Analytic Philosophy Without Naturalism

Edited by Antonella Corradini, Sergio Galvan and E. Jonathan Lowe

Routledge
Taylor & Francis Group

Analytic Philosophy Without Naturalism

In recent years numerous attempts have been made by analytic philosophers to *naturalize* various different domains of philosophical inquiry. All these attempts have had the common goal of rendering these areas of philosophy amenable to empirical methods, with the intention of securing for them the supposedly objective status and broad intellectual appeal currently associated with such methods.

This volume brings together internationally recognized analytic philosophers, including Alvin Plantinga, Peter van Inwagen and Robert Audi, to question the project of naturalism. The articles investigate what it means to *naturalize* a domain of philosophical inquiry and look at how it applies to the various sub-disciplines of philosophy including epistemology, metaphysics and the philosophy of mind. The issue of whether naturalism is desirable is raised and the contributors take seriously the possibility that excellent analytic philosophy can be undertaken without naturalization. *Analytic Philosophy Without Naturalism* examines interesting and contentious methodological issues in analytic philosophy and examines the connections between philosophy and science.

Antonella Corradini is Professor of Philosophy of the Social Sciences at the Catholic University of Milan, Italy.

Sergio Galvan is Professor of Logic at the Catholic University of Milan, Italy.

E. Jonathan Lowe is Professor of Philosophy, University of Durham, UK.

Routledge Studies in Twentieth-Century Philosophy

1 The Story of Analytic Philosophy
Plot and heroes
Edited by Anat Biletzki and Anat Matar

2 Donald Davidson
Truth, meaning and knowledge
Edited by Urszula M. Zeglén

3 Philosophy and Ordinary Language
The bent and genius of our tongue
Oswald Hanfling

4 The Subject in Question
Sartre's critique of Husserl in *The Transcendence of the Ego*
Stephen Priest

5 Aesthetic Order
A philosophy of order, beauty and art
Ruth Lorland

6 Naturalism
A critical analysis
Edited by William Lane Craig and J.P. Moreland

7 Grammar in Early Twentieth-Century Philosophy
Richard Gaskin

8 Rules, Magic and Instrumental Reason
A critical interpretation of Peter Winch's philosophy of the social sciences
Berel Dov Lerner

9 Gaston Bachelard
Critic of science and the imagination
Cristina Chimisso

10 Hilary Putnam
Pragmatism and realism
Edited by James Conant and Urszula Zeglen

11 Karl Jaspers
Politics and metaphysics
Chris Thornhill

12 From Kant to Davidson
The idea of the transcendental in twentieth-century philosophy
Edited by Jeff Malpas

13 Collingwood and the Metaphysics of Experience
A reinterpretation
Giuseppina D'Oro

14 The Logic of Liberal Rights
A study in the formal analysis of legal discourse
Eric Heinze

15 Real Metaphysics
Edited by Hallvard Lillehammer and Gonzalo Rodriguez-Pereyra

16 Philosophy After Postmodernism
Civilized values and the scope of knowledge
Paul Crowther

17 Phenomenology and Imagination in Husserl and Heidegger
Brian Elliott

18 Laws in Nature
Stephen Mumford

19 Trust and Toleration
Richard H. Dees

20 The Metaphysics of Perception
Wilfrid Sellars, Critical Realism and the nature of experience
Paul Coates

21 Wittgenstein, Austrian Economics, and the Logic of Action
Praxeological investigations
Roderick T. Long

22 Ineffability and Philosophy
André Kukla

23 Cognitive Metaphor and Continental Philosophy
Clive Cazeaux

24 Wittgenstein and Levinas
Ethical and religious thought
Bob Plant

25 Philosophy of Time
Time before times
Roger McClure

26 The Russellian Origins of Analytic Philosophy
Bertrand Russell and the unity of the proposition
Graham Stevens

27 Analytic Philosophy Without Naturalism
Edited by Antonella Corradini, Sergio Galvan and E. Jonathan Lowe

Analytic Philosophy Without Naturalism

Edited by Antonella Corradini,
Sergio Galvan and E. Jonathan Lowe

Routledge
Taylor & Francis Group
LONDON AND NEW YORK

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

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

This edition published in the Taylor & Francis e-Library, 2006.

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

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

Library of Congress Cataloging in Publication Data
A catalog record for this book has been requested

ISBN10: 0-415-34945-1 (hbk)
ISBN10: 0-203-69513-5 (ebk)

ISBN13: 978-0-415-34945-1 (hbk)
ISBN13: 978-0-203-69513-5 (ebk)

Contents

Illustrations

Table

Figures

Contributors

Robert N. Audi is Professor of Philosophy and David. E. Gallo Professor of Business Ethics at the University of Notre Dame, Indiana. He works in ethics, theoretical and applied, and including political philosophy; epistemology, including religious epistemology; and philosophy of mind and action. His most recent books are *The Architecture of Reason* (Oxford, 2001) and *The Good in the Right: A Theory of Intuition and Intrinsic Value* (Princeton, NJ, 2004).

Francesca Castellani teaches Logic and Philosophy of the Social Sciences at the University of Trento, Italy. Her main scientific concerns are regarding action theory, intensional logics, logic and common sense. Her latest book is *Modelli di ragionamento pratico* (*Models of Practical Reasoning*) (Milan, 2000).

Robin Collins is Associate Professor of Philosophy at Messiah College, with research interests in science and religion, philosophy of physics, and philosophy of religion. One of his most recent book chapters is 'The Evidence for Fine-Tuning', in *God and Design: The Teleological Argument and Modern Science* (London, 2003).

Antonella Corradini is Professor of Philosophy of the Social Sciences at the Catholic University of Milan, Italy. Her research interests include the epistemology of psychology, the philosophy of mind and action, the philosophy of biology and the social sciences, and the foundations of ethics. Her latest book is *Modelli epistemologici in psicologia* (*Epistemological Models in Psychology*) (Rome, 2004, co-authored by M. Castiglioni).

Roberta Corvi is Professor of History of Contemporary Philosophy at the Catholic University of Milan, Italy. Her research interests concern contemporary Anglo-Saxon philosophy with particular attention to issues of knowledge, rationality and language. Among her publications is *An Introduction to the Thought of Karl Popper* (London, 1997).

William Lane Craig is Research Professor of Philosophy at the Talbot School of Theology, Biola University, La Mirada, California. His research

interests include philosophy of time and philosophy of religion. Among his publications are *Theism, Atheism, and Big Bang Cosmology* (Oxford, 1993) and *God, Time, and Eternity* (Dordrecht, 2001).

Sergio Galvan is Professor of Logic at the Catholic University of Milan, Italy. He works mainly on Gödel's theorems and on philosophical logic. Among his publications are *Introduzione ai teoremi di incompletezza* (*Introduction to the Incompleteness Theorems*) (Milan, 1992) and 'The principle of deontic reflexivity and the Kantian axiom' (*Logique et analyse*, 2001).

Alessandro Giordani is Research Professor of Logic at the Catholic University of Milan, Italy. His research interests concern in particular the application of modal logic to the analysis of issues in the field of ontology and philosophy of science. His publications include *Teoria della fondazione epistemica* (*A Theory of Epistemic Foundation*) (Milan, 2002).

Christian Kanzian is Professor of Philosophy at the University of Innsbruck, Austria. His main research fields are ontology, analytical philosophy, and the history of philosophy. His latest book is *Ereignisse und andere Partikularien* (*Events and Other Particulars*) (Paderborn, 2001).

Michele Lenoci is Professor of History of Contemporary Philosophy at the Catholic University of Milan, Italy. He has devoted his research to the theory of knowledge, ontology and anthropology, especially the range of phenomenological movement and analytical philosophy. Among his books are *Pensiero, linguaggio, verità* (*Thought, Language, Truth*) (Milan, 1986).

Winfried Löffler is Associate Professor of Philosophy at the University of Innsbruck, Austria. His main interests are the structure of arguments in the context of world-views and the philosophy of Bernard Bolzano. A recent publication is 'Epistemically relevant possible worlds' (*Grazer Philosophische Studien*, 2002).

E. Jonathan Lowe is Professor of Philosophy at the University of Durham, UK, specializing in metaphysics, the philosophy of mind and action, philosophical logic, and early modern philosophy. He is the author of *The Possibility of Metaphysics* (1998) and *Locke* (2005).

Hugh J. McCann is Professor of Philosophy at the Texas A&M University. His areas of interest concern action theory and the philosophy of religion. He is the author of *The Works of Agency* (Cornell, 1998).

Uwe Meixner currently teaches at the University of the Saarland, Germany. His research interests are metaphysics, logic, and the philosophy of mind.

He has published many articles and books, among them *Axiomatic Formal Ontology* (Dordrecht, 1997) and *The Two Sides of Being* (Paderborn, 2004).

Alvin Plantinga is John A. O'Brien Professor of Philosophy at the University of Notre Dame, Indiana. He gives seminars in philosophy of religion, metaphysics and epistemology. Among his many books is *Warrant and Proper Function* (Oxford, 1993).

Josef Quitterer is Associate Professor of Philosophy at the University of Innsbruck, Austria. He has published articles on the philosophy of mind and the philosophy of science. He is the co-author of *Unser Selbst: Personale Identität im Wandel neuronaler Prozeße* (*Our Self: Personal Identity Within the Changes of Neural Processes*) (Paderborn, 2003).

Edmund Runggaldier is Professor of Philosophy at the University of Innsbruck, Austria. His main research field is ontology in the Aristotelian tradition. Among his publications is *Che cosa sono le azioni?* (*What Are Actions?*) (Milan, 2000).

Lucia Urbani Ulivi is Professor of Theoretical Philosophy at the Catholic University of Milan, Italy. She works on metaphysics and ontology, the philosophy of mind, and anthropology. Her latest book is *Introduzione alla filosofia* (*Introduction to Philosophy*) (Lugano, 2005).

Peter van Inwagen is John Cardinal O'Hara Professor of Philosophy at the University of Notre Dame, Indiana. His research interests concern metaphysics and the philosophy of religion. His latest book is *Ontology, Identity, and Modality: Essays in Metaphysics* (Cambridge, 2001).

Franz von Kutschera is Professor Emeritus at the University of Regensburg, Germany. His main research interests are the philosophy of mind and ethics. He is the author of *Jenseits des Materialismus* (*Beyond Materialism*) (Paderborn, 2003).

Preface

In recent years, numerous attempts have been made by analytic philosophers to *naturalize* various different domains of philosophical inquiry. All these attempts have had the common goal of rendering these areas of philosophy amenable to empirical methods, with the intention of securing for them the supposedly objective status and broad intellectual approval currently associated with such methods. What exactly does it mean, though, to 'naturalize' a domain of philosophical inquiry?

What similarities and differences obtain between, for example, a naturalization programme in epistemology and one in ethics or action theory? And to what extent can these attempts at philosophical naturalization be judged to be consistent, coherent and supported by solid empirical evidence? Are there any just grounds to challenge the underlying assumptions of such programmes, by appealing to such notions as those of intentionality, rationality, consciousness, free will, and ethical value?

These topics were the subject matter of a major international conference which took place at the Catholic University of Milan in June 2003. The main speakers at the conference were analytic philosophers from both sides of the Atlantic who are both internationally recognized in their respective fields and noted for their record of critical reflection on the credentials of naturalism. The conference was divided into five sections, each devoted to a different area of philosophical inquiry: epistemology, ontology, the philosophy of mind, the philosophy of religion, and ethics. Revised versions of all the papers delivered by the main speakers, accompanied by the replies made to them by their commentators, are collected in this volume, along with an introductory chapter written by its editors.

Acknowledgements

The editors and contributors are extremely grateful for helpful comments and suggestions provided by two anonymous readers for the publishers, as well as to all those at Routledge who helped to bring this project to fruition so smoothly and efficiently.

Chapter 5 includes figures reprinted from W.L. Craig, *Creation Out of Nothing*, Baker Book House Company, 2005.

Introduction

Antonella Corradini, Sergio Galvan and E. Jonathan Lowe

Naturalism probably constitutes mainstream opinion in contemporary analytic philosophy. Its importance, however, lies not merely in its current popularity among analytic philosophers, but also in the fact that the naturalistic approach addresses itself systematically and compellingly to the entire spectrum of philosophical problems, thus presenting a serious challenge to the methods and claims of other philosophical traditions.

The areas of analytic philosophy in which the naturalistic approach has some role to play are many. However, there are four in which it is particularly significant: ontology, the theory of knowledge, the philosophy of mind, and practical philosophy. Moreover, a rich network of themes underlies these areas, forming a common background to them all. These are themes that generate arguments at the heart of each area and, at the same time, emerge from deep interconnections between problems belonging to different areas. Thus, for example, the philosophy of mind is a domain in which certain central theses of ontology, epistemology and practical philosophy are put to the test. As a consequence, if ontological – and, equally, epistemological – claims are guided by naturalistic principles, the philosophy of action will likewise acquire a naturalistic cast, and with it the ensuing conception of the philosophy of mind in general. This point is connected with the fact that the philosophy of mind is an appropriate domain in which to evaluate not only ontological hypotheses of a general kind, but also those with a specific metaphysical relevance for the philosophy of religion. Suppose, for example, that the philosophy of mind delivers conclusions which rule out an analysis of the relationship between the mental and the physical as being one of mere supervenience, thus excluding a physicalist explanation of the mental in these terms. The implication will be that the mental is *emergent* with respect to the physical. But if emergence is taken seriously – that is, if an emergent feature is regarded as one that is genuinely *novel* with respect to the intrinsic potentialities of the subvenient domain – this will mean that such a feature cannot be explained as arising out of the subvenient domain purely in and of itself. This will leave us with just two possibilities: either the novel feature is absolutely inexplicable – that is, it is a brute fact existing without a foundation of any kind, which

conflicts with the principle of sufficient reason – or else it is explicable only within a framework more extensive than is empirically available. Clearly, accepting the latter will mean opting for a transcendent metaphysics, which could even turn out to be a metaphysics of the Transcendent in the strict sense.

Obviously, the emergent status of the mental with respect to the physical is something that might need to be acknowledged even at the level of the philosophy of action and the ethics of moral responsibility. If we are convinced of our responsibility for (at least some of) our actions and choices, then we ought also to be convinced that we are their authors. This, however, means presupposing that we are free, which in turn presupposes a fundamental autonomy of the mental with respect to the physical. It must thus make sense to explain our actions on the basis of the (abstract and axiologically ideal) reasons that are at the root of our intentions and inform our will, not merely on the basis of the causes that generate our bodily movements. But if one recognizes the legitimacy of a rational explanation of our intentional actions, all the more must one accept the irreducibility of the mental to the physical, in accord with the idea, outlined earlier, of the emergence of the mental with respect to the physical.

In conclusion, the four areas of analytic philosophy mentioned at the outset – ontology, the theory of knowledge, the philosophy of mind and practical philosophy, together with the philosophy of religion – are closely interconnected, to such an extent that naturalism in its various forms is a position that deserves only to be accepted or rejected *en bloc*. One cannot, for example, consistently be a naturalist in ontology but a non-naturalist in the philosophy of action. Naturalism is an approach that either permeates every region of philosophical thought or else none.

Now, the thematic and methodological unity of philosophical thought that we have taken for granted so far is accompanied, in many contemporary expressions of analytic philosophy, by the presumption of naturalism. It is therefore quite natural to ask whether the presumption of naturalism, being so widespread, is simply *intrinsic* to the analytical method: that is, whether the structural unity underlying the various areas of philosophical thought is, perhaps, the reflection of a particular model – the naturalistic one – which informs the entire field. Or are we, on the contrary, dealing here with a mere *de facto* alliance, generated by the current intellectual climate, and hence one that is open to reversal once the thematic and methodological unity in question is put to the service of a different presumption? In the context of this issue, the contributions to the present volume may be viewed on two distinct, but related, levels.

