard defines postmodernism as "incredulity toward metanarratives" and contrasts the postmodern attitude toward justification with the modern attitude: "I will use the term *modern* to designate any science that legitimates itself with reference to a metadiscourse of this kind making an explicit appeal to some grand narrative, such as the dialectics of Spirit, the hermeneutics of meaning, the emancipation of the rational or working subject, or the creation of wealth."⁵ Metaphysics, as traditionally conceived, is a paradigmatic "grand narrative."

Why do postmoderns reject metanarratives in principle rather than just rejecting particular metanarratives like those of Hegel, Marx or the Enlightenment? According to Lyotard, it is because every discourse (or language) is embedded in contingent, historical practices (its rules): "Any consensus on the rules defining a game and the 'moves' playable within it *must* be local, in other words, agreed on by its present players and subject to eventual cancellation."⁶ If *all* discourse, *all* language games must be local, then metanarratives are really just pretentious local narratives that happen to be widely believed.

It is certainly possible that a feminist theorist might find Rorty's or Lyotard's rejection of metaphysics intrinsically plausible. One might think with Rorty that metaphysics has served its function and just doesn't do anyone any good: "Whatever good the ideas of 'objectivity' and 'transcendence' have done for our culture can be attained equally well by the idea of a community which strives after both intersubjective agreement and novelty – a democratic, progressive, pluralistic community of the sort of which Dewey dreamt."⁷ Or one might agree with Lyotard that the hallmark of our age is the realization that all descriptions of reality are context-bound and perspectival and that, therefore, no discourse can claim to transcend its own perspective and context.

It is clear, however, that there are no feminist reasons for accepting either explanation of the petering out of metaphysics. For, whatever their intrinsic philosophical plausibility, neither theory reflects feminist insights and neither explains the status of women or their exclusion from culture. And, perhaps most important, neither view provides an adequate theoretical basis for the kind of conceptual critique or social change that is central to feminism.⁸

Rorty's celebration of modern, liberal democracy and his rejection of the utility of any deep criticism of existing social categories ought to make feminists who accept his metaphilosophical views on their intrinsic merit reconsider them in the light of their political implications. In "Feminism and Pragmatism" Rorty

argues that prophetic feminists like Catharine MacKinnon and Marilyn Frye are engaged in a process of invention.⁹ They are creating women as full human persons rather than discovering that that is what women really are and have been all along despite their oppression. For the pragmatist there is nothing that women (or slaves or homosexuals) really are and have been all along. Initially, Rorty's description of prophetic feminism as creating a new reality rather than merely discovering and articulating the biases and limitations of traditional descriptions of reality seems attractive and innocuous. But it is important to read the fine print. In following Rorty and replacing talk of discovery with talk of invention, feminists must also relinquish "the notion that the oppression of women is *intrinsically* abominable" and "the claim that there is something called 'right' or 'justice' or 'humanity' which has always been on their side making their claims true."¹⁰ Feminists who are good pragmatists cannot advocate for feminism or explain its success by appeal to its truth or moral rightness.¹¹ Pragmatism's seductive metaphors of creativity turn out to disarm feminism by undercutting the kinds of criticisms that feminists of all types want to make, and ought to make, of patriarchy.

Similarly, feminists who find the arguments of postmodernism intrinsically convincing should consider the implications for political theory and political change of postmodernism's thoroughgoing relativism.¹² Postmodernism provides meager resources for critical political thought. If all discourse is radically contextualized and local (entirely relative to historical groups or cultures) then feminist criticisms are themselves contextualized, and merely local. But if this is so, what grounds are there to claim that patriarchal discourse (the language game of male power that is our local lingo) ought to change?

Gender and Reason: the Historical Argument

At this point some feminist anti-metaphysicians might object that I have told only half of their story, and it is the other half that provides the *feminist* reasons for the rejection of metaphysics. The major feminist argument against traditional metaphysics is historical. In *The Man of Reason* for example, Genevieve Lloyd makes a historical argument that the concept of transcendence, the ability of reason to transcend material and temporal limits, is male-centered; transcendence means overcoming nature, matter, the feminine; and the feminine itself has been partly constituted by its occurrence within this symbolic structure.¹³

What are the implications for feminists of the fact that metaphysics, the transcendence of reason, was historically conceived of in relation to men and not women? There are three possibilities: (1) One could argue for a clean break with the tradition, a rejection of metaphysics. (2) One could argue that the feminist understanding of the tradition gives good reasons for engaging in a reconception of the central categories of metaphysics like reason and transcendence. (3) One could argue that all that is required in order to address the demonstrated male bias in traditional philosophy is the explicit inclusion of women in the domain of reason.

It is clear that the historical argument does not necessitate drawing the first conclusion. There are three objections. First, it does not follow from the fact that metaphysical categories have reflected male experience in the past that they need do so in the future. Second, the case for male bias cannot be made based on the historical evidence alone because we need to show not only that women were excluded in the tradition, but also that their exclusion has made a conceptual difference. Third, the historical argument (whether supplemented by a conceptual investigation or not) is insufficient to discredit metaphysical inquiry itself.

This last point, that the historical argument is not powerful enough to discredit metaphysical inquiry, does not concern the ideas generated by metaphysical reflection, but rather the activity itself. A feminist might argue that reflective philosophy is a waste of time or inherently elitist, but the historical argument is directed against the fruits of philosophical reflection and not against the activity itself. So, although the historical criticism of the history of metaphysics, unlike Rortyan pragmatism and Lyotard's postmodernism, is explicitly feminist, it does not forge the connection between the metaphilosophical arguments against metaphysics we have considered, and feminist theory.

And, in fact, although some feminist theorists appear to want to reject traditional philosophy altogether by a proposed "break with modernism," what some envision is a project of reconceptualization of the kind suggested in (2) above. "Male-biased (or patriarchal) conceptual frameworks must be replaced by ones that are not male-biased."¹⁴ I take it that to undertake a reconception of basic metaphysical categories is to do metaphysics. In the end, then, feminist criticisms of male-centered metaphysical categories should motivate theoretically inclined feminists to reform or revolutionize metaphysics rather than to abandon it. If feminists do not engage in this project then they are left with (3) – the idea that the traditional categories are fine as they are and that women simply need to be included explicitly.

Let me draw together the themes I have been discussing by considering how a feminist thinker might bolster the historical argument with a deconstruction of metaphysics. The historical story shows that women have been excluded from traditional philosophy in several ways and, consequently, that the values of traditional philosophy (like reason) really reflect male values and norms. These values are purportedly universal, but their simple extension to women seems suspect. Why should we accept male norms as the norms? I have suggested that we feminists need to rethink the traditional categories. Another feminist, impressed by the duplicity of our tradition, might embrace a total deconstruction of the claim to universal norms. She might adopt a Rortyan or Lyotardian metaphilosophy. In doing so, however, she undercuts herself. For she wants to claim that the universal norms of the tradition are illegitimate and ought to be rejected, but neither metaphilosophy provides grounds for her claim. If your metaphilosophy does not countenance the possibility of universal norms, then how can you criticize the tradition for not furnishing them? This is a self-defeating strategy for feminists who would criticize the tradition for gender bias.