On the one hand, the contributions show how it is possible for one to adopt the methodological principles of analytical investigation without accepting the presumption of naturalism. On the other, they demonstrate how the rejection of naturalism in one area implies a consequential change of paradigm in other areas too, entailing the elaboration – though superficially,

perhaps, without a uniform aim – of an alternative model centred on a vision of human beings and the world as constituting a reality that is not purely natural. From the first perspective, the contributions all stem from the question of whether analytic philosophy is intrinsically fated, in virtue of its method, to issue in naturalistic claims, or whether such claims – when and where they emerge – depend on factors that are tied not to the analytical method itself but rather to ulterior assumptions. The authors consider it important to respond to this question – see in particular Peter van Inwagen's contribution – because it reflects a widespread conviction, especially to be found in philosophical circles rooted in the tradition of continental philosophy. This is the conviction that analytic philosophy is inherently apt to favour naturalistic positions, both for reasons of principle and for historical reasons connecting analytic philosophy tightly to the Vienna Circle and the movements that grew out of it. As Edmund Runggaldier's contribution demonstrates, it is emphatically maintained in many quarters that the analytical method is, in particular, inherently reductionistic and scientistic in character. According to the continental critics of analytic philosophy, indeed, the analytical method aims to gather together the ultimate constituents of reality – be they atoms or tropes – and to construct, by means of a process of reduction to the basic entities, the entire complex image of reality as a whole. Furthermore, it supposedly assigns to the empirical sciences the task of determining what these basic entities are and how the modalities of the reduction should be specified, allowing philosophical reflection no cognitive role at all, in the strict sense. Seen in this light, the scientistic and naturalistic bias of analytic philosophy seems quite evident, so much so that it appears to warrant the critical verdict that continental philosophy so often pronounces upon it.

On the other hand, the theoretical virtues of the analytical method are sufficient to dissuade even the severest critic from abandoning it in order to embrace some of the perspectives to which the continental tradition appeals. Two of these virtues in particular are unanimously emphasized by the contributors. First, the analytical method is important on account of its systematicity – a characteristic deemed to be essential and inescapable for a philosophy set upon achieving a synthetic vision. Second, the analytical method is welcoming in its attitude towards science, in regard to both its methods and its results. It is not plausible, indeed, to engage in philosophical investigation as though science did not exist, avoiding a robust and systematic engagement with the fields of knowledge that the various scientific disciplines make available to us. The repudiation of naturalism, if it is a practicable option, should not be arrived at *a priori*, on the basis of a gratuitous assumption, but only *a posteriori*, as a result of a critical examination of scientific results and the contexts of their interpretation.

In answer to the anticipated question as to the plausibility of a non-naturalistic paradigm, the individual contributors advance, in their own independent ways, a model of analytical philosophy that is open to a non-naturalistic

vision of the world. And this is the second perspective from which the contributions of the different authors can be read. From this point of view, the contributions – in conjunction with the comments of the discussants – can be placed in three broad thematic groupings: the first concerning connections with the philosophy of religion, the second concerning connections with the theory of knowledge, and the third concerning connections at once with ontology, the philosophy of mind and practical philosophy.

In Part III on the philosophy of religion, William Lane Craig and Robin Collins develop two closely connected themes. The former presents a particular version of the Kalam cosmological argument. As is well known, this argument was originally advanced by al-Ghazali in a formulation based on two fundamental principles: (1) it is not possible to have an infinite regress in the temporal series of events. This principle, excluding as it does the possibility that the series of past events has no first member, implies that the universe has a beginning; and (2) that which has a beginning cannot be the source of its own existence, that is, cannot be *causa sui*. Together, the two principles imply that the universe requires an external Cause, whose attributes coincide with those of God the Creator. Craig has also developed such an argument in other contexts, where he has advanced various considerations in favour of the principle of the impossibility of an infinite regress in the temporal series of events, including the *a priori* truth that an infinite series of past events would constitute – impossibly – an example of the actual infinite. Craig's contribution in the present volume is characterized, however, by the fact that the reasons that he takes into consideration in order to exclude an infinite regress are all *a posteriori* reasons of a scientific nature. In fact, Craig surveys a large number of alternatives to the Big Bang cosmological model – according to which the universe has a beginning in time – pointing out the inherent difficulties of each of these. The outcome of this critical review is, thus, an empirical confirmation of the standard model and, hence, of the claim that the universe is not eternal, but had a beginning. At this point, one might have expected Craig to move directly to the conclusion of his argument, with a direct application of the principle that *that which has a beginning to its existence cannot be a cause of itself*. Craig, however, does not regard al-Ghazali's second principle as being unconditionally valid. And this is the second point of interest in Craig's contribution. He, in fact, regards the principle as being valid only in the context of a tensed conception of time. An opportunity to illustrate this point is provided by a critical discussion of the hypothesis advanced by J. Richard Gott and Li-Xin Li, according to which the laws of physics do not exclude the possibility that the universe, even though it had a beginning in time, is self-created. Now, in Craig's view, the self-creation hypothesis makes sense in the context of a tenseless (B-theoretical) conception of time, but not in the context of a tensed (A-theoretical) conception. B-theorists maintain that specifically temporal qualifications such as present, past, and future are subjective illusions of consciousness. Present, past, and future events in time

are all equally existent, differing from each other only in their positions, depending on whether they occur before, after or simultaneously with one another. In contrast, A-theorists contend that only the present exists, so that the process of becoming is a genuine passage from non-existence into being and from being into non-existence. Thus, if one adheres to the tensed conception of time, one cannot consider the universe to be capable of causing itself, since this would require the present world (which is not yet existent at the moment of causation) to be caused by the future world (which does not exist). Only by regarding temporal succession as a B-series is one free to suppose that the universe is self-caused. But, then, indeed, it doesn't even make sense to say that the universe has a beginning in time, because in the B-series nothing comes into being or ceases to exist. To say that the universe has a beginning is then just like saying that a street has a beginning, where the street's having a beginning doesn't signify that it comes into being but only that it starts its course in a certain place. Paradoxical though it may sound, according to B-theorists, the universe is eternal even though it has a beginning, and this is Craig's reason for siding explicitly with the A-theorists.

Collins's contribution on the argument from design is connected with Craig's, not only for the obvious reason that they both deal with important arguments of rational theology, but also for a deeper reason. Even though the arguments – Craig's Kalamic argument concerning cosmological models and Collins's teleological argument concerning fine-tuning – draw on empirical considerations, they are both conducted at a metaphysical level or, more accurately, at a level at which metaphysics has a bearing on the empirical. If, for example, we ask why, in formulating Craig's cosmological argument, it is not sufficient to assume the finitude of cosmic evolution beginning with the singularity of the initial moment, we can reply, for reasons given above, that this depends on how one chooses between the A-theory and the B-theory of time, and that such a choice is a metaphysical one. The metaphysical character of such a choice can also be seen from the consequences of accepting one of these theories rather than the other. If one accepts the A-theory, a beginning signifies a coming into existence and, consequently, the existence of the universe is not explained unless one appeals to a cause external to the universe itself. Its existence would not be explained, because in the context of the A-theory the initial moment of the universe is the mark of its ontological contingency – if it exists, it does not do so in virtue of itself – whereas if the universe were ontologically necessary, it would not have need of a transcendent ground. In confirmation of this, if the B-theory were correct, each moment of the temporal series would exist eternally, including the very moment of the initial singularity, so that the series itself would be given eternally – that is, as something ontologically necessary – and there would be no need to seek an external explanation for it. As can easily be seen, the key to Craig's entire analysis of the cosmological argument lies in the point that the universe requires an ontological

foundation. If, then, the universe is ontologically contingent, its foundation is external to it and is transcendent; whereas if the universe is not contingent, then it is its own foundation, but this can be the case only because it is the universe itself that manages to confer upon itself its own ontological necessity.

The ontological point of view – and not only the cosmological – also plays an important role in the context of the argument from design, dealt with in Collins's contribution. Collins examines the argument from design with the phenomenon of *fine-tuning* as his starting point. This phenomenon consists in the highly improbable fact that the values possessed by cosmological constants at the moment of the Big Bang were, to a very precise and finely determined degree, the only ones permitting the evolution of the cosmos in a way suited to the generation of intelligent life. However, the extreme improbability of fine-tuning is enormously diminished – fine-tuning ceases, indeed, to be improbable – given the hypothesis of design, whereas it remains high given the hypothesis of chance. From this and the *likelihood principle*, Collins draws the conclusion that the fact of fine-tuning supplies strong evidence in favour of the hypothesis of design as opposed to the atheistical hypothesis of chance. However, Collins's actual conclusion is, in fact, weaker than the one that has literally just been stated. That conclusion is valid, rather, only if one can exclude the many-universes hypothesis – the hypothesis according to which our universe, with its entire evolutionary history, is just one of many (perhaps infinitely many) universes differing in various ways from our own and existing alongside it. If, indeed, the many-universes hypothesis were true, then the initial coincidences of our own universe would not be at all improbable. On the contrary, they would be ontologically necessary, precisely because the many-universes hypothesis would allow us to say that nothing is possible that is not realized in some universe, and this would apply also to the anthropic coincidences. But if fine-tuning were necessary, it would have maximal probability, and thus it would no longer make sense to apply the likelihood principle to it. As one can see also from the reply to Collins's paper, the argument from design requires, like the Kalamic argument, a point of departure that lies beyond the purely empirical framework supplied by Big Bang cosmology. One must argue against the plausibility of metaphysical hypotheses like the many-universes hypothesis and in that way provide warrant for the contingency of the origin of the universe itself. The need for a perspective that is broader than that provided purely by science follows additionally from the bearing that axiological considerations have on the very estimation of the epistemic probability of fine-tuning under the hypothesis of design. This fact comes out particularly clearly in remarks made in the reply to Collins's paper.

The contributions in Part I dealing with the theory of knowledge present two strongly anti-naturalistic arguments. The first of these arguments is a reworking of the well-known evolutionary argument against naturalism first

expounded by Alvin Plantinga in Plantinga (1993b), which has given rise to a fruitful debate that is still ongoing. According to this argument, it is not possible to explain in terms of evolutionary processes why humans appear to possess a reliable cognitive apparatus, that is to say, an apparatus capable of generating beliefs that are on the whole true. Indeed, the mechanisms on which evolution is based are driven by adaptive constraints, and the revelation of truth does not figure among these. It is not knowledge of truth that is conducive to survival but adaptation of our behaviour to the environment, which only by chance involves knowledge of truth. In other words, adaptiveness is not associated with truth *as such*, but eventually leads to truth only in so far as the latter contributes to bringing about adaptive behaviour. Assuming the naturalistic perspective therefore renders it highly improbable that our cognitive faculties are reliable, from which a sceptical conclusion immediately follows. Plantinga infers, thus, that if one wishes to reject scepticism, one is bound to repudiate naturalism.

Plantinga's argument is the focus of an extremely intense and fertile debate, signs of which are also to be found in the reply to his paper. Some commentators on the argument, for instance, emphasize critically the fideistic presupposition on which it is based. Others regard as implausible the claim that a true belief can, in general, only by accident be useful for evolutionary purposes. Obviously, this could happen in some cases, but not universally and, above all, it is not sensible to suppose that it happens systematically. One could perhaps say that there are certain types of belief – where the type is defined by the content of the belief itself – whose existence is irrelevant to the ends of evolution and hence that such beliefs serve to justify Plantinga's thesis. It is essential, however, to establish that beliefs of this type do genuinely exist, which means that these beliefs will have to be characterized by quite specific and irreducible contents. For example, it is perfectly sensible to say that the truth of certain theoretical beliefs – such as many mathematical beliefs possessing highly abstract contents – is irrelevant as far as survival is concerned and therefore that the genesis of a corresponding cognitive faculty in human beings cannot easily be explained by recourse to the theory of evolution. But the problem at bottom is whether a set of true beliefs characterized in such a way do exist and are irreducible. For only in that case does the problem arise of explaining – in non-evolutionary terms – how a faculty could emerge that is capable of grasping the contents of such a set of beliefs. If the question is posed in these terms, however, Plantinga's anti-naturalistic argument comes in essence to coincide with some of the classical arguments against naturalism – for example, the argument that our cognitive activities exhibit emergent features, such as the capacities for abstraction and *a priori* judgement and others that are not easily accounted for by a naturalistic model.

In contrast to Plantinga's paper, Franz von Kutschera's starts precisely with the analysis of one of the characteristics of human mental activity, our capacity for reflection. This capacity is the source of the distinctively higher

cognitive activities of human beings and von Kutschera devotes the first part of his paper to illustrating some of these, including consciousness in particular as the focus of mental life, both theoretical and practical. The most original parts of von Kutschera's contribution, however, are the two that follow. In the first of these, he argues in favour of a constructivist conception of the notion of a set and, quite generally, a conceptualist conception of abstract entities. The upshot of this line of argument is that being a conceptualist means conceiving of the domain of abstract entities as an open one, capable of unlimited extension through the construction of ever higher levels of reflection. But the very openness of the domain of abstract entities, which takes concrete form in the openness of subjects' beliefs, is a characteristic which, as von Kutschera demonstrates in the final part of his paper, is incompatible with even the weakest notion of the supervenience of the mental on the physical. Without contending that the mental has a strong form of independence from the physical, von Kutschera finally arrives at an argument against the reduction of mental life in all its complexity to mere neurophysiological processes.

The papers in the remaining parts are closely connected with one another in virtue of a common theme which manifests itself in each of the contributions at the different levels on which they operate. This is the idea that anti-naturalistic theses are particularly relevant to human agency and all of the conceptual categories connected with it. It is an idea that is very evident in Edmund Runggaldier's paper, in Part II on ontology. In his contribution, Runggaldier shows how naturalistic ontology, in presupposing a scientific description of reality, gives rise to a dichotomy between theory and practice. Indeed, although it is able to cope with natural properties, by drawing on a scientific account of our objective knowledge of the world, it finds it difficult to adopt the same point of view in dealing with the phenomena of subjective human experience. The problem of the relationship between the theoretical dimension of human knowledge and the practical and subjective aspects of human life impinges directly on the complex of questions concerning agency that forms the heart of the debate about naturalism. What is human action? Is human action free? What does it mean to explain human action? What is the relationship between action and values? The papers in Part II on ontology, in Part IV on the philosophy of mind and in Part V on practical philosophy furnish answers to each of these questions that exhibit a profound unity and a common commitment to anti-naturalism.

The thesis held in common by all the authors in these three parts, and constituting the systematic assumption to which all of their other anti-naturalistic claims can be traced, is the proposition that it is impossible to give an adequate definition of human action without recognizing that such action is conscious and free. There is no human action properly so-called if there is no genuine possibility of choice – in the libertarian sense of the term – between different alternatives. This is, in particular, demanded by the fact that we feel ourselves to be morally responsible, which would not

be justifiable without an acknowledgement that it is in our power to deliberate in favour of this or that choice. But the conditions for free action are many and exacting.

First of all, it is necessary for action to be conscious. Uwe Meixner and Hugh McCann urge in a profound and detailed fashion that consciousness is presupposed by free action. For both authors, it is necessary for one to be conscious of the object of choice, because otherwise one could not have control over the action that occurs as a consequence of a choice. Furthermore, the consciousness involved in the act of deliberation is considered by both to be an irreducible given, not explicable by a reduction of the mental to the neurophysiological level. This is particularly clear in the model proposed by Meixner, in which consciousness plays a role in determining the actions of agents of every kind, including lower animals. A condition of this model is that action occurs in a non-deterministic context. Thus, according to Meixner, to say that an agent x is conscious of p is to say that p is meaningful for x, in the sense that information concerning p, carried by x's state of consciousness of p, puts x in a position to be able to engender or impede, to his own advantage, the occurrence of p. Hence, for Meixner, to be conscious of a state of affairs is to be able to utilize information concerning that state of affairs in acting to one's greatest advantage, which rules out the possibility that the natural world is deterministic. Otherwise, an agent could not make use of the information of which it was conscious and would be constrained to follow the independent course of natural events, which would determine its actions inexorably. In that case, however, we would have to concede that agents – and in particular human agents – have decidedly limited epistemic capacities, since they would be systematically subject to the illusion of supposing that they could intervene in the course of natural events, without in reality having a power to do so.

As we can see, Meixner's paper opens up all the problems of action, various of which are approached by the other authors from several different angles. The diverse aspects that they deal with can all be related to two broad issues connected with the last two questions posed above. On the one hand, there are problems to do with the explanation of human action and, on the other, problems concerning practical knowledge. These two themes are, obviously, intimately interrelated.