As things stand, we could conclude from the historical argument either that traditional categories need to be thought through again or that we need merely

include women in their domain. In order to establish the stronger conclusion, it is necessary to show not only that women have been excluded but also that their inclusion will make a conceptual difference. For, if their inclusion does not make a conceptual difference, then what is meant by calling the tradition “phallocentric” or “male-biased” is uncontroversial, and relatively easy to correct. So, not only does the gender critique of the tradition not provide a conclusive argument against metaphysics, it actually needs to be supplemented by a project that has the goal of showing that the inclusion of our perspective makes a conceptual difference.

Gender, Anti-Essentialism and Postmodern Feminism

But, whose perspective is “our” perspective, and what idea of gender does a gendered reading of the philosophical tradition presuppose? In arguing that the tradition is male-biased or exclusive of women, I have tacitly assumed that the genders have fixed or stable identities. This assumption, which I call “gender essentialism,” is criticized by postmodern feminists, who argue against essentialism quite generally.¹⁵ And, since many postmodern feminists equate essentialism with metaphysics, the argument against essentialism is taken to be an argument against metaphysics as well. The question of essentialism is an important issue within feminist theory itself, since the kind of gendered reading of the history of philosophy that we have been considering in the previous section presupposes fixed or stable gender identities.

Two arguments against gender essentialism are common to all feminist anti-essentialists. I call them the *core argument* and the *exclusion argument*. The core argument points out that gender is a socially constructed category, and concludes that it cannot have essential features (because only biological or natural categories could have essences). The exclusion argument claims that essentialism in feminist thought excludes certain groups of women, just as women as a group disappear in male-centered traditional philosophical theories. Since neither of these arguments is successful, the case that feminist theory must be anti-essentialist has yet to be made. And, in so far as being anti-essentialist and being anti-metaphysical amount to the same thing, the feminist case against metaphysics is not furthered by anti-essentialist arguments. Let us consider each argument in turn.

The Core Argument

Anti-essentialist feminists share two basic premises which are sometimes put together into an argument against essentialism. The first premise is an equation of essentialism and biologism. The second is that gender is socially constructed rather than given by biology or nature.¹⁶ On the assumption that essences are biologically determined or natural and that gender is socially constructed and not biologically determined, an anti-essentialist conclusion concerning gender follows.

Proponents of the core argument provide (or might provide) the following support for the equation of essentialism and biologism. First, although it is strictly speaking false to equate the two, biological descriptions are one way of specifying the essence of women, a way that has predominated in patriarchal thought both in the past and today.¹⁷ The most plausible reading of the anti-essentialist equation of biologism and essentialism is to read it as a rejection of patriarchal conceptions of gender, which have tended to be naturalistic. Feminists have very correctly been suspicious of appeals to women's nature or biology because they have been, and are, used to justify social and political injustice. For this reason, feminist theorists like Simone de Beauvoir have urged that women are made and not born. Thus the social constructivist view of gender is opposed to the naturalist position, which is equated with essentialism. Second, the rejection of biologism (and with it essentialism) can charitably be read as endorsing another view of concepts of gender, the social constructionist view. This is the idea that concepts of gender are not given in biology or nature but are produced by languages, cultures, ideologies, regimes of power, etc. Social constructionist views of gender are uniformly anti-essentialist, although they vary with regard to the question of whether or not there is anything at all given prior to, or underlying, the process of social construction.¹⁸ It is thus assumed that the thesis that gender is socially constructed, in itself, entails a rejection of essences.

Thus the issue at the heart of the core argument boils down to the question of whether the social construction of gender is incompatible with gender essentialism. The social construction of gender would be incompatible with gender essentialism if the fact that gender is social (and not natural) and the fact that it is constructed (rather than given) ruled out either (i) that there is some property necessary to my being a woman (like being nurturing, or being oppressed, or having a uterus) or (ii) that there is a core of properties that must be satisfied if something is to count as a woman. However, the fact that gender is socially constructed, or "man-made," is compatible with both ways of formulating gender essentialism. We can see this by example.

It is important to notice that my argument by example is not intended to establish either the truth of gender essentialism or the falsity of anti-essentialism. Moreover, I do not argue either that it is necessary that the genders exist or that it is necessary that they have the features that they do have. The genders are dependent upon human culture, and both the existence and features of human culture are entirely contingent. Rather, I use the example of an artifact to show that the origin of an object or category makes no difference to the question of whether or not it has an essence. But, if this is so, then the fact that gender is socially constructed rather than natural makes no difference to the question of whether or not it has an essence. Or, in other words, the core argument fails, and the case against essences must be made on other grounds.

Human inventions like machines are constructed, they are social objects and not biological givens. If there are necessary features that an object must have in order to satisfy a machine kind, then social construction is compatible with essentialism of type (ii). Consider the Coke machine. It does not follow from

the fact that being a Coke machine is a socially constructed category, rather than a natural one, that it lacks an essence. In order to be a Coke machine, a machine *must* have the function of providing a Coke in exchange for money.¹⁹ And, further, if an object is a Coke machine, it necessarily has the function of providing a Coke in exchange for money. So, it does not follow from the fact that the functional essence of a Coke machine is not a natural or biological property that it is not an essential property. Essentialism is compatible with social construction.

Similarly, it does not follow simply from the fact that gender is determined within and by a social context that gender essentialism is false. Our Coke machine example shows that it does not follow from the fact that women's essence is not biological (but rather – take your pick – psychological, legal, pornographic, symbolic) that women could not have an essence. Now, this result is modest. All I have established is that anti-essentialists are wrong when they argue, or assume, that the social construction of gender is incompatible with gender essentialism. This leaves wide open many important questions about gender essentialism, including whether there are other compelling arguments against it; whether there are any considerations in favor of it; and a whole host of epistemological issues concerning how gender essences might be determined, given the sorry spectacle of patriarchal accounts of the essential feminine. In what follows, I consider the second anti-essentialist argument which has been most influential among feminists.

The Exclusion Argument

In her enormously influential book *Inessential Woman* Elizabeth Spelman argued that feminist theory has reproduced the exclusionary practices of male thinking in the Western philosophical tradition. Spelman shows, through an examination of the roots of exclusionary, categorical thinking in Plato and Aristotle, that certain classes of women and men are erased. For example, in the distinctions Aristotle draws between men and women, and citizens and slaves, the class of women slaves is invisible. The act of categorization, of deciding what distinctions matter, is itself a political act because it reflects the interests and position of the categorizer. And this point holds when applied to feminist theory itself. Spelman argues in great detail, and persuasively, that when feminist theorists have made gender relations their central topic of inquiry, and have not considered other social classifications to be significant, they have tacitly been exploring the gender relations of their own race (white) and class (middle). The process of marginalization of race and class in feminist theory, in fact marginalizes certain women. The allegedly universal theory of women is really a theory of some women just as the allegedly universal theory of human nature is really a theory of some humans (men).

Moreover, the error of taking one group of women to be women as such is not simply an empirical mistake that can be easily corrected. Which women are the subject of gender theory is not arbitrary at all, but reflects the privileged position of the theorizer. Feminist essentialism not only excludes many groups

of women from explicit consideration in its theorizing, it creates a kind of double-privilege in that it excludes many women from both the object of inquiry and the role of inquirer, while at the same time placing other women at center stage.

Spelman is committed to the core argument. Concepts of gender are socially constructed. "But do we have gender identity in common? In one sense, of course, yes: all women are women. But in another sense, no: not if gender is a social construction and females become not simply women but particular kinds of women."²⁰ Spelman believes that essentialism is particularly ill-suited to inquiry into concepts of gender because these concepts are socially constructed rather than natural. But, as I argued above, the fact that gender is socially constructed does not count against essentialism.

According to Spelman not only is gender a social construction, but it is also a culturally relative concept. If gender is culturally defined, Spelman argues, then it must make a difference to the concept of gender what culture the women in question inhabit. Once we see that gender is culturally defined then we also see that the race, the class, or the nationality of the woman will make a great difference to what being a woman means for her. The idea is that African-American women and white women live in different cultures and that these different cultures specify the meaning of gender differently. So, it is particularly ironic that feminist inquiries into gender marginalize factors like race and class that are acknowledged to be social factors germane to concepts of gender. But what kind of a difference does difference in culture make to the concept of gender?

In order for Spelman's cultural relativism to entail anti-essentialism it has to be maintained in a strong sense. She must hold that there are no; or no significant or interesting, features common to women across cultures. The cultural relativity has to pertain to the core of gender concepts and not to their periphery. For example, pointing to the difference between foot-binding in China and cosmetic surgery in the United States does not establish significant cultural relativity in notions of gender, because the differences are not as significant as the similarity in the way that women mutilate themselves in pursuit of a culturally defined notion of physical beauty. To put the point another way, if the differences between foot-binding and cosmetic surgery are what is meant by cultural relativity, then the essentialist feminist can accept that concepts of gender are culturally relative.

For, cosmetic surgery and foot-binding, while certainly very different procedures in very different cultures, nonetheless can both be classed together as practices in which women mutilate themselves in order to satisfy a culturally defined concept of beauty. They are fundamentally alike although vastly different in detail. Their similarity only emerges at a certain level of abstraction and generality; feet are not normally the object of cosmetic surgery in the West. Since Spelman is suspicious of the philosopher's desire to classify by similarities and ignore differences, she might object to this way of approaching the question of cultural relativity. As a response it is important to consider how this level of generality allows feminist theory to gain in explanatory power. When feminists classify foot-binding and face-lifts together, as essentially the same, they

obviously make no claim to “carve nature at the joints”; these are undeniably social customs. But they do enrich the explanatory power of their position. We understand something important about women when we see both practices in terms of their similarity. What kinds of abstraction and what level of generality are appropriate in a given case is, of course, an on-going question for both essentialist and anti-essentialist feminists.²¹

Spelman's book is a cautionary tale about the political and intellectual abuses of a facile essentialism rather than an argument against the possibility of significant common characteristics among women. It invites feminist theorists to listen to women who inhabit different cultures, and to recognize the political and normative dimensions of the phrase “different cultures.” It argues eloquently for attention to be paid to the political, both with regard to the objects of feminist inquiry and to the subjects who are inquirers. Finally, however, the question of whether or not there are significant similarities among women is left open as Spelman details the false assumptions, projections, and power relations that accompany the investigation of gender.

The exclusion argument does not provide the missing argument that allows one to move from the social construction of gender to anti-essentialism. Even supplemented by a weak thesis of cultural relativity, the exclusion argument fails to forge the missing link. Hence, we can conclude that the exclusion argument is compatible with either essentialism or anti-essentialism. Neither the core argument nor the exclusion argument justifies the position that feminist theory must be anti-essentialist. Hence, there is no reason for feminists either to give up the idea that there are stable gender identities, or to give up gendered reading of the history of philosophy. And, in so far as essentialism is equated with metaphysics, these arguments against essentialism fail to establish that the rejection of metaphysical thinking is a necessary feature of feminist theorizing.

Notes

- 1 The position that postmodernism's rejection of traditional metaphysics and epistemology is (with a few modifications) the appropriate perspective for feminist theorizing is articulated in the introduction and many of the essays of Linda J. Nicholson, ed., *Feminism/Postmodernism* (New York: Routledge, 1990). A genuine rarity in feminist philosophy is Ann Garry and Marilyn Pearsall, eds, *Women, Knowledge, and Reality: Explorations in Feminist Philosophy* (Boston: Unwin Hyman, 1989), which contains three essays under the heading “Feminist Metaphysics.”
- 2 The first part of this paper is excerpted from “Feminist Metaphysics,” *A Mind of One's Own*, ed. Louise Antony and Charlotte Witt (Boulder, Col.: Westview Press, 1991). The feminist contribution to metaphysics is discussed in the second part of that paper.
- 3 Nancy Fraser and Linda Nicholson, “Social Criticism without Philosophy: An Encounter between Feminism and Postmodernism,” in *The Institution of Philosophy*, ed. Avner Cohen and Marcelo Dascal (La Salle, Ill.: Open Court, 1989), p. 285.
- 4 Richard Rorty, *Objectivity, Relativism and Truth* (Cambridge: Cambridge University Press, 1991), p. 7.

- 5 Jean-François Lyotard, "The Postmodern Condition," *After Philosophy*, ed. Kenneth Baynes, James Bohman, and Thomas McCarthy (Cambridge, Mass.: MIT Press, 1987), p. 73.
- 6 *Ibid.*, p. 89.
- 7 Rorty, *Objectivity, Relativity, and Truth*, p. 13.
- 8 For a useful discussion of this point see Fraser and Nicholson, "Social Criticism without Philosophy," pp. 289–90.
- 9 Richard Rorty, "Feminism and Pragmatism," *Michigan Quarterly Review*, 1, 2 (Spring 1991).
- 10 *Ibid.*, p. 237.
- 11 *Ibid.*, p. 250.
- 12 Rorty argues in "Postmodernist Bourgeois Liberalism," in *Objectivity, Relativity, and Truth*, that postmodernism is not relativistic. "To accuse postmodernism of relativism is to try to put a metanarrative in the postmodernist's mouth. One will do this if one identifies 'holding a philosophical position' with having a metanarrative available" (p. 202). Rorty makes this comment in response to the charge that postmodernism is relativism and that relativism is self-refuting. In this essay, Rorty labels himself a "postmodernist" and defends his view from the charge of relativism (and self-refutation). But it is unclear that this defense will work for Lyotard. For Lyotard – in so far as he holds that all languages are necessarily local (an apparently absolute statement) – does embrace relativism and is open to the charge of self-refutation. For a criticism of Lyotard's thought as inadequate for progressive political change, see Seyla Benhabib, "Epistemologies of Postmodernism: A Rejoinder to Jean-François Lyotard," in Nicholson, ed., *Feminism/Postmodernism*, pp. 107–30.
- 13 Genevieve Lloyd, *The Man of Reason* (Minneapolis: University of Minnesota Press, 1984), p. 104. For a survey of the different ways in which feminists have described reason as male, see Karen J. Warren, "Male-Gender Bias and Western Conceptions of Reason and Rationality," *APA Newsletter on Feminism and Philosophy*, 88, 2 (March 1989).
- 14 Warren, "Male-Gender Bias and Western Conceptions of Reason and Rationality," p. 52.
- 15 The arguments that follow are excerpted from my "Anti-Essentialism in Feminist Theory," *Philosophical Topics*, vol. 23, no. 2, which contains a more complete treatment of these issues.
- 16 In this paper I use biology and nature interchangeably, even though there are differences between the two. For a useful discussion of this distinction, as well as other terms associated with the essentialism/anti-essentialism debate, see Elizabeth Grosz, "Sexual Difference and the Problem of Essentialism," in *The Essential Difference*, ed. Naomi Schor and Elizabeth Weed (Bloomington: Indiana University Press, 1994).
- 17 The list runs from Aristotle to Freud. Feminist historians of philosophy cite Aristotle's attempt to provide a biological basis for the difference between men and women in his theory of *pneuma*. See Nancy Tuana's "Aristotle and the Politics of Reproduction," in *Engendering Origins*, ed. Bat-Ami Bar On (Albany, N.Y.: SUNY Press, 1994). Freud seems to fall in the same camp with his emphasis on the central role of genitalia in establishing gender identity. Today scientists are busy trying to find differences in the brain to explain certain differences between men and women.
- 18 Both Butler and Wittig are anti-essentialist feminist theorists, but Wittig posits an authentic form of sexuality underlying the distortions of social construction. See