The issue of the explanation of human action is principally examined by Jonathan Lowe and Hugh McCann. If action is understood to be conscious, the product of freely executed deliberation, then it is evident that action cannot be explained in accordance with models deployed in the natural sciences. More particularly, one must distance oneself from a causal model of explanation, in which causation is understood as a relation of production or succession among natural events. One must assume a teleological model, in which the explanation of action hinges on the intentions of the agent and the agent's beliefs about the world. This is insisted upon with particular force and clarity in Lowe's paper, which devotes several pages to deepening

the differences between causal explanation – which is valid in the domain of the natural sciences, including cognitive science itself – and explanation by means of reasons, which is valid in the field of intentional action. The conclusion that Lowe comes to is that human action is not suited to causal, but rather to rational, explanation. This allows us to formulate the problem of free will correctly, as one concerning a choice between alternatives that is not determined by causal mechanisms beyond the agent's control, but is left up to the agent, who ponders over various reasons – without being determined by any of them – before coming to a final decision. In this way one avoids the danger of having to say that our decisions occur merely by chance as a consequence of denying that they are caused. Denying that human action has a causal explanation is not equivalent to regarding our actions as the results of chance, understood simply as the absence of causes. Such a conclusion is not necessary, because the will makes its choices on the basis of reasons and all that is required for the purposes of freedom is that the subject should have control over the final decision. Of course, maintaining that the subject has control over the decision is not trivial, but it is the very least that is required – particularly in McCann's view – in order to secure a non-naturalistic vision of human beings. Clearly, the theory of rational action-explanation defended by Lowe and McCann also embodies a characteristically anti-naturalistic assumption of the sort that Meixner highlights. Just as it would make no sense to attribute to human beings a consciousness incapable of intervening in the course of natural events – for the simple reason that, as remarked above, this would be a consciousness perpetually doomed to illusion – so likewise it makes no sense to suppose that one can explain action rationally without postulating that the acting subject is capable of control. Similarly, this is for the simple reason that explaining in terms of reasons means supposing that reasons can be pondered over, evaluated and, in a word, controlled, right down to the final act of choice, which in its turn is an act of control *par excellence*.

With these last observations we come to the second problem area connected with the general theme of action: the place of values and preferences in the whole fabric of reasons underlying choice of action and the role played, from this point of view, by practical knowledge. The analysis of action advanced in the papers by Lowe, Meixner and McCann leads to an explanation of action in terms of reasons. From a logical point of view, this means that the explanation of human action proceeds according to a scheme conforming with principles of practical inference, rather than the Hempelian scheme or anything like it. But this in turn implies that the premises of any argument capable of explaining an action are – reducing the schema to its bare bones – volitional and epistemic in nature. There are no empirical laws in the *explanans*, and the nexus between *explanans* and *explanandum* is constituted by a logic specifically capable of connecting volitional and epistemic premises to a description of the action. Such a logic is the *logic of volition*, characterized precisely by the principle that unimpeded

volition analytically entails the realization of the willed action. On the other hand, the beliefs, desires and ideals of the agents are important factors bearing on their action: for they choose what action to perform in the light of their own beliefs, desires and ideals. How, then, is it possible that such factors should have a bearing on the choice of action and, as a consequence, be efficacious for the purpose of its performance, without its being the case that the action is actually caused by them? As was remarked above, this is possible because a genuine action is something that the agents choose to do in the light of their beliefs and desires and not because of them. If, therefore, the logic of volition is the logic of explanation in terms of reasons, we must show exactly how such reasons interact with the logic of volition.

Important pointers towards such an account are supplied, not only in the papers by Lowe and McCann, but also in Robert Audi's, as well as in the replies to McCann and to Audi himself. McCann urges in particular that choosing for a reason is a teleological relation, not a causal one. It follows that if an agent's desire to do something, together with relevant beliefs of the agent, are understood by the agent as reasons for doing that thing, then it is not the agent's mental states of desire or belief as such that have explanatory force, but rather their contents. Such a content is not an event or a state, but a 'thought', which is not an entity of an appropriate type to explain anything in causal or statistical terms. The mental contents of reasons for an action make reference to the end to which the action is directed, to the extent that this is anticipated by the agent as being positive or desirable, and to the most opportune means sufficient in the circumstances to realize that end. Finally, the explanatory value of reasons resides in the particular logical relationship that obtains between the relevant contents.

Here it is particularly important to distinguish between the content of the volitional premise and the contents of other relevant premises. This distinction gives a great practical significance to many of the points made in Audi's paper and the reply to it. In fact, the content of the volitional premise is traceable, in the final analysis, to the acting subject's set of preferences. Simplifying somewhat, these can be represented in terms of (somehow) ordered structures, whose elements are states of affairs that are objects of evaluation on the part of the agent. Obviously, the evaluation of such states of affairs – and the consequent ordering of them which can for this reason be termed a preferential one – is determined by the subjective preferences of the agent. But often in the process of forming these, an important role is played by the ordering of values in which the subject believes as values of an objective or ideal kind. Now, it is not easy to provide a purely naturalistic account of such values: for, being deemed objective, they cannot be treated as values that are subjectively traceable to individual desires and, being ideal and therefore abstract, neither can they be seen as natural states of affairs rooted in the biological structure of the human species. Such values need to be regarded as possessing an epistemological status that secures both their autonomy with respect to natural

qualities and their supervenience upon them, noting that a supervenience relation in this domain is not to be understood in a nomic sense. The complex of problems connected with a non-naturalistic conception of values, applicable not only within the theory of action but also in the broader context of practical philosophy and metaethics, is the target of overall concern in Audi's paper and the reply to it.

Returning to the question with which we began this introduction, we can, following our overall evaluation of the different contributions, confirm our initial thoughts concerning the relationship between naturalism and the analytical method. The question at issue involved two aspects. On the one hand, it had regard to the problem of whether the presumption of naturalism is inherent in the analytical method or whether, on the contrary, it is tied to the latter in a contingent way ungrounded by principle. On the other hand, it inquired whether there is a non-naturalistic approach that is fundamentally unified, global in scope and methodologically comparable with that of naturalism, but constituting an alternative to the latter. The answer is positive in both cases. The analytical method is not united to the naturalistic approach, so we can coherently claim to be both analytic philosophers and non-naturalists. And, indeed, the most significant thing to emerge from the papers in this volume is that all of them contribute, in a unified and convergent fashion, towards elaborating the fundamental theoretical outlines of an alternative picture to that of contemporary naturalism. The common feature of all the contributions, which gives precise sense to the anti-naturalistic picture that they disclose, seems to be their affirmation of the *openness of knowledge and reality*: the openness of reality to the transcendent, the openness of knowledge to forms of cognition free from the constraint of adaptation, the openness of the mind to the construction of mental structures that are ever more complex and abstract, the violation of the causal closure of the physical world, the emergence of the mental with respect to the physical, the openness of the will to free choice – in sum, the openness to ontological, epistemic and practical *novelty*. In contrast to anti-naturalistic conceptions of a traditional kind – which are often representative of philosophical paradigms incommensurable with that of contemporary naturalism – an anti-naturalistic picture of this sort is set within the framework of analytic philosophy and its characteristic method. The great advantage of this is that the differing models of contemporary naturalism and anti-naturalism are able to confront one another. Moreover, the unity of the non-naturalistic model, together with the global nature of the themes that serve to contrast it with the natural sciences, enables us to keep alive the dialogue with science while at the same time avoiding the insidious danger, rooted in the scientific method, of abjuring all knowledge of a synthetic kind.

Part I

Epistemology

1 Reflection

Franz von Kutschera

The topic of this paper is an intellectual capacity which we constantly use but seldom think about: the capacity to reflect upon what we think or do, on what we plan, feel or wish. Hence, the difficulty of reflecting on reflection lies not in a lack of acquaintance with the phenomenon but rather in the invisibility of the familiar. The first part of the paper outlines the dimensions of the phenomenon. It will be less concerned with a general definition of reflection than with its varieties, its ubiquity and importance. The second part describes what can be termed the reflective structure of the mind and its constructs and pleads for a conceptualistic understanding of abstract objects. The third part, finally, bases a simple argument against materialistic reductions of the mental on this reflective structure.

The dimensions of reflection

As we are interested here not in the physical phenomenon but only in the psychological one, 'to reflect on something' means in general to think about it. More specifically, reflecting means turning the attention from the objects of the outer world to the subject and its activities and states, to its perceiving, sensing or aspiring. Still more specifically, reflection consists in topicalizing previous mental acts and states. Therefore the structure of a reflection can be described as leading from one act or state to other intentional acts that have the first one as their object.[1] I listen to a symphony, for instance, and then realize that my thoughts have wandered and that I have taken in only very little of the music. I accept an invitation and then reflect on whether it was right to do so, or on what it was that made me accept the invitation. I make a claim and then consider whether it was justified. I feel unhappy and reflect on the reasons. I want to make a trip into the mountains and ask myself whether, in view of all the unfinished work on my desk, it would not be better to stay at home. In all these cases, to reflect means to disengage from what one is doing, feeling or wanting, to take a step back, bring it into focus and think about it.

The capacity for reflection is a trait of the human intellect whose theoretical and practical importance is hard to overestimate. If we do something,

want something or feel something, we are more or less aware of it. To gain a clear idea of it, however, we have to reflect on it. Only the objects of intentional attitudes lie in the clear light of our consciousness and our attention. If I contemplate a tree, for example, my attention is normally focused only upon the tree and its properties. I am not altogether unaware of myself and of my contemplative activity, but these are not my topic and remain out of focus. We see things as in the light of a torch which illuminates only what is in front of us, but not ourselves. To perceive ourselves, and what we do, we must make ourselves the topic of a reflection. The capacity for reflection, therefore, is a precondition for forming a clear and distinct notion of ourselves.

Reflection, furthermore, is a precondition for a critical evaluation of our own actions. A scientist has constantly to reconsider the justification of his assertions and his theories, and in everyday life we must examine our own actions regarding their expediency or their moral permissibility. In order to control what we say, feel and want, we must reflect on it. Reflection, then, is also a precondition for our ability to gain that distance from ourselves necessary for judging and readjusting our own behaviour and for evaluating single actions as well as our entire way of life, our situation, our aims, dispositions and abilities. A person's intellectual quality depends on how she develops and employs this reflective ability, the extent to which she is able to judge matters involving herself objectively.

Reflection, finally, is the path to self-knowledge. We have just reflected on reflection, and thereby gained some insights about ourselves. In the history of philosophy, however, the idea of reflecting on reflection as a way to achieve deeper self-knowledge has taken on a more specific form. A reflection of this kind starts, in my own case, with an act which I perform: a promise, let us say. In a second step I reflect on this act and inquire, for instance, into my motives for making this promise. Then I reflect again on what I have just done: that is, I reflect on my activity of reflecting. In this third step I may focus my attention on different aspects of the second reflection, especially on its motivation again. Let us assume that in first reflecting on my promise I wanted not only to recognize and enumerate my motives but also to evaluate them and see whether they were justified and whether I could subscribe to them on reconsideration. Then the criteria for my evaluation in the first reflection cannot have been simply those preferences that I followed in making the promise: for the evaluative purpose they must rather have been my notions about objective values or norms. I cannot decide on the rightness of an action and that of the preferences on which it was based by referring to these preferences themselves; I can only do so by looking to what I believe to be objectively right. The second reflection shows up this kind of evaluation and may thus help me suddenly to recognize that the orientation to what I believe to be objectively right is more fundamental for me than other preferences and aims.[2]

Theoretically as well as practically, therefore, the capacity for reflection is extremely important. To reiterate once again, we are not concerned here just

with reflection in the broad sense of topicalization, without which, of course, we could not investigate or describe anything; our theme is reflection in the specific sense of a topicalization of our own previous mental states and activities, and the fact that we can grasp these only mediately by passing from performance to reflection.

Our ability for reflection depends largely on language. Language is not only a means for the expression and communication of thoughts but thinking and speaking are inseparable and are developed in close correspondence. This was first emphasized by Wilhelm von Humboldt (Humboldt 1960, vol. VII: 53), who pointed out that linguistic articulation is essential for distinctness of thought. There are two reasons for this. First, expressing a thought linguistically means giving it an objective form in which it can be perceived optically or, in spoken language, acoustically. Now, our cognitive competences are best developed for comprehending the outer world. Only in sensory perception are we confronted with well-defined objects; only by sensory perception can we make fine and accurate distinctions, which are much more difficult in the field of mental phenomena. Linguistic expression, then, allows further distinctions to be made in whatever is expressed by correlation with sensible differences. It also makes reflection on them much easier. It is not possible to distinguish pure emotions or impressions by placing them side by side and comparing them, for, roughly speaking, the mental realm admits of no simultaneity but only of succession. If they are described linguistically, however, the descriptions can be compared and analyzed even if the emotions and impressions themselves have passed away.

Second, the linguistic expression of mental states and activities makes it possible to talk about them. Because we live in a common world, agreement with others is a criterion for reality. The expression of our opinions opens the way to additional, intersubjective criteria for their correctness and saves us from having to rely on our own impressions alone. In one sense, the development of a common language presupposes common experiences, of course, but in another sense, common experience presupposes a common language since it is only by this that we can determine what contents of experience we share.

The reflective structure of the mental

In my introductory remarks I said that I wanted to argue for a conceptualistic understanding of abstract objects such as concepts, propositions and sets. First, however, I shall start from such an understanding, for it is by viewing abstract objects as constructions of the human mind that I can best represent what I term *the reflective structure of the mental*. Let us take *concepts* first. First-order concepts are defined on a set U_0 of objects that are not themselves concepts. In the case of number-theoretic concepts such as 'is a prime number' or 'is smaller than', U_0 would be the set of natural numbers.

If we now reflect upon such first-order concepts and take them as new objects and add them to U_0, thereby going from U_0 to a more inclusive set U_1, we can also define second-order concepts on this new set, such as 'is a transitive relation' or 'is an instance of the property ... '. We can then repeat this step and count also second-order concepts among our objects, arriving at a still larger set U_2. On this we may then also define third-order concepts such as 'applying to all second-order concepts', and so on. In this way we arrive at a hierarchy of concepts and their domains as envisaged in (cumulative) type theory. This hierarchy of concepts remains open, because we can always enlarge our last universe of discourse by adding the concepts that we have used to characterize its objects. A set of all concepts does not exist, therefore.

If we are talking about concept formation, we take a *conceptualistic* stance where abstract objects are constructions of the human mind. The opposite stance is that of *Platonism* or realism with respect to universals. For it, abstract objects are not mental constructs but exist independently of human thinking; they are, as Frege said, not created but only grasped by us. The strongest argument for conceptualism is that Platonism gets entangled in contradictions. If concepts exist ready-made out there, there must be the set of all concepts. But then there must also be the set of all properties that do not apply to themselves. An instance of such a property is being a frog, which being a concept is not a frog. But if we ask whether the property of not being self-applying is self-applying or not, we get a contradiction because it is evidently self-applying if and only if it is not self-applying.

Such paradoxes have been discussed most extensively for sets rather than concepts, because it is in terms of sets that they can be represented most simply. As a fundamental mathematical discipline, moreover, set theory is of prime importance. If we start from the classical concept of set, a *set* is the extension of a property. The set of frogs, for example, is the extension of the property of being a frog and contains as elements exactly those things that are instances of this property, i.e. frogs. This conception of set is also the one held by the founders of set theory, Gottlob Frege and Georg Cantor. Their – as hindsight has it – 'naïve' set theory rests on just two simple principles. The first is the axiom of comprehension: 'Every property has an extension.' The second is the axiom of extensionality: 'Sets having exactly the same elements are identical.' From these together it follows that every property has exactly one extension, that it determines exactly one set of its instances.

Although these two simple principles are immediate consequences of the concept of a set as the extension of a property, they still give rise to a host of contradictions, the paradoxes of naïve set theory. The simplest one is Russell's: there are sets that are not elements of themselves, as our set of frogs testifies. The Russell set – let us call it R – is the set of all such sets, i.e. the set of all sets that are not elements of themselves. Now, is R an element of itself or is it not? It contains itself as an element if and only if the defining

property of not containing itself applies to it, i.e. if it does not contain itself, and that again is a contradiction.

In fact, Russell's paradox follows from the principle of comprehension alone. But wherein lies the mistake if this principle, as has been emphasized, follows directly from the normal, classical understanding of set? To see that, we must realize that the construction of the paradoxes presupposes the existence of the set of all sets, the distinctive principle of Platonism. The founders of set theory, Cantor and Frege, were both staunch Platonists, as are the majority of contemporary mathematicians. To them the metaphysical reality and eternity of abstract objects seemed overwhelmingly plausible and, therefore, also the existence of a closed universe of sets. However, Platonism in regard to sets is really not very convincing. According to the classical notion, a set is an extension of a concept. Now concepts are forms of conceiving which we create, which come into use, and pass out of it. They are therefore naturally thought of as mental constructs. But then this conceptualism which is so natural for concepts should also be extended to their extensions.

I do not want to present these brief intuitive considerations as a refutation of Platonism. The main argument is simply the inconsistency of Platonistic theories of sets, concepts and propositions, and the fact that there are intuitively well-founded and consistent conceptualistic alternatives.[3] In a conceptualistic theory, concepts and sets form an open hierarchy whose elements are all *grounded*. In the case of sets, this means that there are no infinite element-chains. Therefore no set can be an element of itself. Russell's set does not exist, and neither does the universal set, which would have to be an element of itself.