Butler's discussion of Wittig in *Gender Trouble* (New York: Routledge, 1990), pp. 16–22.

- 19 A broken Coke machine has that function, even though it is incapable of performing it.
- 20 *Inessential Woman: Problems of Exclusion in Feminist Thought* (Boston: Beacon Press, 1988), p. 113; see also pp. 134, 136, 172.
- 21 Two examples of recent feminist writing on diverse subjects in which the question of generality is a central concern are Susan Bordo, "Material Girl: The Effacements of Postmodern Culture," in *The Female Body*, ed. Laurence Goldstein (Ann Arbor: University of Michigan Press, 1991), and Marilyn Frye, "The Possibility of Feminist Theory," in her *Willful Virgin* (Freedom, Cal.: Crossing Press, 1992). I suspect that both of these writers consider themselves to be of the anti-essentialist persuasion. Yet they both think that some form of generalization about women is fundamental to the political projects of feminism.

INDEX

- Abbott, Edward A., 120
abstract entities, 8, 45–7,
 49–51, 385, 391, 394
Achilles paradox, 120–3,
 128–30, 132, 133, 134–5,
 139, 142–4
act of will, 361, 362
agent causation, 226, 265,
 359–62, 377–80
alphabet of being, 43
amoeba, 183
analysis, 468
 failure to grasp an, 344, 347
angels, 371–2
Anscombe, G.E.M., 22, 201
Anselm, St., 79, 80, 364, 413,
 416, 443–50, 452
anthropic principle, 419–20
anti-realism, *see* realism and
 anti-realism
appearances, sensible, 267–71,
 281–9
adverbial theory of, 261, 273,
 284–9
and mistaken judgment, 268
as parts of the brain, 288
as processes, not substances,
 284–5, 288–9
justificatory role of, 286
multiple relation theory of,
 268–9, 271
object theory of, 268–71
of straight stick in water, 270
parts of, 286–7
sensum theory of, 270–1,
 273–4
 see also consciousness; reality
 and appearance;
 secondary qualities;
 sensa; sense data; visual
 field
Aquinas, St. Thomas, 53, 105,
 181, 293, 331, 357, 417,
 431
Archimedes' point, 388, 394,
 457, 472, 477
Aristotle, 1, 25, 51, 120–1,
 123, 133, 245, 252, 282,
 283, 318–19, 329, 356,
 359, 360, 487
Armstrong, D.M., 261, 263
arrow paradox, 131, 132,
 137–8, 145
atheism, 451
atoms, 128, 146
 logical, 393
 of space, 132
 of time, 132
becoming, 79, 88, 93, 96
being, 415–18
 and non-being, 416
Benardete, José 19
Bergson, Henri, 136, 145
Berkeley, George, 47, 107, 135,
 275
Big Bang, 164, 419–21
Bigelow, John, 216
Black, Max, 65, 139, 141–3
Boccaccio, 156
Bohm, David, 1
Born, Max, 255
boundaries, 117, 295
 see also lines; points; surfaces
Bowne, Borden Parker, 292
Boyle, Robert, 275
Bradley, F.H., 288
brain,
 and consciousness, 293, 333
 artificial, 341
 divided, 296, 310–11, 315,
 322–6
 transplantation of, 294, 308,
 316–17, 319–20, 325
brain zap, 307
Bridgman, P.W., 144, 145
Broad, C.D., 18, 19, 242, 256,
 261, 373, 383
Browning, Robert, 192
Buchan, John, 196
Buddha, 312, 314–15
bundle theory, 11, 17, 53–8,
 65, 311–17
Butler, Joseph, 173, 178, 181,
 292–3, 299, 301, 302, 303
calculus of individuals, 45, 393
calculus, infinitesimal, 107,
 132, 135–7, 139, 144
Calvinism, 172–3
'can', 166–8, 343, 347–51,
 358, 367–9
Carnap, Rudolf, 5–6, 389, 394,
 395–8, 399–402, 457
Carroll, Lewis, 82
Case, Thomas, 288
cat,
 grin left behind by Cheshire
 cat, as trope, 51
 skinning a philosophical, 214
ways to kill a, 233
categories,
 metaphysical, 4–5
 ontological, 43
Catholicism, 172–3
Cauchy, Augustin Louis, 132,
 133, 136
causal loops, 164, 432; *see also*
 causation, backward
causation, 21–2, 60–3, 221–58
analysis of, 230–40, 377–8,
 427
and common sense, 230–5
and counterfactual
 dependence, 163, 308
and derivativeness, 247–8
and events, 351–2, 359–62,
 377, 422, 427
and explanation, 379
and necessary and sufficient
 conditions, 237, 239,
 246–7
and necessity, 222–5, 227–9,
 244–58, 352
and nomic subsumption, 22
and ordinary language, 249
and presentism, 215
and science, 227, 243

- and standing conditions, 233, 237, 250
 as *a priori* notion, 246
 as ambiguous notion, 237
 as relation between tropes, 48
 backward, 163–4
 concept of, as derived from willing, 226, 228–9
 definition of, 222, 361–2
 immanent, 238, 359–62
 indefinability of, 242
 non-necessitating, 255
 of past, 369
 singular, 258
 transeunt, 238, 264, 359–62
 unobservability of, 247–8
 cause and effect, 434–6, 438
 as logically connected, 245
 constant conjunction of, 22, 223–5, 246
 contiguity of, 221, 222–5, 227–9, 236, 241–3
 temporal order of, 222–5
 uniqueness of, 228–9, 232, 241
- Cavell, Stanley, 388–9, 390
 Chalmers, David, 261–2, 263
 change, 85–8, 171–3, 186, 267, 298–9, 371–2
 analysis of 24–5, 95, 96, 160, 194
 as confined to minds, 195–6
 Cambridge, 160
 continuous, 227
 impossible without A-series, 68
 of events, 164–6
 of intrinsic properties, 21, 110, 205–8, 269–70
 of parts, 21, 175, 177, 180, 181, 189, 291–2, 299
 of past, 202, 252
 of relational properties, 25, 68, 110, 205
 of sensible appearances, 267
 of the past, 164–6, 169
 unreality of, 138
 characteristics, 25–39, 57; *see also* properties; qualities
 Chisholm, Roderick M., 21, 22, 216, 261, 262, 263, 264, 265, 376, 379, 435
 Chomsky, Noam, 373
 Clarke, Samuel, 413, 431, 437
 class, 32–3, 45, 201; *see also* sets
 cognition, 24–5, 27
 color, 60
 and science, 270–1, 276
 as intrinsic property, 272, 279
 resemblances among, 279,
see also secondary qualities; visual field
- compatibilism, 256–7, 264, 343–55, 369, 373
 compresence, 43, 54–8
 concepts, 27, 30, 38, 39, 187
 definable and indefinable, 450
 temporal, anthropocentricity of, 94, 101
 conceptual scheme, 383, 387, 391–8, 403–10
 of common sense, 393
 scientific-philosophical, 393
 conceptualism, 27, 31, 34, 38
 concrete entities, 45–7, 49–51, 391, 394
 concurrence, 43, 44, 45
 consciousness, 333, 336–7, 341
 hard and easy problems of, 334–6
 constant conjunction, 22
 convention, 180, 183, 314
 and fact, 391–2
 Copleston, Frederick, 433, 436–8
 cosmological argument, 413, 431–9
 counterfactuals, 168–9, 308
 Crick, Francis, 335–6
- Davidson, Donald, 258, 396
 de Beauvoir, Simone, 475, 486
 dearths, 210, 213–14
 deconstruction, 457, 469–71
 Dedekind, Richard, 132
 Democritus, 281–2
 Dennett, Daniel C., 336
 Derrida, Jacques, 389
 Descartes, René, 51, 122, 196, 249, 417, 444, 445, 449, 450
 descriptions, 189, 200
 determinables and determinates, 37–9
 determinism, 100–1, 196, 250–8, 264, 343, 356–9, 366, 380
 and determinateness, 251
 definition of, 369
 theological, 357–8
 dichotomy paradox, 123, 130–1, 132, 134–5, 139, 142
 Dick, Philip K., 218
 dispositions, 155–7, 234–5, 277, 393
 indeterministic, 375
 to act, 380
see also powers
 divine foreknowledge, 367
 divine providence, 357–8
 double aspect theory, 291
see also dualism, mental state
 dualism, Cartesian, 196, 327
- mental state, 262–3, 291, 296–7, 333–41; and interactionism, 327, 335–40; and Jackson's 'knowledge argument', 335
 substance, 195–6, 262–3, 295, 296
 Thomistic, 331
 Dummett, Michael, 202
- Earman, John, 113
 Eddington, Sir Arthur, 353
 Eddy, Mary Baker, 194
 Edwards, Jonathan, 184, 357, 362
 Einstein, Albert, 193, 253
 Einsteinian local time, 57
 enantiomorphism, 19, 108, 110
 energy, 418
 Enlightenment, 469–71, 477
entia non grata, 210–11, 396
entia nonsuccessiva, 292
entia per se, 295
entia successiva, 291–2
 essence, individual, 211
 essentialism, 472, 475, 480–1
 gender, 485–9
 eternity, opposed to endless duration, 448
 ethical statements, as metaphysical, 465
 expressive function of, 464
 non-verifiability of, 463–5
 Eucharist, 172–3
 Euclid, 117
 events, 53, 55, 69, 77, 83, 85–8, 104, 107, 161, 164–6, 227
 as changes, 232
 as logical constructions, 199–200
 as tropes, 48
 future, 88–93, 100, 212
 momentary, 187
 nonactual, 214
 past, 212
 recurrence of, 236
 evil, problem of, 420, 422, 423, 428
 excluded middle, law of, 91
 existence, 7, 18, 100, 413, 415–39
 and non-existence, 415–18
 and perfection, 416–17
 and reduction, 397, 399
 as a perfection, 444–5
 as predicate, 79, 445
 as relative, 383
 as relative to conceptual scheme, 394–8, 403–10
 as relative to point of view, 107, 409–10
 cessation of, 88–9

INDEX

- existence (*cont'd*)
coming into, 87–8, 416, 435, 437–8, 448–9
contingent, 448
criteria of, 400, 402–10
degrees of, 416, 417–18, 441–2, 444–5
eternal, 416, 437–8, 448
fictitious, 80–1
in the understanding, 441–4
intentional, 273–4
necessary, 416, 422, 431, 434, 448–9; as a perfection, 446–7, 449
sum total of, 92–3, 210
see also explanations of existence
- existential quantifier, 12; *see also* quantification
- existentialism, 373
- explanations,
causal, 419, 421–2, 427–8, 433–9
intentional, 429
of existence, 415–38
and axiarchic view, 422–6, 428
- facts, 48–9, 89–93
as truthmakers, 91
brute, 413, 424–7, 428
negative, 90
- fatalism, 100, 105, 167–8, 196–7, 353–4
- Fawkes, Guy, 322
- feminism, 457
- feminist theory, 475–8, 480–9
- Feynman, Richard, 255, 258
- fictional characters, 210; *see also* existence, fictitious
- Flatland, 110, 111, 116, 119
- Flax, Jane, 5, 457
- Foucault, Michel, 389
- foundationalism, 389
- Fraser, Nancy, 481
- free will, 105, 196, 264, 343–80
analysis of, 343–55
and control, 375, 378–9
and deliberation, 371, 373
and indeterminism, 256
as contradictory notion, 373, 376
- freedom, as negative concept, 365–70
- Frege, Gottlob, 122, 397
- Friedberg, Richard, 143
- Frye, Marilyn, 483
- Gardner, Martin, 113, 114, 115
- Gassendi, Pierre, 445, 449
- Geach, Peter, 19–20, 21, 216, 263
- gender relations, 476–7
- as social construction, 485–9
- general terms, 25, 34–5, 38, 39, 189, 190, 385–7
- geometry, 109
Euclidean, 202, 394
non-Euclidean, 202
of four-dimensional spaces, 117
- Gilson, Etienne, 388, 457
- Ginet, Carl, 369
- Glaucon, 156
- God, 51, 150–1, 153, 164, 196, 357, 362, 379, 417–18, 419, 420, 421, 423, 431–52
and time, 105, 416
as cause of universe, 428–9
as self-caused, 426
as simplest possible substance, 427
as unlimited, 447
- concept of, as self-contradictory, 450
- existence of, as brute fact, 413
- nature of, 372
- necessary existence of, 413, 431
- Goldbach's conjecture, 157
- Goodman, Nelson, 154
- grand unified field theory, 428
- 'grue,' 154–5
- Grünbaum, Adolf, 143, 157
- Guanilo, 444
- guilt, 451
- Gyges, ring of, 156
- hallucination, 272–5
- Hameroff, Stuart R., 336
- handedness, left- and right-, 19, 83, 108–11, 114, 115
- Hart, Johnny, 110, 111
- headless woman illusion, 278, 279, 280–1
- Hegel, G.W.F., 417
- Heracleitus, 24, 186, 189
- Hinckfuss, Ian, 19
- Hintikka, Jaakko, 202
- history, 472–3
- Hobart, R.E. (Dickinson S. Miller), 264
- Hobbes, Thomas, 174, 203, 245, 363, 367
- Holbach, Baron Paul von, 366, 373
- holes, 7–8, 49
- Horwich, Paul, 18
- Hume, David, 21–2, 47, 154, 179, 187, 245–6, 248–9, 250, 348, 362, 367, 431, 433, 434–6, 468
- Humpty Dumpty, 111, 189
- hypersphere, 119
- ideas, 47
and impressions, 223–4
- identity of indiscernibles, 28
- identity, 298–9
and constitution, 176, 178, 402–10
as a relation, 188–9
- numerical, 42, 59, 298–9, 390–1; and diversity, 52–8
- of portion of stuff, 299
- over time, 94–5, 186–90, 291–5, 386–7; and causal continuity, 22; and continuity of change, 189, 235; and spatio-temporal continuity, 235, 300; Aristotelian theory of, 318–19, 329–31; criteria of, 234–6, 326, 400, 402–10; perdurance and endurance, 204–6
- qualitive, 42, 298–9
- imagination, 34
- immanent causation, 22, 264, 265, 309
- immaterial substances, 371, 427
- incompatibilism, 369–70, 373, 374, 376
- incongruent counterparts, 109, 111–14, 115
- indeterminism, 63, 250–8, 264, 343–7, 350–1, 356, 359, 361, 366, 370–80
as incompatible with freedom, 346–7
compatible with four-dimensionalism, 100–1
- indexicals, 409
- indiscernibility of identicals, 298; *see also* Leibniz's law
- individuals,
dependent and independent, 431–4, 447
nonactual, 211, 214
- induction, 188, 427
- infinite regress, 73, 75, 77, 97, 402, 416
of agent-causings, 378
- infinite series, 433
summation of, 121–2, 132–5, 139, 141
- infinitesimals, 228; *see also* calculus, infinitesimal
- infinity,
actual, 125
and collections, 124–9
countable and uncountable, 152
rival conceptions of, 148–54
- infinity machines, 125–9, 139–44, 157
- information, 339–40