Since I have mentioned a hierarchy of propositions I must briefly describe it. First-order propositions are those that we can form without presupposing other propositions. Second-order propositions may presuppose only first-order ones, and so on. A hierarchy of this kind is needed, for example, to avoid the paradox of self-applying propositions that Alfred Tarski formulated for sentences. A universal proposition of the form 'All propositions have property *E*' – for instance, 'All propositions asserted by Max are wrong' – is called *self-applying* if and only if it has the property *E* itself; in our example, if the proposition is either not asserted by Max or is false (Tarski 1949: 80, fn. 11). Now, is the proposition that all propositions are not self-applying self-applying or not? Again we find that it is self-applying if and only if it is not self-applying. This construction corresponds closely to the one in the paradox of the concept of non-self-applying concepts or in Russell's paradox. In the case of propositions, we do not need principles of comprehension or extensionality for the construction, only the Platonistic assumption that propositions are not something we form but something that exists from eternity, independently of human thought, which, in the case of propositions, is no more plausible than in the case of concepts and their extensions. If the set of all propositions exists, then there are self-applying

and non-self-applying ones among them, and the proposition that all propositions are not self-applying makes sense and must be contained in the propositional universe. Conceptualistically, however, the construction collapses. If we have a set *P1* of propositions, we can make assertions about all the propositions in *P1* and form the corresponding propositions. Since the formation of such propositions presupposes those in *P1*, they cannot belong to *P1* themselves. If we add these propositions to *P1* and thereby obtain a larger set *P2*, the sense of the word 'all' is changed if applied to the elements of *P2* instead of to those of *P1*. The proposition 'All propositions in *P1* are not self-applying' is therefore not in *P1*. It does not refer to itself, and thus the contradiction evaporates.

In the hierarchical construction of abstract objects, the reflective structure of the mental and its fundamental role become especially clear. It represents our ability to create new objects out of previous acts of conceiving and judging, and thereby make them eligible for detailed study. The essential openness of the hierarchies of abstract objects, moreover, mirrors the creativity of the human mind. Set theory is not only a mathematical discipline; it is also the most developed theory of mental operations. Philosophy of mind should therefore pay it the attention it deserves, and that is much more than it has yet received.

An argument against materialism

Materialism has many variants. It is a metaphysical Proteus, difficult to pin down because whenever one tries to get a grip on it, it changes its form. If we want to argue against materialism in general, it is best to think of it as a thesis about the *supervenience* of the mental on the physical. Since there are many concepts of supervenience, however, it is against the weakest of the supervenience claims still breathing some materialist spirit that a refutation should be directed. I refer to the claim of a *global supervenience* of mental on physical states of affairs.[4] Even then, there are *prima facie* still several versions to be dealt with: besides theses of analytical supervenience, also such of nomological supervenience. It is doubtful, however, whether a claim of nomological supervenience of the mental on the physical can still be classified as materialist, because this claim can also be upheld by a form of dualism, namely psycho-physical parallelism as put forward, for example, by Arnold Geulincx, Nicole Malebranche and, most prominently, Gottfried Wilhelm Leibniz. Parallelists, to be sure, maintain in addition that there is no causal interplay between the two realms which a materialist would not accept, of course, but global supervenience is neutral with respect to causal relations. Nomological global supervenience merely says something about the coexistence of mental and physical states, while causal relations always connect earlier with later events. A distinctive materialistic content is therefore only to be found in claims about an analytic global supervenience of the mental on the physical.

Let us concentrate on *doxastic* states of affairs, i.e. propositions that somebody believes that this or that is the case,[5] as a special case of mental states of affairs. The claim of a supervenience of these states of affairs on physical ones, and therefore also the general claim of a supervenience of all mental states of affairs on physical ones, cannot even be formulated if we accept the conceptualistic understanding of propositions argued for in the previous section. Since with respect to every set of propositions new doxastic propositions may be formed – namely, propositions that this or that proposition in the set is or is not believed – there is just no set of *all* doxastic propositions that could be supervenient on the set of physical ones. If we start with physical propositions that are not concerned with beliefs, first-order doxastic propositions are about beliefs in such physical propositions. Second-order doxastic propositions are about beliefs in the first-order ones or in connections between them and physical propositions, and so on. In this way we arrive at an open hierarchy of doxastic propositions, and the claim of a supervenience of *all* doxastic propositions on physical ones cannot even be meaningfully formulated, as I said; there is no universe of discourse for this 'all' to make sense. Only a thesis about the supervenience of all the doxastic propositions *of a certain order* on physical ones can be formulated, but the belief that such a supervenience holds is then in each case a higher-order proposition for which the claim does not hold. This, I think, is a very strong argument against materialism.

If, in spite of my plea for it, one does not want to presuppose a conceptualistic understanding of propositions or states of affairs, it is possible to argue simply thus: If T is a set of propositions, a possible doxastic attitude of a person P towards them is determined by the subset of those propositions of T which P believes. There are just as many possible doxastic attitudes of P towards the propositions in T as there are elements in the power set of T, the set of all the subsets of T. Now the power set of a set always has a higher cardinality than the set itself, i.e. it always contains more elements than the set itself, and the greater the original set, the greater the difference of their cardinalities. If the set T has n elements, the power set of it has 2^n elements, and this also holds for infinite cardinalities. On the other hand, the analytic global supervenience of a set of propositions S on a set of propositions T implies that there are not more S-propositions than T-propositions, and therefore the set of propositions that person P has this or that doxastic attitude towards physical propositions cannot possibly be supervenient on the set of physical propositions.

Now, materialism often only assumes a *nomological* global supervenience of the mental on the physical. But then in the realm of what is possible according to the laws of nature, there still could be no more doxastic attitudes towards the physical than physical states of affairs. Therefore even a nomological global supervenience would radically restrict what we can believe about the physical.

Thus far I have presupposed a *descriptive* concept of belief according to which even inconsistent propositions may be believed and not all the logical consequences of what one believes must be believed, too. In my paper (von Kutschera 1994), however, I used a *normative* concept of belief implying perfect rationality. For this there is a well-developed logic: *doxastic* or, if the concept of knowledge is included, *epistemic logic*.[6] For the normative concept of belief there are very strong connections between those propositions that can be believed simultaneously by the same person; in fact, they are so strong that there is a proposition from which they all follow. The argument from cardinality consequently collapses. In restriction to a propositional language, furthermore, the higher-order beliefs are logically equivalent to first-order beliefs, so that the hierarchy of beliefs also collapses. But even in this radical simplification, a global supervenience of doxastic on physical propositions still has absurdly implausible consequences. Let us make the very plausible assumption:

(*) *For every analytically possible proposition it is analytically possible that it is believed.*

Then the assumption of an analytic supervenience of doxastic on physical propositions implies that it is analytically necessary that everybody is omniscient, i.e. that everybody believes exactly those propositions that actually hold (von Kutschera 1994: 107). If we assume only a nomological supervenience and also, as materialists do, that the fundamental laws of nature are physical laws,[7] we still obtain from (*) that everybody is nomologically omniscient.

My arguments have presupposed that a global supervenience of *S*- on *T*-propositions is so defined that two possible worlds agreeing as to what *T*-propositions hold in them also agree as to all *S*-propositions. David Lewis has advocated a weaker concept of analytical global supervenience of *S*- on *T*-propositions. Aside from restrictions that need not concern us here, it may be so defined that one of the two possible worlds must be the real one. Then the claim is that all possible worlds that are physically indiscernible from the real one are also psychologically indiscernible from it.[8] Aside from the fact that such a claim cannot give rise to psycho-physical laws, together with the assumption (*), it still has the absurd consequence that everybody is *in fact* omniscient.

The result is, therefore, that even the weakest claim of a supervenience of the mental on the physical which might still be called 'materialistic' is untenable. If no such supervenience holds, this of course does not mean that there are mental processes which are not accompanied by physical ones, especially by brain processes. It only means that we cannot always distinguish neurologically between psychologically different events, and that neurological classifications have sometimes to remain coarser than psychological ones.

Notes

1 The term 'object' is used here in a very broad sense, that of 'topic', so that propositions are also objects in this sense.
2 Evidently a third reflection on the second one would not produce new results. For this practical form of a reflection, see Kutschera (1999: 312ff.). The idea of a reflection on reflection as a way to self-knowledge first appears, as far as I am aware, in Augustine, *Confessiones*, Book VII, § 17, where he writes:

> Quaerens enim, unde adprobarem pulchritudinem corporum sive caelestium sive terrestrium et qui mihi praesto esset integre de mutabilibus iudicanti et dicenti: 'Hoc ita esse debet, illud non ita', hoc ergo quaerens, unde iudicarem, cum ita iudicarem, invenerem incommutabilem et veram veritatis aeternitatem supra mentem meam conmutabilem.

3 See Boolos (1971), and for another version of a conceptualistic set theory, see von Kutschera (2001).
4 Even a dualist who assumes that physical criteria are sufficient for personal identity would accept a weak supervenience of the mental on the physical. See von Kutschera (2003: 30ff.).
5 I do not distinguish between states of affairs and propositions in this paper. For a distinction and its problems, see von Kutschera (2003:106ff.).
6 For a comprehensive exposition, see Lenzen (1980).
7 Or propositions of a certain maximal order according to our hierarchical model.
8 See Lewis (1983: 364), where the restriction is that only such worlds be considered that contain no natural kinds – incorporal spirits, for example – not occurring in the real world. See also Chalmers (1996: Chapter 2).

Reflection, self-consciousness and intentionality

Comment on von Kutschera's paper

Michele Lenoci

Von Kutschera's paper is based on the following points: (1) the theoretical and practical importance of reflection is so described as leading from one intentional act or state to other intentional acts that have the first one as their object. This is significant both for the critical evaluation of our own actions, and in order to acquire a better self-knowledge. (2) Language is essential to reflection. It enables us to communicate with others, and, above all, to compare different mental phenomena that admit of no simultaneity, but only succession. (3) The reflective structure of the mind implies a hierarchical construction of abstract objects (concepts, propositions and sets). This conceptualism is opposite to the Platonism, which is inconsistent and gets entangled in contradictions. (4) If we consider a variant of materialism, that claims the global supervenience of the mental over the physical state, we can say the conceptualist interpretation of the propositions supports a strong argument against materialism: it follows that, from the neurological point of view, we are not always able to distinguish between different events on the psychological plane; consequently, the neurological classifications will always be coarser than psychological ones.

In the considerations that follow, I intend to cover the following themes: (1) the relationship between reflection and self-consciousness from the phenomenological point of view; (2) the possibility of an act of consciousness directed primarily on itself, which, according to some, implies antinomies similar to the Russell paradox of sets; (3) the need to intend mental states as an open hierarchical structure; and (4) some remarks on the criticisms of Platonism and on the proposed confutation of materialism.

Reflection and self-consciousness

Von Kutschera briefly refers to the self-consciousness that accompanies our intentional mental activity and which should be clearly distinguished from reflection. He states:

> I am not altogether unaware of myself and of my contemplative activity, but these are not my topic and remain out of focus. We see

> things as in the light of a torch which illuminates only what is in front of us, but not ourselves. To perceive ourselves, and what we do, we must make ourselves the topic of a reflection. The capacity for reflection, therefore, is a precondition for forming a clear and distinct notion of ourselves.
>
> (von Kutschera, p. 16)

This theme has been examined from a phenomenological point of view, which in this case is very near to the analytical one, in light of descriptive psychology, by philosophers of the school of Franz Brentano (Brentano 1924: 170–183), while Sartre dedicated to this topic some interesting reflections in the first few pages of *L'Être et le Méant* (1943: 16–23). The thesis advanced in these contexts is that every intentional act primarily directed towards a distinct object or a distinct situation is also aware of itself, and so directed towards itself.

There are basically two arguments in support of this thesis: the first, primarily of a phenomenological nature, refers to the evidence of inner experience. If I grasp an object that is distinct from me, that is, if I take the other object *as* other, i.e. different from me, I must be aware of that subject, of the ego, with respect to which the other is grasped as such, i.e. as different. Second, it should be noted that where the intentional act is not also aware of itself, it could not, *subsequently*, in a moment of reflection, recognize as its own that act on which our attention is directed and that we want to focus on. In other words, if the self-consciousness that accompanies every mental state were missing, it would not be possible to distinguish reflection from an act of fantasy or imagination orientated toward a certain mental state. I can be certain that the mental act on which I am reflecting is not a product of my imagination if, granted the reliability of my memory (proven or supposed), I was, in some way, already aware of that act at the moment of its accomplishment. It is obvious that, from the phenomenological point of view, the case of self-consciousness assumes a different form from that of reflection and implies different problems. In the latter case, that of reflection, the act of consciousness aims at a mental state that is distinct from itself, while in the former, that of self-consciousness, an act would be, in some way, aimed at itself. As we know, many thinkers have seen in the identity between subject and object the privilege of self-consciousness with respect to other mental states. But this is not the topic that interests us at this moment.

Antinomies of self-consciousness

The following question then arises: in which conditions can there be an act that is aware of itself, that is at the same time and from the same point of view both subject and object? From a more detailed examination of the issue, it would seem that self-consciousness, which appeared both as a

datum of experience and a condition of subsequent reflection, was, in fact, impossible for two reasons. Such identification would not be consistent with all those conceptions for which the intentional act would always be beyond itself, towards something other than itself and different from itself. This alterity would occur even in reflections, in which the knower belongs to a particular moment of my experience, while the known belongs to another one, completely distinct from the former. In the case of self-consciousness on the other hand, it is required that the perceiver be the perceived, that the subject identify himself completely with the object. This idea appeared many times in philosophy, in different contexts and even within highly heterogeneous perspectives. Response to this objection held that the feared difficulty did not exist, since thought never changes, nor modifies its object, with the result that the subject can perceive itself, without its characteristics being altered in becoming an object. However, it must be recognized that, in self-consciousness, it is the subject that sets itself as something *other* than itself and different from itself, in order to be thought about, and, as *different*, it is no longer itself. In this way, the subject undergoes a change not at the moment in which it is thought, but as premise and presupposition in order to be thought of. In other words, it is not the thinking of the object (which in self-consciousness is the subject itself) that changes the object, but the need of becoming an object, in order to be thought of, that causes a transformation within the subject. On the other hand, for the very reason why the object should continue to remain a subject, we have a subject (object) which would intend an object (subject). Self-consciousness would therefore be possible only at the price of renouncing the main characteristic of mental states: transcending ourselves.

The second reason against the possibility of self-consciousness is the fact that it produces an antinomy similar to the Russell paradox of the set of sets which are not elements of themselves. It was Mally in one of his essays (Mally 1914: 38–39) who highlighted this point. He asks himself if a thought can perceive itself. If the proposition 'thought (or thinking) D perceives itself' has a sense, then it would be true or false, if referred to any particular thought D. The same can be said for the contrary. We could then have a thought G directed to 'a thought that does not perceive itself' or to 'a D that does not perceive D' and here, we could therefore raise the question previously held correct: that is, if it perceives itself or not. At this point we have the antinomy. If the answer is affirmative, that is, if G perceives G, then G is in fact, a thought of the kind thought in G, i.e. a thought that *does not* perceive itself; if, vice versa, we assume that G does not perceive G, then G is, in fact, a thought of the kind that does not perceive itself, that is, a thought that is grasped and perceived by G. The affirmative answer to our question is therefore equal to the negative one and vice versa: since each significant statement is either true or false, it follows that 'G perceives G' has no sense.

The plausibility of self-consciousness

Faced with these findings that seem to remove all plausibility to self-consciousness as a prior condition to reflection, two different solution strategies have been implemented. Some (such as Brentano and Sartre) have highlighted that the intentional act always involves turning *primarily* and *thematically* towards an object that is distinct from the act, towards which it aims and focuses its attention. But the intentional act is not merely punctual, that is, it is not such as to be able to have only one object at a time. On the contrary: beyond the primary object, there is always, so to speak, in a lateral, secondary and non-thematic way, another object, albeit always present to the knowing subject. This object constitutes the act itself, which is thus always known. Each intentional act is therefore always accompanied by the awareness of itself, as if by its own shadow. The second attempt at solution is of more direct interest to us, since it converges with the proposal by von Kutschera to perceive the reflective mental structure as stratified. It has been stated (Meinong 1917: 10–27), that each intentional act requires an object as its logical *prius*, so that the thought, of a thought, of a thought, *ad infinitum*, would be empty and would never have an object. It follows that in order to be directed towards a mental state, an act of consciousness must set itself on a higher level, for which reason it cannot include itself among the states of which it is aware. It follows that, in this case, we would have self-consciousness only through reflection and it would not accompany every intentional act like its shadow.

Some reflections on the criticism to Platonism and materialism

Two brief reflections on the criticisms directed to Platonism and on the confutation of materialism. Platonism is declared implausible since, if the concepts are taken as subsisting, with the result that we would only have to grasp them, there should be a set of all concepts, and this, as we know, would give rise to an antinomy of sets, which is peculiar to the naïve set theory. At this point we should ask ourselves if the conception outlined here is the only possible one for Platonism. We could think (as would Meinong 1904) that in the universe of concepts (objects) there is also room for contradictory or paradox concepts and that all that is needed is an adequate logic to deal with them. In this case, we would be giving priority to ontology and there would be the need to adapt to and to deal with this ontological structure, which is anomalous in many respects. Or, we could say that among the structure rules that govern the (presumably perfect) platonic world, a set is prohibited from being an element of itself, so that we would have a stratification of concepts (and of propositions) according to a hierarchical configuration.

With regard to the confutation of materialism presented in the third part of von Kutschera's paper, it surely places an emphasis on a real element,

given by the different and higher cardinality of the doxastic propositions with respect to the physical propositions. However, we could say that, although we cannot formulate the supervenience of *all* doxastic propositions on physical ones, we could still think that even the doxastic propositions of higher order are supervenient on (other) physical propositions. The fact that we cannot formulate the supervenience of *all* doxastic propositions, does not prevent us, *if we proceed into infinite*, from stating the supervenience of each doxastic proposition of level (n-1).

2 How naturalism implies skepticism

Alvin Plantinga

My topic is naturalism, skepticism, and the connection there between. I propose to argue that there is a connection between the two, and that indeed the former commits one to the latter. But first a word about each.

Varieties of skepticism

The term 'naturalism', of course, is used in a wide variety of senses. For example, there is the naturalism in ethics with which G.E. Moore did battle; there is naturalism in art and literature; there is naturalism as the pursuit of one who studies nature. I use the term as follows: a naturalist is someone who thinks there is no such person as God or anything much like God. Naturalism thus entails atheism. The converse, however, doesn't hold; it is possible to be an atheist but not rise to the heights (or perhaps sink to the depths) of naturalism. Thus a Hegelian might well be an atheist; she will not be a naturalist, however, because Hegel's absolute is too much like God. The same goes for a serious follower of Plato, with his Idea of the Good, and similarly for Aristotle with his First Mover who thinks just about himself. Naturalism, as I'm thinking of it, includes the views of people like John Dewey, Bertrand Russell, Daniel Dennett, Richard Dawkins, Steven Pinker, and the like. We might call it 'Atheism Plus' or perhaps 'High Octane Atheism'. According to Oxford philosopher John Lucas, naturalism is the orthodoxy of the western academy. That may be a bit strong; there are also those myriads of postmodern anti-realists with respect to truth. But naturalism is at the least very popular these days. And I propose, as I say, to argue that naturalism commits one to skepticism.

Second, skepticism. Unlike 'naturalism', 'skepticism' doesn't have a host of different senses. Still, there are several brands of skepticism. Consider, for example, contextualism in epistemology, and in particular the contextualist response to skepticism. According to Keith DeRose, an eminent authority on contextualism, 'contextualist theories of knowledge attributions have almost invariably been developed with an eye toward providing some kind of answer to philosophical skepticism' (DeRose 1999: 185). The basic idea of the contextualist response to the skeptic is as follows. The terms 'know',

'knowledge' and their like are in fact multiply ambiguous, in that in different contexts, different standards apply for their correct application. We are in an epistemology seminar; the topic is Descartes's evil genius who creates us only to deceive us. In this context, the standard for the application of the term 'know' is very high; one, or one's belief, must be in an exceptionally good epistemic position to warrant application of the term 'know'. It may well turn out, therefore, that in that high-standard sense of the term, I don't know that I'm not being deceived by such a Cartesian demon. In that same sense I don't know that I'm not a brain in a vat, holding mainly false beliefs; and if I don't know that, then presumably I don't know any proposition that entails that I am not thus envatted. In other more ordinary and everyday contexts, however, the standards for the correct application of the term aren't nearly as stringent. If you ask me whether I know my mother's phone number, I will reply, quite correctly, that I do. Also I know my name and where I live. In those contexts, the term 'knowledge' may properly apply to a given belief, even if it is not in an epistemic position good enough to warrant the term in its high-standard seminar sense.

'According to contextualists', says DeRose, 'the skeptic, in presenting her argument, manipulates the semantic standards of knowledge, thereby creating a context in which she can *truthfully* say that we know nothing or very little' (1999: 185). The skeptic, by raising her skeptical doubts, creates a context in which the standards for the proper use of the term 'know' go way up; hence she can correctly say that no one has knowledge. More exactly, her use of the sentence 'No one has knowledge' expresses a truth in that context. But of course this is compatible with its being the case that when, in an ordinary context, I say, 'I know that Tom was at the party; I saw him there' that sentence also expresses a truth.

Well, suppose the skeptic succeeds in confusing me; I apply the wrong standards in an everyday context, and form the belief that no one knows much of anything. What I am then committed to is the thought that our beliefs don't meet the high standards (for justification, warrant, rationality, whatever) required for the proper application of the term 'knowledge' in the seminar context. That is, there are some high standards of epistemic justification or warrant – those required by the application of the term 'knowledge' in the seminar context – that hardly anyone's beliefs ever meet. Of course that is entirely compatible with thinking I've got very good reason for many of my beliefs.

Now this is not the kind of skepticism that has traditionally been urged, or lamented, by the Humes and Sextus Empiricists of our tradition. It is also not the sort of skepticism to which I claim the naturalist is committed. We can see why by considering the Hume of the conclusion of Book I of the *Treatise* (Hume 1951: 263ff.). Here he isn't coolly announcing, as an interesting fact about us, that few if any of our beliefs meet those very high standards of justification or warrant to which the contextualist refers.

Instead, he finds himself in a crisis which is both epistemic and existential; he simply doesn't know what to believe. When he follows out what seem to be the promptings and leading of reason, he winds up time after time in a black coal pit, not knowing which way to turn:

> Where am I, or what? From what causes do I derive my existence, and to what condition shall I return? Whose favour shall I court, and whose anger must I dread? What beings surround me? and on whom have I any influence, or who have any influence on me? I am confounded with all these questions, and begin to fancy myself in the most deplorable condition imaginable, inviron'd with the deepest darkness, and utterly depriv'd of the use of every member and faculty.
> (Hume 1951: 269)

Of course, this is Hume in his study, some time before he emerges for that famous game of backgammon. Nature herself, fortunately, dispels these clouds of despair:

> [she] cures me of this philosophical melancholy and delirium, either by relaxing this bent of mind, or by some avocation, and lively impression of my senses, which obliterate all these chimeras. I dine, I play a game of back-gammon, I converse, and am merry with my friends.
> (ibid.)

Still, the enlightened person, Hume thinks, holds the consolations of nature at arm's length. She knows she can't help acquiescing in the common illusion, but she maintains her skepticism of 'the general maxims of the world' and adopts a certain ironic distance, a wary double-mindedness: 'I may, nay, I must yield to the current of nature, in submitting to my senses and understanding; and in this blind submission I shew most perfectly my sceptical disposition and principles' (ibid.: 269). This is the irony of the human condition: those who are enlightened can see that nature inevitably leads us to believe what is probably false, or arbitrary, or at best extremely dubious; they also see, however, that even the best of us simply don't have it in them to successfully resist her blandishments. We can't help believing those 'general maxims', or if we can, it is only for brief periods of time and in artificial situations. No one can think Humean thoughts about, say, induction, when under attack by a shark, or when clinging precariously to a rock face high above the valley floor. (You won't find yourself saying, 'Well, I do of course believe that if this handhold breaks out, I'll hurtle down to the ground and get killed, but of course (fleeting sardonic, self-deprecatory smile) I also know that this thought is just a deliverance of my nature and is therefore not really to be taken seriously'.) Still, in other circumstances one can take a sort of condescending and dismissive stance with

respect to these promptings of nature; in reflective moments in my study, for example, I see through them. As a rational creature I can rise above them, recognizing that they have little or nothing to be said for them. Indeed, I see more: this skepticism is itself a *reflexive* skepticism; it arises even with respect to this very thought; this very doubt, this feeling of superiority, this seeing through what our natures impose on us, is itself a deliverance of my nature and is thus as suspect as any other. The true skeptic, says Hume, 'will be diffident of his philosophical doubts, as well as of his philosophical conviction' (ibid.: 273).[1] The true skeptic, we might say, has a defeater for each of his beliefs, including those that lead to skepticism; and it is this kind of skepticism to which the naturalist is committed.

Reduction and supervenience

Most naturalists accept *materialism* with respect to human beings: the claim that human beings are material objects, and material objects with no immaterial parts – no immaterial soul, or mind, or self, for example. From this perspective it is not the case that a human person is an immaterial substance or thing that is connected with or joined to a material body; nor is it the case that a human being *has* an immaterial soul or mind. Instead, so the materialist thinks, a person *just is* her body, or perhaps some part of her body, or perhaps some other material object constituted by her body. I *am* my body (or perhaps my brain, or some part of it, or some other part of my body).

Now what sort of thing will a *belief* be, from this materialist perspective? Suppose you are a materialist, and also think, as we ordinarily do, that there are such things as beliefs. For example, you believe that Proust is more subtle than Louis L'Amour. What kind of a thing is this belief? Well, from a materialist perspective, it looks as if it would have to be something like a long-standing event or structure in your brain or nervous system. Presumably this event will involve many neurons connected to each other in various ways. There are plenty of neurons to go around: a normal human brain contains some 100 billion neurons. These neurons, furthermore, are connected with other neurons at synapses; a single neuron can be involved in many synapses. The total number of possible brain states, then, is absolutely enormous, much greater than the number of electrons in the universe. Under certain conditions, a neuron fires, i.e., produces an electrical impulse; by virtue of its connection with other neurons, this impulse can be transmitted (with appropriate modification) down the cables of neurons that constitute effector nerves to muscles or glands, causing, e.g., muscular contraction and thus behavior.

So (from the materialist's point of view) a belief will be a neuronal event or structure of this sort, with input from other parts of the nervous system and output to still other parts. But if this is the sort of thing beliefs are,

they will have two quite different sorts of properties. On the one hand there will be *electro-chemical* or *neurophysiological* properties (NP properties, for short). Among these would be such properties as that of involving n neurons and n* connections between neurons, properties that specify which neurons are connected with which others, what the rates of fire in the various parts of the event are, how these rates of fire change in response to changes in input, and so on. But if the event in question is really a *belief*, then, in addition to those NP properties, it will have another property as well: it will have to have a *content*.[2] It will have to be the belief that *p*, for some proposition *p*. If it's the belief that Proust is a more subtle writer than Louis L'Amour, then its content is the proposition *Proust is more subtle than Louis L'Amour*. If it is instead the belief that Cleveland is a beautiful city, then its content is the proposition *Cleveland is a beautiful city*. My belief that naturalism is all the rage these days has as content the proposition *Naturalism is all the rage these days*. (That same proposition is the content of the German speaker's belief that naturalism is all the rage these days, even though he expresses this belief by uttering the German sentence 'Der Naturalismus ist diese Tage ganz groß in Mode'; beliefs, unlike sentences, do not come in different languages.) It is in virtue of having a content, of course, that a belief is true or false: it is true if the proposition which is its content is true, and false otherwise. My belief that all men are mortal is true because the proposition which constitutes its content is true, but Hitler's belief that the Third Reich would last a thousand years was false, because the proposition that constituted its content is (was) false.

Given materialism, therefore, beliefs would be long-standing neural events. As such, they would have content, but also neurophysiological properties (NP properties). Now how is it that we human beings have come to have beliefs, and how is it that those beliefs come to have the content they do in fact have? Naturalists (and of course others as well) ordinarily believe that human beings have come to be by way of evolution; they have evolved according to the mechanisms specified in contemporary evolutionary theory. (The prime candidates are natural selection operating on some source of genetic variability such as random genetic mutation.) We have something of an idea as to the history of those neurophysiological properties: structures with these properties have come to exist by small increments, each increment (ignoring spandrels and pliotropy) such that it has proved to be useful in the struggle for survival. But what about the *content* of belief? If a belief is a neuronal event, where does its content come from? How does it get to be associated in that way with a given proposition?

Materialists offer two main theories here. According to the first, content *supervenes upon* NP properties; according to the second content *is reducible to* NP properties.[3] Suppose we think about the second theory first. Consider the property of having as content the proposition *Naturalism is all the rage these days*, and call this property 'C'. On the present suggestion, C *just is* a certain combination of NP properties. It might be a disjunction of such

properties: where P_1 to P_n are NP properties, C, the property of having the content in question, might be something like (where 'v' represents 'or')

$$P_1 v P_3 v P_8 v \ldots P_n.$$

More likely, it would be something more complicated: perhaps a disjunction of conjunctions, something like (where '&' represents 'and')

$$(P_1 \& P_7 \& P_{28} \ldots) \text{ v } (P_3 \& P_{34} \& P_{17} \& \ldots) \text{ v } (P_8 \& P_{83} \& P_{107} \& \ldots) \text{ v. } \ldots$$

We could put this by saying that any content property is a Boolean combination of NP properties, that is, a combination constructed from NP properties by disjunction, conjunction and negation. And to say that content properties are reducible to NP properties is just to say that every content property is some Boolean combination of NP properties. In fact, if we think that any Boolean combination of NP properties is itself an NP property, we could say that content properties just are NP properties – a special sort of NP property, to be sure, but still NP properties. So, on this theory, content properties – e.g., the property of having *Naturalism is all the rage these days* as content – are or are reducible to NP properties.

That's one of the two materialistic proposals; the other is that a content property isn't a NP property, or a Boolean combination of NP properties, but rather *supervenes on* NP properties. What does that mean; what is this 'supervention'? The basic idea is that a set of properties S supervenes on a set of properties S* just if any pair of objects which agree on the S* properties must also agree on the S properties. For example, beauty (of a picture, a face) supervenes on molecular constitution; any two pictures (or faces) with the same molecular constitution will be beautiful to the same degree. Content properties supervene on NP properties, then, if and only if any two objects or structures with the same NP properties must have the same content properties. You couldn't have a pair of structures – neuronal events, say – that had the same NP properties but different contents.[4] Content is a *function* of NP properties.

We can put this officially as follows:

> (S) Necessarily, any structures that have the same NP properties have the same content.

This is a *weak* form of supervenience; a stronger one could be put as:

> (S+) For any possible worlds W and W* and any structures S and S*, if S has the same NP properties in W as S* has in W*, then S has the same content in W as S* has in W*.

Those who think that content properties supervene on NP properties for the most part think, I believe, that the former supervene on the latter in the stronger

sense (S+) (and hence also, of course, in the weaker sense (S)). For present purposes, however, it doesn't matter which sense we employ. But what about that 'necessarily'? Here this supervention suggestion divides into two branches. On the first branch, the necessity in question is broadly logical necessity. According to the other branch of the supervenience theory, the necessity in question isn't broadly logical necessity, but something more obscure – something we could call 'causal' or 'natural' or 'nomic' necessity.

Naturalism and reliability

Return now to the question that led us into reduction and supervenience: how does it happen that those neural structures, the ones that constitute belief, have *content*? Where does it come from and how do they get it? The basic idea is something like this. As we go up the evolutionary scale, we find neural structures with greater and greater complexity. Near one end of the scale, for example, we find C. *elegans*, a small but charismatic worm with a nervous system composed of only a few neurons. (The nervous system of C. *elegans* has been completely mapped.) We human beings are at the other end of the scale; our brains contain many billions of neurons connected in complex and multifarious ways. And now the idea is that as you rise in the evolutionary scale, as you progress through more and more complex neural structures, at a certain point content shows up. At a certain level of complexity, these neural structures start to display content. Perhaps this starts gradually and early on (possibly C. *elegans* displays just the merest glimmer of consciousness and the merest glimmer of content), or perhaps later and more abruptly; that doesn't matter. What does matter is that at a certain level of complexity of neural structures, content appears. This is true whether content properties are reducible to NP properties or supervene on them.

So (given materialism), some neural structures at a certain level of complexity acquire content; they thus become beliefs. And the question I want to ask is this: what is the likelihood, given naturalism, that the content that thus arises is in fact *true*? In particular, what is the likelihood, given N, that the content associated with *our* neural structures is true? More generally, what is the likelihood, given naturalism, that our cognitive faculties are reliable, thereby producing mostly true beliefs?