- inherence, 43
 instantiation, 31, 37
 instants, 53, 57, 132, 136, 145,
 208
 introspection, 155; *see also*
 consciousness; mental
 states, privileged access to
- Jackson, Frank, 335
 James, William, 413, 439
 Jesus Christ, 172–3
 Johnson, Samuel, 373
 Johnson, W.E., 92
 judgments,
 about the future, 89–93
 types of, 88, 92
- Kant, Immanuel, 108–10,
 111–15, 246, 256, 363,
 364, 367, 388, 394, 431,
 445
- Keynes, J.M., 43
 Kierkegaard, Søren, 451–2
 Koch, Christof, 335–6
 Kotarbiński, Tadeusz, 200
 Kripke, Saul, 19, 153, 154,
 157, 202, 390–1
- language, 470, 474, 482
 and metaphysics, 54
 ordinary, 99; and formal
 logic, 193; misleading
 features of, 94, 96, 171,
 312, 355
 tensed and tenseless, 78,
 94–101, 102–3, 104–7,
 193, 198, 208–9
 translation, 385–8, 396
- laws of nature, 22, 114, 160,
 227–9, 257–8, 337, 366,
 374, 419, 420, 421
 and determinism, 353–4
 and events, 242
 and free will, 256–7
 and human action, 363
 and natural kinds, 238–40,
 250
 bridging, 338
 elegance of, 424
 indeterministic, 370–3
 invariability of, 223–5,
 228–9, 230–1, 241,
 244–7, 258
 psycho-physical, 337–40,
 353–4
 reducibility of, to powers,
 428
 simplicity of, 424
- Leibniz's law, 298–9
 Leibniz, G.W., 28, 43, 53, 80,
 109, 111–12, 292, 298,
 363, 394, 422, 431, 450
- Lemmon, E.J., 202
- Leslie, John, 422
 Lesniewski, Stanislaw, 393
 Levine, Joseph, 337
 Lewis, C.I., 48–9
 Lewis, David, 12, 19, 21, 22,
 206–16, 263, 390–1
 Lewis, Stephanie, 12
 lines, 295
 Lloyd, Genevieve, 483
 Locke, John, 27, 47, 58, 176,
 275, 276, 294, 301, 302,
 303, 320–1, 328, 362, 434
- logic, 92, 193–4, 209, 467
 and contemporary physics,
 107, 194
see also tense logic
- logical constructions or fictions,
 57, 178, 191, 200
- logical positivism, 6, 457, 463;
see also verificationism
- logical types, 46
 Lorentz, H.A., 97
 Lucretius, 284
Luz bone, 292
 Lyotard, Jean-François, 480,
 481–2, 484
- MacKinnon, Catharine, 483
 Malcolm, Norman, 413
 Markosian, Ned, 218
 materialism, 7–8
 mental state, 262–3, 287–9,
 296–7, 333–41
 substance, 262–3, 291–5
- mathematics,
 arithmetic, 152–7
 pure and applied, 141, 144
- matter, 171, 293, 299, 390–1,
 418
 and form, 318–19, 330–1,
 402–8
 as extended, 330
 infinite divisibility of, 330
- Maxwell, Clerk, 255
 McGinn, Colin, 373, 376
 McKay, Donald, 310, 311
 McTaggart's paradox, 18, 215
 McTaggart, J. McT. E., 18,
 74–9, 192, 194, 197, 200
- Meinong, Alexius, 80, 88
 Melden, A.I., 360
 Mellor, D.H., 18, 209
 memory,
 as causal notion, 304–6
 factual and event, 307
- mental images, 272–5; *see also*
 imagination
- mental states, 262–3, 271,
 291–2
 as causes of behavior, 264
 privileged access to, 297–8
see also double aspect theory;
 dualism, mental state;
 materialism, mental state
- mercy, 451
 mereological essentialism, 21,
 171–2, 292
 mereology, 393–8, 399–400
 Merricks, Trenton, 218
 metaphor, 26
 metaphysical statements
 as meaningless, 5–7, 390–1,
 468–9
 as neither true nor false, 466
 expressive function of, 465–6
 non-verifiability of, 461–3,
 465–6
- metaphysics,
 as fictive activity, 475
 as illness, 416
 as the thoroughly empirical
 science, 40
 death of, 388–9, 457, 473
 deceptive character of, 466
 definition of, 1–5
 impossibility of, 5–7
 Mill, J.S., 43, 246, 367
 mind,
 as trope, 47
 causal theory of, 261, 272–3,
 276, 280
 dispositional theory of, 155
 materialist theory of, 298
see also dualism; immaterial
 substance; materialism;
 self; soul
- Minkowski, H., 97
 mirror imagery, 108
 mixed modes, 48
 modality, 202
 actualism with respect to, 211
 and counterpart theory, 64,
 390
 and God, 112
 and logical impossibilities,
 421
 and logical necessity, 425,
 446
 and logical possibility, 140–1,
 152, 422
 and necessary truths, 421
 and necessity of past and
 present, 252
 and possible worlds, 80–1,
 420–2
 and possibility, 191
 and predicates, 390
- modally challenged
 philosophers, 59, 65
- Moebius, A.F., 109, 115
 monad, 43
 monism, 423
 Moore, G.E., 358
 motion,
 at-at theory of, 137–8, 228
 continuity of, 127, 138,
 144–7
 instantaneous, 136–8, 228