We commonsensically assume that our cognitive faculties are for the most part reliable, at least over a large area of their functioning. I remember where I was last night and that my elder son's name is not Archibald; I can see that the light is on in my study, that the flower garden is overgrown with weeds, and that my neighbor put on weight over the winter. I know a few truths of mathematics and logic, mostly pretty simple, no doubt, but still ... The natural thing to assume, and what we all do assume (at least before we are corrupted by philosophy (or neuroscience)) is that when our cognitive faculties aren't subject to malfunction, then, for the most part and

over a wide area of everyday life, the beliefs they produce in us are true. We assume that our cognitive faculties are reliable. But what I want to argue is that the naturalist has a powerful reason against this initial presumption and should give it up.

By way of entering this argument, suppose we conduct a thought experiment. Consider a hypothetical species that is cognitively a lot like us: members of this species hold beliefs, make inferences, change beliefs, and the like. And let us suppose naturalism holds for them; they exist in a world in which there is no such person as God or anything like God. Our question, then, is this: what is the probability that their cognitive faculties are reliable? Consider any particular belief on the part of one of these hypothetical creatures. That belief, of course, is a neural structure of a given sort, and one sufficiently complex to generate content. We may add, if we like, that this structure occurs or takes place in response to something in the environment; perhaps it is a certain pattern of firing of neurons in the optical portion of the brain, and perhaps this pattern arises in response to the appearance of a predator in the middle distance. And a certain proposition has somehow come to be associated with this structure, so that the structure acquires belief content and is a belief.

Now what is the probability (given naturalism) that this proposition is *true*? Well, what we know about the belief in question is that it is a neurological structure that has certain NP properties, properties, the possession of which is logically or causally sufficient for the possession of that particular content. We are assuming also that this structure arises in response to the presence of that predator, and we can also assume, if we like, that this structure is a reliable indicator of that kind of predator. This structure, we may suppose, arises when and only when there is a predator in the middle distance. But even so, of course, the content generated by this structure, on this occasion, need have nothing to do with that predator, or with anything else in the environment. Indication is one thing; belief content is something else altogether, and we know of no reason why the one should be related to the other. By way of something like a necessary accident, content simply arises upon the appearance of neural structures of sufficient complexity. But we can see no reason why that content need be related to what the structures indicate, if anything. The proposition constituting that content need not be so much as *about* that predator.

So what, then, is the likelihood that this proposition, this content, is true? Given this much, shouldn't we suppose that the proposition in question is as likely to be false as true? Shouldn't we suppose that the proposition in question has a probability of roughly one-half of being true? Shouldn't we estimate its probability, on the condition in question, as in the neighborhood of 0.5? That would be the sensible course. Neither seems more probable than the other; hence we estimate the probability of its being true as 0.5.

But am I not relying upon the notorious Principle of Indifference? We are trying to estimate the probability that the content in question is true, given

that it is generated by adaptive neural structures; I say that given this condition, for all we can see, it is as likely to be false as to be true; so we should judge that probability to be around 0.5; isn't that to endorse some version of the Principle of Indifference? And hasn't that principle been discredited?[5] Not really. The Bertrand paradoxes show that certain incautious statements of PI come to grief – just as Goodman's grue/bleen paradoxes show that incautious statements of a principle governing the projection of predicates or properties come to grief. But of course the fact is we project properties all the time, and do so perfectly sensibly. In the same way, I think, we often employ a principle of indifference in ordinary reasoning, and do so quite properly. We also use it in science, for example, in statistical mechanics.[6] Of course, problems arise where there are equally natural or plausible ways of selecting the relevant possibilities and where these different ways carry incompatible probability assignments with them.

But aren't we forgetting something important? These hypothetical creatures have arisen, presumably, by way of evolution. They have come to be by way of something like natural selection working on some process of genetic variation – perhaps random genetic mutation. Presumably, then, it has proven adaptively useful for creatures of that sort to display that neural structure in the circumstances in which this creature finds itself. This structure's arising in those circumstances has (or had) survival value; it contributes to the reproductive fitness of the creature in question, perhaps by helping cause the right sort of behavior (fleeing, or wary watchfulness, maybe). Whatever exactly the appropriate action is, the neuronal event in question is useful because it is a cause (part cause) of that behavior. And doesn't that mean that it's likely that the content associated with this structure is in fact a true proposition?

It is crucially important to see that the answer to this question is NO. This neuronal event or structure has NP properties such as sending electrical signals to other parts of the nervous system as well as to muscles and/or glands. By virtue of these NP properties, it causes adaptive behavior such as fleeing. This neuronal structure also displays NP properties that are sufficient, causally or logically, for the presence of content. As a result of having that neuronal event with that particular constellation of NP properties, the creature in which this event is to be found also believes a certain proposition. But what reason is there to think that proposition *true*? Granted, the structure in question helps cause adaptive behavior. But that doesn't so much as slyly suggest that the content that gets associated with the structure – by way of a logical accident, so to say – is *true*. As far as its causing the right kind of behavior is concerned, it simply doesn't matter whether the content, that associated proposition, is true or false. At this point, as far as the truth or falsehood of the content that arises, natural selection just has to take pot luck. (Not that it minds – it's interested, so to speak, just in adaptive behavior, not in true belief.) Natural selection selects for structures that have adaptive NP properties; as it happens, these structures

are of sufficient complexity to generate content; but there isn't even the faintest reason to think that content true. Given naturalism, it would be sheer coincidence, an enormous cosmic serendipity, if the content that is associated with adaptively useful neurophysiological properties should also turn out to be all or mostly true content. Naturalists who think content supervenes on neurophysiological properties (and that would be most naturalists) tend to assume automatically (at least when it comes to us human beings) that the content in question *would be* true; but why think that? This assumption is at best a piece of charming but ingenuous piety. Given naturalism, the belief in question is as likely to be false as to be true.

So, with respect to the relevant facts about the origin and provenance of this particular belief on the part of this hypothetical creature, the probability of its being true – i.e., the probability that the content of the neural structure in question should be a true proposition – would have to be estimated as about 0.5. The associated content in question could, of course, be true; but it could also, and with equal likelihood, be false. What, then, is the probability that the cognitive faculties of these creatures will be *reliable*? A reliable belief-producing faculty will produce a considerable preponderance of true belief over false belief. We ordinarily think our cognitive faculties are more reliable in some circumstances than in others; we are good at such things as remembering what we had for breakfast or perceiving whether there are any trees in the backyard; we are less good at determining (without artificial aids) whether a mountain goat we see at 500 yards has horns. We are also less reliable when working at the limits of our faculties, as in trying to determine what happened in the first 10^{-33} seconds after the Big Bang. (Given all the disagreements, perhaps we are also less reliable when it comes to philosophy.) But any reasonable degree of reliability, as we ordinarily think of it, requires producing a substantial preponderance of true beliefs. A thermometer that didn't produce many more true than false readings (in normal circumstances and within the appropriate limits of error) would not be reliable.

And the same sort of thing goes for the reliability of cognitive faculties; they too are reliable, and reliable in a certain area, only if they produce a preponderance of true beliefs over false. Going back to those hypothetical creatures, what we've seen is that the probability, on the relevant condition, that any given belief of theirs should be true is in the neighborhood of 1/2. This means that the probability that their faculties produce the preponderance of true beliefs over false required by reliability is very small indeed. If I have 1000 independent[7] beliefs, for example, the probability (under these conditions) that three-quarters or more of these beliefs are true (certainly a modest enough requirement for reliability) will be less than 10^{-58}.[8] And even if I am running a modest epistemic establishment of only 100 beliefs, the probability that 3/4 of them are true, given that the probability of any one's being true is 1/2, is very low, something like 0.000001. So the chances that this creature's true beliefs substantially outnumber its

false beliefs (even in a particular area) are small. The conclusion to be drawn is that it is very unlikely that the cognitive faculties of those creatures are reliable.

So far what we've seen is that, given naturalism and the supervenience of content upon neurophysiological properties, it is unlikely that the cognitive faculties of these creatures are reliable; this is true even if we add that the content of their beliefs is generated by structures with NP properties that are fitness-enhancing, adaptively useful.

That's how things stand if content *supervenes* upon NP properties. But what about the other option, reductionism? What if content properties (for example, the property of having as content the proposition *Naturalism is all the rage these days*) just *are* NP properties, or complex clusters of NP properties? In this case we get the very same results. To see why, consider, again, a given belief on the part of a given member of that hypothetical group of creatures. That belief, of course, is a neuronal event, a congeries of neurons connected in complex ways and firing away in the fashion neurons are wont to do. This neuronal event displays a lot of NP properties. Again, we may suppose that it is adaptively useful for a creature of the kind in question to harbor neuronal structures of the sort in question in the circumstances in question. The event's having the NP properties it does have is fitness-enhancing in that by virtue of having these properties, the organism is caused to perform adaptively useful action – fleeing, for example. But some subset of these NP properties together constitute its having a certain content, constitute its being associated, in that way, with some proposition. What is the probability that this content is true? What is the probability that the associated proposition is a true proposition? The answer is the same as in the case we've already considered. The content doesn't have to be true, of course, for the neuronal structure to cause the appropriate kind of behavior. It just happens that this particular arrangement of adaptive NP properties also constitutes having this particular content. But again: it would be a piece of enormous serendipity if this content, this proposition, were *true*; it could just as well be false. So the probability that this content is true would have to be rated at about 1/2, just as in the case of supervenience. If this is true for each of the independent beliefs of the organism in question, the probability (on naturalism) that the cognitive faculties of these creatures are reliable would have to be rated as low. The conclusion to be drawn so far, then, is that given naturalism, it is unlikely that these creatures have reliable cognitive faculties.

Now the next step in the argument is to note that of course what goes for these hypothetical creatures also goes for us. Suppose naturalism (construed as including materialism) is in fact true with respect to us human beings: there is no such person as God or anything like God. Then the probability that our cognitive faculties are reliable is low, just as in the case of those hypothetical creatures. For us, too, the main possibilities would have to be supervenience (logical or causal) and reduction or identity. In our case, too,

if we focus on any particular belief – say, the belief that naturalism is all the rage these days – on the part of a particular believer, we see that this belief (given materialism) will have to be a neuronal event of some kind. This event will be of sufficient complexity to generate content (by supervenience or reduction); somehow a proposition gets associated with it as its content. We may suppose, if we wish, that it is adaptively useful for creatures like us to harbor structures of that kind in the circumstances in which the believer finds herself. It would be the merest coincidence, however, if the content generated by the structure in question should be *true* content, if the proposition which is the content of the belief in question should turn out to be a *true* proposition. That means that the probability of this belief's being true would have to be judged to be in the neighborhood of 1/2, not much more likely to be true than to be false. But then it will be exceedingly improbable that the whole set of this believer's beliefs should display the preponderance of true belief over false required by the reliability of her cognitive faculties. So our case is like that of those hypothetical creatures; in our case too the probability that our cognitive faculties are reliable, 'P(R/N)', is low.

Naturalism and defeat

But now let's take one more step: a person who accepts naturalism and recognizes that P(R/N) is low, thereby acquires a *defeater* for R. A defeater[9] for a belief B I hold – at any rate this kind of defeater – is another belief B* I come to hold which is such that given that I hold B*, I can no longer rationally hold B. For example, I look into a field and see what I take to be a sheep. You come along, identify yourself as the owner of the field, and tell me that there aren't any sheep in that field and that what I see is really a dog that at this distance is indistinguishable from a sheep. Then I give up the belief that what I see is a sheep. Another example: on the basis of what the guidebook says I form the belief that the University of Aberdeen was established in 1695. You, the university's public relations director, tell me the embarrassing truth: this guidebook is notorious for giving the wrong date for the foundation of the university. (It was actually established in 1595.) My new belief that the university was established in 1595 is a defeater for my old belief. In the same way, if I accept naturalism and see that P(R/N) is low, then I have a defeater for R; I can no longer rationally believe that my cognitive faculties are reliable.

The problem isn't that I don't have enough *evidence* for R, to believe it rationally. The fact is I don't *need* evidence for R. That's a good thing, because it doesn't seem possible to acquire evidence for it, at least if I have any doubts about it. For suppose I think up some argument for R, and on the basis of this argument come to believe that R is indeed true. Clearly this is not a sensible procedure; to become convinced of R on the basis of that argument, I must of course believe the premises of the argument, and

also believe that if those premises are true, then so is the conclusion. But if I do that, I am already assuming R to be true, at least for the faculties or belief-producing processes that produce in me belief in the premises of the argument and belief that if the premises are true, so is the conclusion. As the great Scottish philosopher Thomas Reid says,

> If a man's honesty were called into question, it would be ridiculous to refer to the man's own word, whether he be honest or not. The same absurdity there is in attempting to prove, by any kind of reasoning, probable or demonstrative, that our reason is not fallacious, since the very point in question is, whether reasoning may be trusted.[10]

My accepting any argument for R, or any evidence for it, would clearly presuppose my believing R; any such procedure would therefore be viciously circular.

More important, however, is the following. We all naturally assume R, and assume it from our earliest days as cognitive agents. Now rationality is best explained in terms of proper function: a belief is rational, in a given set of circumstances, just if a rational person, one whose cognitive faculties are functioning properly, could hold that belief in those circumstances.[11] But then clearly, it is perfectly rational to assume, without evidence, that your cognitive faculties are functioning reliably. We rational agents do this all the time, and do not thereby display cognitive malfunction. You might wind up in a care facility for believing that you are *Napoleon*, but not for believing that your cognitive faculties are functioning reliably. It is therefore perfectly rational to believe R, and to believe it in the basic way, i.e., not on the basis of propositional evidence.

But that doesn't mean that it is not possible to acquire a defeater for R; even if a belief is properly basic it is still possible to acquire a defeater for it. In the above example about the sheep in the field, my original belief, we may suppose, was basic, and properly so; I still acquired a defeater for it. Here is another famous example to show the same thing. You and I are driving through southern Wisconsin; I see what looks like a fine barn and form the belief *Now that's a fine barn!* Furthermore, I hold that belief in the basic way; I don't accept it on the basis of evidence from other propositions I believe. You then tell me that the whole area is full of barn facades (indistinguishable, from the highway, from real barns) erected by the local inhabitants in an effort to make themselves look more prosperous than they really are. If I believe you, I then have a defeater for my belief that what I saw was a fine barn, even though I was rational in holding the defeated belief in the basic way. It is therefore perfectly possible to acquire a defeater for a belief B even when it is rational to hold B in the basic way. This is what happens when I believe naturalism, and come to see that P(R/N) is low: I acquire a defeater for R. I can then no longer rationally accept R; I must be agnostic about it, or believe its denial.

Perhaps we can see more clearly here by considering an analogy. Imagine a drug – call it XX – that destroys your cognitive reliability. Some 95 percent of those who ingest XX become cognitively unreliable within two hours of ingesting it; they then believe mostly false propositions. Suppose further that I now believe both that I've ingested XX a couple of hours ago and that

P(R/I've ingested XX a couple of hours ago)

is low; taken together, these two beliefs give me a defeater for my initial belief that my cognitive faculties are reliable. Furthermore, I can't appeal to any of my other beliefs to show or argue that my cognitive faculties are still reliable. For example, I can't appeal to my belief that my cognitive faculties have always been reliable in the past or seem to me to be reliable now; any such other belief is now just as suspect or compromised as R is. Any such other belief B is a product of my cognitive faculties: but then in recognizing this and having a defeater for R, I also have a defeater for B.

Two final matters. First, perhaps you believe the thing to think about P (R/N) is not that it is low, but that it is *inscrutable*. How, you ask, can we possibly tell what that probability would be? Return to the question of the probability that a belief is true, conditional on N and its supervening on or being reducible to adaptive NP properties. There I said that this probability should be thought of as in the neighborhood of 1/2 (in which case it would be unlikely *in excelsis* that the creature's true beliefs should exceed its false with a preponderance sufficient for its cognitive faculties being reliable). But maybe the right answer is that we just can't tell what that probability is: it's inscrutable.