INDEX

- motion (*cont'd*)
paradoxes of, 129–47
motives,
as causes of action, 345–7,
349–55, 357, 363–4,
375
as inclining but not
necessitating, 363–4,
380
mysterianism, 334, 373, 376
- names, 52, 189, 200
natural kinds, 23, 31, 243, 249,
372
and causal characteristics,
234–5, 238–40
necessary being, 413
necessity, *see* modality
Necker cube, 119
negative dimensions, 117
Nerlich, Graham, 113
New Realists, 282
Newton, Sir Isaac, 53, 109,
111, 112, 250, 253, 428
Newtonian mechanics, 254–5,
257–8
Nicholson, Linda, 481
nominalism, 9–11, 34, 45
null object, 393
numbers,
as timeless objects, 92, 94,
95, 98, 99
reducibility of, to sets, 397–8
- objects, 393, 394, 398
as composed of events, 50,
85, 187, 197
coincident, 403–6, 408
extended and unextended,
295, 330
- O'Connor, Timothy, 22, 264–5
Occam's razor, 288, 397
occasionalism, 196
omnipotence, necessary, 449
omniscience, necessary, 449
onomatoid, 200, 201
ontological argument, 77, 79,
413, 416–17, 441–52
ontological commitment, 9, 11,
19, 198, 210–16, 262,
396–7, 407
ontological parasites, 295
ontology, 7, 40
ostensive definition, 55, 114,
187–8
- panpsychism, 340
paraphrase, 8, 10, 19, 200,
210–16, 262, 395–7, 399
Parfit, Derek, 263, 324–6, 413,
427–9
parity, fall of, 114
Parmenides, 194, 416
particles, 56, 57, 83, 393, 425
- and dual aspect theory, 293
as discontinuous and
identityless, 50
as strings of events, 58
particular, 52–3
nearly-bare, 212
parts,
abstract, 41–3
concrete, 41–3, 50
see also change of parts; sum
Peacocke, Christopher, 156
Peano postulates, 153–4
Peirce, C.S., 48, 122, 132, 182
Penrose, Roger, 336
perception, 281–9
analysis of, 268–9
and conceptual scheme, 391
of parts of physical object,
286
- perfections,
as simple properties, 450
persistence,
intact and nonintact, 175,
177, 180, 181
- person-stages, 160, 164, 190–1,
201, 300, 303, 309
- personal identity, 174, 181–3,
186, 263, 296, 311–17,
331
and amnesia, 328
and bodily identity, 318–20
and causal continuity, 163,
308–9, 313–17; *see also*
immanent causation
and disembodiment, 296,
327–31
and fission, 183, 305–6,
323–7
and memory, 182–3, 295,
297–8, 301–9, 320–7,
328
and psychological continuity,
162–3, 308–9, 313–17,
320–7
as all-or-nothing, 182–3,
313–17, 324–7
as unanalysable, 301
'best candidate' theories of,
323–4
criteria of, 295, 297–8
ego theory, 311–16
psychological theory of, 263
the simple view of, 326, 327
- personalism, 292–3
- persons,
as composites of soul and
body, 331
as logical constructions out of
events, 200, 312
no-self view of, 312
- philosophers, profession of, 26
- physical impossibility, 128
- physical objects, 262, 263,
291–5, 385–8, 390–1
- concept of, 267
criteria of identity for, 297,
318–19
living, 171
- physical states, 262–3
- pig, ways to roast a, 233
- places, 28, 59, 61; *see also* space-time, regions of
- Plato, 25, 156, 293, 431
- Platonism, 9–11, 25, 38, 39,
59, 422
- plurality of causes, 233–4
- Poincaré, Henri, 118
- point-instant, 57
- points, 56, 57, 117, 118–19,
123, 132, 136, 144–5, 295
reducibility of, to sets of
regions, 391, 394
- polarity, 417
- Port-Royal Logic, 21
- possibility, *see* modality
- postmodernism, 389, 457, 469,
478, 481–9
- powers, 22, 223, 226–7, 377,
427, 428
and laws of nature, 348–9
to act, 348–50, 362
- pragmatism, 482–3
- predication, 46, 197
- present,
as moving spotlight, 84
specious, 84, 86, 89, 138
- presentism, 18, 21, 62–3, 80–2,
205–6, 209, 355
- Price, H.H., 17, 18
- primary qualities, 275–80
- prime mover, 361, 362–3, 379
- principle of causality, 253
- principle of sufficient reason,
432–9
as necessary truth, 435–6
as presupposition of rational
inquiry, 436–9
- Prior, A.N., 18–19, 21, 212,
216, 265
- processes, 48, 186–7, 204–5,
227
- properties, 8–11, 17–18,
58–65, 103
accidental, 318
essential, 318–19
fundamental, and laws of
nature, 337–8
incompatible, 194, 207,
212–13, 267
- instances of, 17, 40, 59, 300;
see also tropes
- intrinsic, 60, 112–14, 115,
119, 191, 205–8,
212–13, 261, 279
- monadic, 207–8
- of sensa, 271
- relational, 61, 261
- see also* qualities; universals

- propositions, 49
as changing, 71, 208
Protagoras, 282
psychologism, 467
psychology, 466–7
Putnam, Hilary, 5, 383, 399–
402, 457
Pythagorean theorem, 147

qualia, 44, 341
qualities, 24–5, 34, 53–8, 191,
221–3
positive and negative, 450
simple, 450
quantification, 193, 386–7
quantum theory, 1–2, 229,
253, 254, 256, 425, 428
Quine, W.V.O., 8–9, 11, 19,
21, 193–4, 197, 198, 383,
390, 391, 395–7, 401, 457
quus, quaddition, 153–7

randomness, 256, 424–5
realism, 398
and anti-realism, 6, 389–98,
457; with respect to
competing ontological
doctrines, 39
and cookie cutter metaphor,
393–4, 409
direct, 276
in epistemology, 34
internal, 392–3, 398
metaphysical, 395–8, 402
physical, 288
reality, 80, 390, 475
and appearance, 1–2, 6
as ambiguous term, 271
as mind-independent, 470
as socially constructed, 477
degrees of, 416, 417–18, 442
explosion of, 403–6, 408
God's eye view of, 383
sum total of, 433
two concepts of, 462–3
unconceptualized, 392
recurrence, 23–9, 54, 56
Reichenbach, Hans, 391
Reid, Thomas, 22, 265, 301,
302, 303, 312, 321, 322,
357, 362, 363
reincarnation, 296, 327
relations, 24–5, 34, 205, 207,
221–3
external, 43, 112
internal, 43, 51, 112
of appearing, 268–9
relativism, 483
concerning secondary
qualities, 282
cultural, 394, 488, 489
linguistic, 383, 391–2,
399–402
of conceptual scheme, 393–8,
- 403–10
ontological, 383, 386–8
relativity, 19, 97–8, 107, 188,
193, 199, 215
resemblances, 37, 41, 45
as internal, 279
exact and inexact, 27–31,
37–9, 44
first-order and second-order,
35
philosophy of ultimate, 17,
27, 31–9
resurrection, 296
revisionism, philosophical, 389
rhetoric, 471
Rorty, Richard, 389–90, 392,
480, 481, 483, 484
Rowe, William, 413
Russell, Bertrand, 17, 22, 43,
70, 137, 200–1, 232,
241–3, 247, 285, 300,
397, 433, 436–7
Ryle, Gilbert, 199