There may be something to this objection. But all the argument as stated really requires is that the probability in question not be very high; that it isn't very high seems clear enough. Suppose, however, that this probability really is completely inscrutable: we haven't the faintest idea what it is. As far as we can tell, it could be as high as 1; it could also be zero; and it could be anything in between. We still get the same result. If this probability is inscrutable, then so will be P(R/N); but *N&P(R/N) is inscrutable* is a defeater for R, just as is *N&P(R/N) is low*. Consider an analogy. You learn that your cousin Sam, whose cognitive faculties you have always assumed to be reliable, has ingested XX. You know that *some* proportion of those who ingest XX become wholly unreliable; but you don't know what that proportion is; as far as you are concerned, P(Sam's faculties are reliable/Sam has ingested XX) is inscrutable. It could be as low as zero; it could be as high as 1; and it could be anything in between. Under these conditions you have a defeater for your assumption that Sam's cognitive faculties are reliable. You would also have a defeater for R if you believed you had ingested XX and that P (R/I've ingested XX) is inscrutable. So what the argument really requires is only that P(R/N) be low or inscrutable.[12]

Finally, there is one more wrinkle, or perhaps fly in the ointment.[13] Consider someone who is cognitively normal, and who comes to believe that she has ingested XX, that reliability-destroying drug mentioned above. This person may very well continue to assume that her cognitive faculties are functioning properly. She may very well carry on her cognitive life in the usual way, even if she becomes convinced she's contracted mad cow disease, a disease, as she believes, that renders its victims cognitively unreliable. And of course the same goes (in spades) if she believes N and sees that P(R/N) is low. But (and this is the crucial point) in so doing, might she not be functioning perfectly properly, without so much as a hint of dysfunction or malfunction? The answer certainly seems to be Yes. If so, however, then given my account of defeat (in terms of proper function) she doesn't have a defeater for R in the belief that she has ingested XX or has contracted mad cow disease, and my argument fails.

Here I can only gesture at the response.[14] The first thing to see is that one who really rejects R is in a state of cognitive disaster. And some modules of our cognitive design plan are aimed, not at the production of true beliefs, but at the production of other worthwhile conditions, including, presumably, avoidance of cognitive disaster. For example, if you fall victim to a usually fatal disease, you may somehow think your chances are much better than is indicated by the statistics you know; this is the so-called 'optimistic overrider'. Your faculties may be functioning perfectly properly in producing this belief; this particular bit of the cognitive design plan is aimed, not at producing true beliefs about the possible course of your disease, but beliefs that will maximize your chances of recovery. Still, in some sense those statistics really do give you a defeater for your belief that in all likelihood you will recover. What they give you is a *Humean Defeater*. You have a Humean defeater for a belief B in a given situation if: (1) the production of B is governed by a bit of the design plan that is aimed, not at the production of true belief, but at some other state of affairs (such as recovery from disease or the avoidance of cognitive disaster); and (2) if only truth-aimed processes were at work in this situation, you would have an ordinary rationality defeater for B. One who believes she's taken XX has a Humean defeater for R, as does someone who thinks she has mad cow disease. My claim is that the naturalist who sees that P(R/N) is low has a Humean defeater for R.

I therefore have a defeater for R. But if I consider R and do not believe it, then I have a defeater for any belief I take to be a product of my cognitive faculties. Naturally enough, that would be *all* of my beliefs; all of my beliefs are products of my cognitive faculties. The result so far, then, is that if I believe N (construed as including materialism) and I also see that the probability of R with respect to N is low, then I have a defeater for each of my beliefs. Therefore, if you believe N and see that P(R/N) is low, you will be enmeshed in that virulent, bottomless, self-reflexive sort of Humean skepticism mentioned above. No doubt you can't really reject R in the heat

and press of day-to-day activities, when you are playing poker with your friends, or building a house, or climbing a cliff. But in the calm and reflective atmosphere of your study, you see that you do in fact have a defeater for R. Of course, you also see that the very reflections that lead you to this position are also no more acceptable than their denials; you have a universal defeater for whatever it is you find yourself believing. This is that really crushing skepticism, and it is this skepticism to which the naturalist is committed.

Notes

1 And this leads to the scandal of skepticism: if I *argue* to skepticism, then of course I rely upon the very cognitive faculties whose unreliability is the conclusion of my skeptical argument.
2 It is of course extremely difficult to see how a material structure or event could have content in the way a belief does.
3 Note that if content properties are reducible to NP properties in the sense of 'reducible' suggested below, then they also supervene upon them. Note also that for present purposes I ignore so-called 'wide content'. If we were to take wide content into account, we'd say that content supervenes, not just on NP properties, but on NP properties together with certain properties of the environment. The same would go, *mutatis mutandis*, for the suggestion that content is reducible to or identical with NP properties. In the interest of simplicity, I ignore wide content; nothing in my argument below hinges on this omission.
4 So the second possibility is really a special case of the first: if content properties are reducible to NP properties, then clearly structures with the same NP properties will have the same content properties.
5 See, e.g., van Fraassen (1989: 293ff).
6 According to Weatherford: 'an astonishing number of extremely complex problems in probability theory have been solved, and usefully so, by calculation based entirely on the assumption of equiprobable alternatives' (1983: 35). See also Collins (1998).
7 'Independent': it could be that a pair of neural structures with content were such that if either occurred, so would the other; then the beliefs in question would not be independent. Similarly when the content of one neural structure entails the content of another: there too the beliefs in question won't be independent.
8 My thanks to Paul Zwier, who performed the calculation.
9 Of course there are several kinds of defeaters; here it isn't necessary to canvass these kinds. The kind of defeater presently relevant would be a *rationality* defeater, and an *undercutting* rationality defeater. In addition to rationality defeaters, there are also *warrant* defeaters; these too come in several kinds. For more on defeaters, see Bergmann (2000, 1997), and see Plantinga (2002: 205–211).
10 Reid (1983: 276).
11 See Plantinga (1993a: 133–137).
12 The first clause of (D) should thus be amended to '(1) S sees that P(A/B) is low or inscrutable'.
13 As William Talbott pointed out to me.
14 For a fuller version of the response, see Plantinga (2002: 205–211).

Methodological and/or ontological naturalism

Comment on Plantinga's paper

Roberta Corvi

Plantinga's paper is very rich, subtle and deep and in it there are many points worthy of discussion. I shall therefore confine myself to a few points that set up the foundations of his complex and teeming reflection.

First, I would like to qualify more precisely the notion of naturalism in the sense used by Plantinga. Then I shall dwell upon the relationship between naturalism and 'true' skepticism. Finally, I shall touch upon the problem of the truth of our beliefs and of the warrant they offer.

Plantinga uses a very clear-cut and strict interpretation of naturalism identifying this philosophical stance with the concept according to which, at least implicitly, God or any supernatural entity does not exist. Yet, philosophical literature on the matter testifies to a wide constellation of forms taken by contemporary naturalism (Goldman 1994; Koppelberg 1999). Dirk Koppelberg's analysis seems to provide an interesting point of view on the subject. According to him, we can distinguish three basic varieties of naturalism: (1) metaphysical or ontological naturalism; (2) analytical naturalism; and (3) methodological naturalism.

As far as contemporary ontological naturalism is concerned:

> the starting-point of contemporary ontological naturalism is Wilfrid Sellars' famous dictum: 'in the dimension of describing and explaining the world, science is the measure of all things, of what is that it is, and of what is not that it is not'. For the naturalist there is no extra-scientific way to metaphysical insight and understanding. So the business of ontological naturalism is to draw out and analyse the metaphysical commitments and implications of scientific theories.
>
> (Koppelberg 1999: 32)

Analytical or conceptual naturalism 'maintains that a philosophical doctrine is naturalistic if its key terms are analysed by means of some privileged set of terms' (ibid.: 33), which can be drawn either from the vocabulary of physics alone or also include terms from biology, psychology and even social sciences.

The last sort of naturalism goes back to the famous essay *Epistemology Naturalized*, published by Quine in 1969 and consists of the combination of

three theses: (1) 'philosophy does not found or ground the sciences'; (2) 'philosophy has no epistemically privileged status among the sciences, rather it is continuous with them'; (3) 'the use of scientific investigations and results is relevant and indispensable for philosophy' (Koppelberg 1999: 35).

I will set aside analytical naturalism, since it only refers to the vocabulary used to express a theory and it is therefore a philosophical criterion to evaluate if a doctrine is naturalistically acceptable. The question is now to establish if there can be methodological naturalism, without metaphysical implications.

Plantinga himself criticized 'methodological' naturalism, which holds that

> for any study of the world to qualify as 'scientific', it cannot refer to God's creative activity (or any sort of divine activity). The methods of science, it is claimed, 'give us no purchase' on theological propositions – even if the latter are true – and theology therefore cannot influence scientific explanation or theory justification.
>
> (Plantinga 1996a)

In other words, science is regarded as religiously neutral and, as Bas van Fraassen maintains, absolutely indifferent to any metaphysics, because 'there is no non-empirical claim which matters at all to the process of science' (van Fraassen 1996: 175).[1] However, in Plantinga's opinion, 'scientific theories are often not religiously or metaphysically neutral' (Plantinga 1996a).

Elsewhere (Plantinga 1993b: 45 ff.), Plantinga defended only epistemological naturalism, consistent with a 'supernaturalistic setting'.[2] So my question is if, in Plantinga's opinion, there can be a sort of naturalism, which turns out to be only epistemological and if this naturalism may be rightly regarded as a methodological naturalism of a quinean kind. If not, we ought to infer, like Earl Conee, that naturalism inevitably implies a metaphysical attitude, which leads us to deny God's existence. For, 'if naturalistic epistemology is what it appears to be – an application of a naturalistic world view and methodology to epistemic topics – then it is incompatible with any sort of supernaturalism' (Conee 1996: 184).

However – and I come to my second point – it is in fact a form of naturalism that has important outcomes on the epistemological ground, as just mentioned: naturalism implies skepticism, a topic examined by Plantinga in his thorough and rigorous study.

Now, the very connection between naturalism and skepticism reminds me of Strawson's essay entitled *Skepticism and Naturalism*, where the author states an opposite point of view, i.e. that naturalism saves us from skepticism. Strawson quotes the *Treatise of Human Nature*, where Hume as a naturalist disarms Hume the skeptic.

In his essay, Strawson distinguishes two kinds of naturalism: 'reductive (or strict) naturalism' and 'nonreductive naturalism'. His first definition of

naturalism views human beings and human actions 'simply as objects and events in nature, natural objects and natural events, to be described, analysed, and causally explained in terms in which moral evaluation has no place' (Strawson 1985: 40). Instead, the second kind of naturalism, also called 'liberal' or 'catholic', admits that:

> our natural disposition to belief, on the points challenged by the skeptic, is absolutely compelling and inescapable; ... Where Nature determines us, we have an original non-rational commitment which sets the bounds within which, or the stage upon which, reason can effectively operate.
>
> (ibid.: 39)

Really none of the two definitions agrees exactly with the meaning used by Plantinga, but the naturalism he criticizes resembles the reductive kind defined by Strawson, since it coincides with materialism. It must be added that, even if Strawson sympathizes with non-reductive naturalism, he thinks that both – indifferently – have anti-skeptic potential (ibid.: 94, 3).

Strawson once more presents Hume's argument, which in his opinion, was continued and modified by Wittgenstein in *On Certainty*. Both for Hume and for Wittgenstein, skeptical doubts are '*idle*, powerless against the force of nature, of our naturally implanted disposition to belief' (Strawson 1985: 13).

In particular, according to Hume, the existence of body and the general reliability of induction are convictions ineradicably impressed by nature in our minds, free from doubt, even if they are not grounded beliefs. We have just heard Plantinga's opinion about this kind of 'reflexive' skepticism, which is not the one implicit in naturalism. For, 'true' skepticism 'has a defeater for each of his beliefs, including those that lead to skepticism' (Plantinga, p. 32).

Wittgenstein's position is different, because he, as Strawson himself acknowledges, never appeals to nature unequivocally, even though he confirms the distinction between what 'we must take for granted in all our reasoning' and what is really matter for inquiry. About this point Wittgenstein's thought is extremely clear: 'I have a world-picture: is it true or false? Above all it is the substratum of all my enquiring and asserting' (Wittgenstein 1974: § 162).[3] Therefore, 'the *questions* that we raise and our *doubts* depend on the fact that some propositions are exempt from doubt, are as it were like hinges on which those turn' (ibid.: § 341).[4] To sum up, 'at the foundation of well-founded belief lies belief that is not founded' (ibid.: § 253).

On one side, there are those elements of our system of beliefs which are subject to empirical control, on the other, there are the scaffolding, the structure, the background which support both our convictions and our doubts – a substratum which is surely natural and even 'animal'[5] but derives from social practice, too. Indeed,

> I did not get my picture of the world by satisfying myself of its correctness; nor do I have it because I am satisfied of its correctness. No: it is the inherited background against which I distinguish between true and false.
>
> (ibid.: § 94)

I do not intend to discuss here if the interpretation that Strawson suggests of Wittgenstein's thought which he calls 'social naturalism' is or is not correct.[6] Rather I would like to ask if, according to Plantinga's close examination, the philosophical analysis yielded by Wittgenstein might be, as it seems to me, useful to face the issue of the reliability of our beliefs or if, on the contrary, Wittgenstein's position might be a form of 'true' skepticism, because it has a defeater for beliefs that lead to skepticism too. For, Wittgenstein assumes and accepts that there are no grounded reasons to think the contents of our beliefs are reliable, but, in spite of this conviction, also claims we cannot help having beliefs.

This remark leads me to my last point concerning the controversial issue of the truth of our beliefs and of the warrant they have. If we try to discuss the issue in naturalistic terms, we have to ask how the naturalist understands truth. Is there place for truth as a normative idea? If there is, could a human being as a natural entity grasp it?

I think that we have to answer no to the first question; otherwise, the acknowledgement of a truth independent from human beings would imply an overcoming of naturalism, according to Plantinga's definition, and it would get nearer to the hegelian model, which Plantinga has just showed not to be naturalistic. Therefore, in such a framework, asking if human beings can grasp something that does not exist, would not make sense.

Therefore, if according to the naturalist there is not a truth to which our knowledge has to conform, why should he worry about the truth of the propositions which express the contents of our beliefs?

Another variety of the naturalistic attitude might be Keith Lehrer's, which we find in the book published in honour of Plantinga. Lehrer admits that 'the way we function in forming beliefs is the result of the evolutionary process fails to support the conclusion that those beliefs are true'. Then he points out: 'Plantinga proceeds by agreeing that evolution does not yield the conclusion that our beliefs are true, assuming our faculties have arisen from an evolutionary process, to the conclusion that they are not likely to be true' (Lehrer 1996: 27). According to Lehrer, it is not appropriate to conclude that our beliefs are not likely to be true from the fact that we have no reasons to state that our beliefs are true. As we have just heard, Plantinga replies that the result is the same (Plantinga, p. 32). Yet, the naturalist might answer that the evolutionary theory is not able to exclude that human beings have retained features not necessary to survival. Or, in other words, truth could not be a value on the biological level, in which natural evolution is placed.

The naturalist might take another path and try to naturalize the concept of truth, as well. On that account, he might claim that we call true those beliefs which prove useful for survival, for the improvement of the species and, generally, of life conditions owing to their adaptive value (Ramsey 2002: 16–19). In this case, the naturalist does not oppose Plantinga's thesis, rather, he denies the premise on which it is founded.

I shall come back to this point after I make a final remark on Plantinga's thought: even if we admit that naturalism implies skepticism, how does our non-naturalistic choice save us from skepticism?

If I did not misunderstand the thought Plantinga has only mentioned here, but developed widely in his doctrine of proper function (Plantinga 1993b: Chapters 1, 2, 11 *passim*), the existence of God and of his plan explains the proper function of human cognitive faculties, when they work in a proper cognitive environment. Indeed, God has created us human beings 'in his own image and, since he is an intellectual being who has knowledge, we resemble him also in our cognitive faculties'.[7] In this perspective, our faculties work properly, when they are working in the way thought by the being who designed and created them. So, since our knowledge and our beliefs are generally reliable, we have to admit that God exists like a warrantor of their reliability, but the skeptic, naturalist or not, might remark again that the reliability of our beliefs has not been demonstrated.

Summarizing, it seems to me that Plantinga's theory is founded upon a premise that is not undeniable. In my opinion, this is no scandal, for, as Aristotle said,[8] it is impossible to be able to demonstrate everything. Yet, I think that the choice of the unfounded premises is problematic, because there are a lot of starting points, as the history of philosophy teaches us. For example, Popper referred to 'an irrational faith in reason'; we, as believers, can place our trust in God, as Plantinga suggests; the naturalist, who denies the supernatural world, can refuse the premise according to which the truth is to be intended in a way that does not coincide with biological utility.

To conclude, I think that the issue is about the principles by means of which we choose or acknowledge not demonstrated premises of our demonstrations, all the more so as, according to Wittgenstein,[9] this basic framework is not unchangeable and absolute but varying and unsettled.

Acknowledgements

I want to thank Professor Margherita Giulietti, who helped me with my English and to avoid a lot of mistakes. It goes without saying that I alone am responsible for the final draft of this paper.