Sagan, Carl, 2–3
Salmon, Wesley, 19
Santayana, George, 43
scholasticism, 47, 51, 52
Schopenhauer, Arthur, 415
Schrödinger, Erwin, 253
Scotus, Duns, 105, 432
secondary qualities, 272,
275–80
and epiphenomenalism, 276
and science, 280
as lacking 'grain,' 278
as subjective, 24
dispositional and sensible use
of terms describing the,
282–3, 285, 288
dispositional theory of, 277
Gestalt theory of, 277–8
reduction of, to primary,
277–80
self, 312, 470
and character, 344–7, 352
bundle theory of, 311–17
self-referential incoherence, 6–7
Sellars, Wilfrid, 98, 100, 278
semantic ascent, 401–2
Seneca, 171
sensa, 261, 270–1; *see also* sense
data
sense data, 261, 273, 276, 281,
283–4, 288
as tropes, 47
indeterminate nature of,
274–5
see also sensa
sensible species, 47
sets, 12–13, 32–3, 44, 45,
136–7, 139, 397–8; *see also*
classes
Shannon, Claude E., 339

ship of Theseus, 174, 180, 189
Shoemaker, Sydney, 22, 263
Sidgwick, Henry, 349
similarity; *see* resemblances
simple substances, 371; *see also*
monad
singular terms, 385–7
skepticism, 119
Cartesian, 249
Smart, J.J.C., 19, 102, 106,
197, 209, 288–9
Smith, Quentin, 212, 213
Socrates, 31
solipsism, 196
Sosa, Ernest, 383
soul, 186, 330–1
as trope, 47
see also immaterial substance
space, 83
absolute, 19, 109–11,
112–14
atoms, 147
continuity of, 127, 138,
144–7
directions in, 193
discrete, 144–7
Euclidean, 147
fourth dimension of, 19,
109–11, 114, 115–19
infinite divisibility of, 128,
141, 228
quantized, 145–7
relational theory of, 28, 53,
109–11, 112–14
substantial theory of, 53, 61
space-time, 53, 57
as four-dimensional, 63
regions of, 390–1
speaking strictly and speaking
loosely, 3, 7, 8, 171, 174,
177–8, 180–2, 188, 294,
301
Spelman, Elizabeth, 487–9
Spencer, Herbert, 416
Spinoza, 101, 245, 447
stadium paradox, 131–2, 145–6
state of affairs, 48–9
Stebbing, L. Susan, 22, 264
Strong, C.A., 291
stuff, 249, 299, 402
immaterial, 329–30
Suarez, Francisco, 361
subject,
in propositions, 171–3, 200
logical and grammatical, 90,
93, 197, 198
substance, 427
as peg on which to hang
predicates, 54
immaterial, 47, 329–30
intellectual, 331
see also bundle theory;
concrete entities;
monad; physical objects

INDEX

- substratum, 58, 65
sum, 44, 45
 mereological, 393–8,
 399–400
supervenience, 191, 263, 403–5
surfaces, 49, 295; *see also*
 boundaries
Swinburne, Richard, 263, 413

Taylor, Richard, 436–8
teletransportation, 314–15
temporal copula, 75, 79; *see also*
 language, tensed and.
 tenseless
temporal parts, 21, 50, 83, 89,
 95, 160, 164, 184,
 186–91, 192–5, 197–8,
 201, 204–8, 215–16, 263,
 300, 385
and substance dualism, 190
instantaneous, 186, 188,
 190, 192, 195
see also person-stages;
 temporal phases
temporal phases, of an object's
 history, 24; *see also* person-
 stages; temporal parts
temporal succession, 22
tense logic, 19, 82, 104–5, 107,
 199–202, 209
theory of everything, 337–8
Thomson, J.J., 140–3
Thomson lamp, 140–3
tie, characterizing, 92
time,
 absolute, 69, 199
 absolute theory of, 78
 and causation, 222
 and hyper-time, 97
 and spatial analogies, 82–4,
 102, 201–2, 203, 204,
 207, 300
 as a 'growing block,' 18–19,
 87–93
 as A-series, 68–73, 74–9
 as B-series, 68–73, 75–9
 as fourth dimension of
 space-time, 19, 83, 160,
 164, 169, 192–6, 203,
 300
 as involving change, 68, 72
 as two-dimensional, 202
atoms, 147
branching, 169
continuity of, 138–9, 144–7
direction of, 83
discrete, 132, 138–9, 144–7
earlier or later than in, 67–9,
 75–9
higher-order, 87
infinite divisibility of, 133,
 228
moments of, 75–7
passage or flow of, 19, 96–7,
 101–3, 107, 215
past, present, and future,
 18–19, 67–73, 75–9,
 80–2, 94–101, 103,
 104–7, 194, 198, 200,
 211, 213–16
personal and external, 160,
 164–5, 167–9
quantized, 145–7
travel, 159–69, 195–6,
 202–3
two-dimensional, 159–60
unreality of, 67–73, 104, 194
see also language, tensed and.
 tenseless
timelessness, 77
token reflexivity, 95, 99–101,
 103
token, 48
topic-neutrality, 199, 278, 280
trope, 17, 43–52, 59
 pain as, 47
 mixed, 48
 geometrical figure as, 49
truth, 457, 470, 474, 477
analytic, 435
 as correspondence with fact,
 89–91
 as relative to place, 409
eternal, 211–12
fundamental bearers of,
 208–9, 216
synthetic *a priori*, 57, 113,
 435–6
type, 48

unity,
 diachronic, 300, 311–17
 of consciousness, 300,
 311–12, 315–17; *see also*
 brains, divided
 synchronic, 300, 311
universalia ante rem, 25, 26,
 31, 38, 39; *see also*
 Platonism
universalia in rebus, 17, 25–7,
 34, 38–9; *see also*
 universals, immanent
universals, 17–18, 23–39, 40,
 58–65, 204
arguments for the existence
 of, 7–11
first-order and second-order,
 35–6
immanent, 59–65, 331; *see*
 also universalia in rebus
instances of, 46, 53–4, 57,
 210, 331
see also properties; qualities
universe,
 alternative, 214; and many-
 worlds hypothesis,
 419–21, 423, 424, 428
complete inventory of, 271
eternally recurring, 100
fine-tuning of, 419–20
 symmetrical, 58–65
unsensed sensibilia, 285
utilitarianism, 423

vacua, 428
vagueness,
 and borderline cases, 30–1,
 179–80, 182
 linguistic, 189
value, 422
Van Cleve, James, 19
van Inwagen, Peter, 264, 265,
 374, 376
verificationism, 328, 459–69;
 see also logical positivism
Vienna Circle, 6, 463, 468
visual field, 57–8, 248, 339,
 340

Weber, Max, 477
Weierstrass, Karl, 132, 137
Wells, H.G., 195
Weyl, Hermann, 146–147
Weyl tile argument, 146–147
Wheeler, John A., 340
White, Morton, 392
Whitehead, A.N., 43, 97, 121,
 122, 145
Wiggins, David, 305, 316–17,
 320
Williams, Bernard, 322, 325
Williams, D.C., 17, 18, 210
Witt, Charlotte, 457
Wittgenstein, Ludwig, 19, 153,
 155, 189, 199, 388, 390,
 450, 468–9
Wright, Crispin, 157

Zeno, 121–3, 133, 135–6, 138,
 141–2, 144–5, 157, 416
Zimmerman, Dean W., 17, 21