Notes

1 In a footnote he adds: 'the *theses* in question [beliefs about God] do not affect science; if we consider *only* the content of these theses, they play no role at all', although they might 'trigger wishes, intentions and eventually actions' (van Fraassen 1996: 180).

2 See Plantinga (1993b: 194):

> the account of warrant I propose is an example of *naturalistic* epistemology: it invokes no kind of normativity not to be found in natural sciences; the only kind of normativity it invokes figures in such sciences as biology and psychology.

See also Plantinga (1996b: 352–357).

3 See also Wittgenstein (1974: § 415): 'certain propositions seem to underlie all questions and all thinking'.

4 In the following section he adds: 'it belongs to the logic of our scientific investigations that certain things are *in deed* not doubted' (ibid.: § 342).

5 See Wittgenstein (1974: § 356): there is a '*comfortable* certainty', when 'I should not understand where a doubt could get a foothold nor where a further test was possible' and 'I want to conceive it as something that lies beyond being justified or unjustified; as it were, as something animal' (ibid.: § 359). See also ibid. § 357.

6 As far as this matter is concerned, see Coliva (1999: 169–187).

7 Plantinga claims that:

> natural organisms have indeed been designed by a conscious and intentional designer: God . . . In setting out to create *rational* creatures: creatures with reason or *ratio*; creatures that reflect his capacity to grasp concepts, entertain propositions, hold beliefs, envisage ends, and act to accomplish them. Furthermore, he proposed to create creatures who reflect his ability to hold *true* beliefs. He therefore created us with that astonishingly subtle and articulate battery of cognitive faculties and powers discussed in the preceding chapters.
>
> (1993b: 197)

8 According to Aristotle:

> Not to know of what things one should demand demonstration, and of what one should not, argues want of education. For it is impossible that there should be demonstration of absolutely everything (there would be an infinite regress, so that there would still be no demonstration).
>
> (*Metaphysics* G, 4, 1006 a, 6–9)

9 See Wittgenstein (1974: § 96):

> It may be imagined that some propositions of the form of empirical propositions, were hardened and functioned as channels for such empirical propositions as were not hardened but fluid; and that this relation altered with time, in that fluid propositions hardened, and fluid ones became fluid.

See also ibid.: § 97–99.

Part II

Ontology

3 Aristotelian substances and the theoretical/practical dichotomy

Edmund Runggaldier

Introduction

Why do most analytic philosophers – as it seems – accept naturalism? Why are many intellectuals – especially in Continental Europe – convinced that it is not possible to be a committed analytic philosopher and at the same time anti-naturalist?

I contend that it is due: (a) to the method of analysis: The last components into which compound entities are analysed, i.e. the ultimates of analysis, are taken to be the *basic entities*; and (b) to the theoretical/practical dichotomy: Presuppositions of practical philosophy are excluded from ontology which is considered to be a mere theoretical discipline.

Both (a) the thesis that the *ultimates of analysis* are the basic and only entities of reality and (b) the consequences of the theoretical/practical dichotomy, undermine substance ontology and favour naturalistic ontologies.

Substances (οὐσίαι) are made of stuff (ὕλη) but are for Aristotle not identical with it, even though they occupy the same space-time region. However, Aristotle did struggle with different intuitions on the ontology of stuff. His conjecture that stuff has substantiality culminated in the modern account of 'materia'. This too favoured a naturalistic ontology of stuff and the consequent abandonment of substance ontology. Typical for this shift in the account of 'materia' is Suárez and one of the fathers of analytic philosophy on the Continent, Bolzano.

Contrary to the naturalistic positions I assume that in ontology we have to account for the presuppositions of practical philosophy as well, such as the assumptions of personal identity, agency, the subjective perspective and indexicality.

These aspects of personal agents do not require a dualistic ontology. There is only one reality, but more than one approach to it: the objective scientific method, on the one side, and the indexical or subjective approaches as presupposed by practical rationality, on the other. Substance ontology allows for both, and thus has various advantages over either a naturalistic monistic or a dualistic ontology.

Continental attitudes towards analytic philosophy

Today, it is almost impossible to find a common denominator of philosophical positions among analytic philosophers. There doesn't seem to be unanimity among them. In mainland Europe, however, analytic philosophy is still identified with early analytic philosophy, especially with the thesis that conceptual and linguistic *analysis* is the proper method in philosophy. Names still associated with it are Russell, Wittgenstein and Carnap who have been important in practising, clarifying and defending this thesis.

Philosophical analysis

In those days, philosophical analysis was very largely a matter of exhibiting the inter-relationship of different types of propositions, e.g. of showing how statements about material things (substances) are related to observation-statements (Ayer 1936: 31). The model for analysis might be seen in Wittgenstein's *Tractatus* and in Carnap's *Der logische Aufbau der Welt*. Their central question was: how does the truth value of complex propositions depend on that of elementary or basic propositions?

Some of the early analytic philosophers claimed that linguistic analysis provides a vital tool for laying bare the logical form of reality. The early Wittgenstein in particular contends that the structure of language reveals the structure of the world. Every meaningful sentence is analysable into atomic constituents that designate the fine-grained constituents of reality. The foremost practitioner of analysis, thus understood, has probably been Russell. Taking reality as one great analysable complex, he asked: What are the *ultimate constituents*? Analysis reveals not only what the last constituents of language are, but of reality as well.

A fact is thereby a complex of particulars, qualities, or relations that makes a proposition true or false. Atomic facts correspond to atomic propositions – a molecular proposition being one that contains other propositions as its components linked by truth-function words. The function of analysis is to resolve all descriptive complex propositions into their elementary ones and these into their ultimate linguistic symbols, representing the ultimate simples of the world.

This view of the analytic method created the suspicion that analytic philosophy is committed to the thesis that the ultimates of analysis are *the basic entities* of reality. What is given and real are these elements. Everything else is reducible to them and can be constructed from them.

Especially Russell's *theory of descriptions* nourished this suspicion. In fact, it served as a model to the early analytic philosophers in explaining away suspicious entities. Words for classes, relations, and Aristotelian substances were dealt with in the same way as descriptions. A typical case for analysis in this sense might be Quine's elimination of singular terms by transforming everyday statements into the canonical notation of quantified language.

Talk of Aristotelian substances has to be replaced by talk lacking its strong ontological commitments.

The method of analysis favoured a monistic ontology whose entities are the ultimates of analysis. These might be either *atoms* or some kind of *tropes* accessible to science. Analysis, thus understood, fostered the conviction that analytic philosophy is in some sense continuous with science. Opposition to 'analytical thinking' was from the very beginning based upon the suspicion that it inevitably reduces the immediately felt unities like persons or Aristotelian substances to a series of mere assemblages or sets of ultimate viz. atomic elements knowable by scientific methods. If confronted with analytic philosophy, many intellectuals in Continental Europe still think that analytic philosophers are bound by the mentioned consequences of the method of analysis and are thus committed to some form of *naturalism*.

There is little agreement about what naturalism is. However, characteristic of naturalism is the affirmation of a continuity between philosophy and empirical science (Papineau 1993: 1). Contemporary naturalism, so it seems, might be characterized as a cluster of various philosophical positions committed to the view that real entities may be described in scientific terms and explained in terms of event-causality. If this is so, naturalism implies an ontology in which the only entities allowed are those that are accessible to physical description and explanation: 'Any entities that are taken to exist should bear a relevant similarity to entities that characterize our best physical theories, their coming-to-be should be intelligible in light of the naturalist causal story, and they should be knowable by scientific means' (Craig and Moreland 2000: xii).

If the ultimates of analysis are taken to be some kind of atoms and if these are considered as basic or as the only entities there are, one's own ontology inevitably will be a naturalistic ontology of stuff. Many intellectuals in Continental Europe, in fact, seem to be convinced that because of the method of analysis in Russell's and Wittgenstein's sense analytic philosophers, generally, are committed to naturalism. But decisive for this attitude towards, analytic philosophy seems to me the theoretical/practical dichotomy.

The theoretical/practical distinction

Early analytic philosophers persistently stressed that many problems in classical philosophy and in everyday life are due to confusion between different kinds of statements or language-forms, especially between descriptive/theoretical statements, on the one hand, and normative/practical statements, on the other. There is a clear difference between reasoning undertaken in order to determine what is *true* and reasoning undertaken in order to determine what to *do*. Theoretical problems are addressed to us as *knowers*, and concern questions of what is the case, whereas practical problems are addressed to us as *agents* and concern what one should do. The

latter are solved by choosing the right action; the former by one's forming the right belief. Theoretical reasons are reasons for believing, practical reasons are reasons for acting (Audi 1989).

Neglect of the theoretical/practical distinction might lead to what Moore called the 'naturalistic fallacy'. To commit this fallacy is to confuse ethical propositions with either psychological propositions or descriptive definitions: what is meant to be an ethical proposition, that something is good, becomes a proposition about people's desires or their usage of words. To say what one should do is not to be confused with describing states of affairs.

Practical and theoretical reasoning have both been much discussed by early analytic philosophers, but separately. The search for the ultimate constituents of reality or of the basic elements falls for them exclusively within the theoretical domain. In order to ascertain what the basic entities are, one has to reason theoretically and to exclude practical aspects. The search for the ultimate constituents is a search for truth. It has to set aside the subjective perspectives and the first person approach to reality.

On the other hand, practical reasoning is based on the *first-person* perspective: I have to decide what to do; I am confronted with the problem of what to do; I have to reason in order to choose an action or a series of actions. This ego-centric approach has to be excluded from theoretical reasoning, as it is excluded from science. Science is not based on any privileged access to reality. Its striving for objectivity aims at a point of view – as it were – from nowhere.

Against the background of early analytic philosophy it seems therefore evident that ontological questions fall exclusively within the realm of theoretical philosophy and science. Both strive for objectivity and exclude teleological reasoning. Thus, it seems plausible that in the tradition of analytic philosophy ontology should conform to the methods of science and exclude that kind of reasoning which falls within the practical domain. All this deepens the suspicion that ontology within the analytic tradition cannot but be naturalistic.

Naturalistic ontologies tend to be monistic. Most of them are either four-dimensional-event-ontologies, as Quine conceived them, or trope-ontologies. Certainly there is no place in them for Aristotelian substances as subjects of physical and intentional properties, and no place for agent causality: for a naturalist in general all features of the world are entirely caused by and realized in systems of micro-elements, i.e. ultimates of analysis. Their properties and their behaviour are sufficient to determine everything that happens. It is a bottom-up picture of causation. If apparently there is top-down causation, this only works because the top level is already caused by and realized in the bottom levels.[1]

Among many continental intellectuals opposition to analytic philosophy is due – so it seems to me – to the suspicion that any analytic ontology cannot but be naturalistic. One of the main reasons for their attitude might be – as I said – the method of analysis and the other reason the mentioned

exclusion of the presuppositions of practical considerations from ontology due to the theoretical/practical dichotomy.

I assume that ontology has to account for the presuppositions of practical philosophy as well, i.e. of agency, subjectivity and intentionality. Ontology should not fall exclusively within the realm of theoretical philosophy. I want, thus, to argue for the acceptance of Aristotelian substances as part of a strategy to avoid, on the one hand, monistic naturalistic ontologies and, on the other, dualism. Substances as basic entities are not identical with their stuff and can be subjects of both bodily and intentional properties. Their acceptance in one's own ontology implies neither the abandonment of the method of analysis in philosophy nor the neglect of the theoretical/practical distinction. It presupposes, however, that the ultimates of analyses do not coincide with the basic or fundamental entities of reality.

Historical background to Aristotelian substances

Classical ontology with Aristotelian substances was developed by Aristotle himself. Substances (οὐσίαι) are for him *ultimate* subjects of predication and ontologically independent. They are made out of stuff (ὕλη) but are not identical to it, even though they occupy the same space-time region. However, Book Z of Aristotle's *Metaphysics* seems to support an ontology of stuff as well. Aristotle's intuition that stuff might be more than pure potentiality culminated in a different account of the meaning of 'materia' in the seventeenth and eighteenth centuries. This shift in meaning favoured an ontology of stuff and the consequent abandonment of substances. It ultimately favoured a mere naturalistic ontology, on the one hand, and dualism, on the other.

The problem of stuff (ὕλη)

What is substance (οὐσία)? Aristotle first considers what others mean by it: if different people mean different things by οὐσίαι, they will propose different candidates for it. Are substances Anaxagorean stuffs or Democritean atoms? The earliest Greek thinkers had taken stuff, and in particular elemental stuff, to be substance: partly just one stuff, partly the four elements.

It is known that Aristotle denies that stuff is substance. It was plain to him that knowledge of atoms, however subtle it might be, could not possibly be basic (Barnes 1995: 100f.). Botanists, for instance, do not consider atoms: the objects of botany are animals. Aristotle conceives of stuff as *potentiality*: to say that wood is stuff of boxes is to say that wood has the potentiality or capacity to become a box. He contrasts potentialities with *actualities*: δυνάμεις *versus* ἐνέργειας. Capacities are not as real as actualities or their realizations. Thus they cannot be basic. As mere potentialities they are 'posterior' to actualities. Actuality is prior to capacity even in time (Book Θ 8, 1049b 18–25). And if, in general, actuality is prior to potentiality, then substance must be prior to stuff.

Aristotle thus maintains that the physical parts of a body are less basic than the body itself. A finger, for example, is defined by reference to the whole body: 'a finger is such-and-such a part of a man. Hence the parts which are of the nature of matter, and into which, as its matter, a thing is divided, are posterior to it' (Book Z 10, 1035b 10ff.). In order to explain what it is to be a finger, we must make reference to bodies.

However, the naturalistic intuition that stuff too has substantiality is not completely alien to Aristotle. At the end of Book Z 3 he lists matter (ὕλη) as one of three possible sorts of substance (1029a 30ff.). Perhaps Aristotle himself recognized some problems in the exclusion of matter as subject (Gill 1989: 39). The particular bronze out of which the statue is made seems a particular of some kind: a τόδε τι and separate, since it can exist apart from the statue and possess its own properties. The form of a statue depends upon bronze for its existence, while the bronze has no need for the form of the statue; the bronze can exist on its own before it acquires the form and can continue in existence after the form has been removed. Matter is something in its own right. That matter is the ultimate subject seems to have a certain plausibility.

According to the dictionary in Book Δ, substances have to satisfy two conditions: they must be ultimate subjects of predication and separable, i.e. ontologically independent: 'Things are called substances in two ways: a substance is whatever is an ultimate subject, which is no longer said of anything else; and a substance is a this so-and-so which is separable' (Book Δ 8, 1017b 23ff.). A subject is that of which other things are predicated but which is not itself predicated of anything else (ibid.: 1028b 36f.). But what counts as ultimate subject? The composite of stuff and form or stuff as a component thereof? Book Z 3 indicates some uncertainty about what type of entity should count as an ultimate subject.

Matter satisfies the subject- and separation-criterion and therefore succeeds as substance in a way. Besides, if matter persists throughout the generation, career, and destruction of a composite, then matter has a nature distinct from that of the form whose temporary presence gives the composite its particular identity. Yet if composites are primary substances, they must be conceptually primary entities. Thus, the kind of unity needed to account for change conflicts with the unity required of those entities that are primary substances. This is the problem Gill calls 'the paradox of unity' (1989: 6f.).

Still, ultimately, Aristotle rejects the naturalistic view of the 'antiqui' that matter is substance, even though, as seen in Book Z 3, he tends to share some of their intuitions. For Aristotle, overall, matter has a lesser claim than the composite. Typical substances, i.e. basic entities, are organic composites, such as human beings and horses. Such entities are both ontologically and conceptually primary. Separation (χωριστῶς) and thisness (τόδε τι) are for Aristotle crucial features of substance, and so the composites of matter and form are better candidates than matter alone to be substance.

The intuition that matter too has a certain substantiality got strong support in the seventeenth and eighteenth centuries – on the basis of the criteria of subjecthood and ontological independence. There has been a shift in the account of the concepts of matter and substance favouring, on the one hand, a naturalistic ontology and, on the other, dualism. I want to mention two authors within the Aristotelian tradition who are representative of this shift. One is Suárez who decisively influenced the work of modern Scholastics and of philosophers such as Descartes, Leibniz, Wolff, and Schopenhauer. The other is Bolzano, whom some consider the early father of the modern analytic movement on the Continent.

Modern shift in the understanding of 'materia'

The problems of the interpretation of the Corpus Aristotelicum were the subject of endless discussions throughout the Middle Ages and the modern period. Widely discussed among them was the problem of the ontological determination of ὕλη/*materia*. Scholastics dealt with the problem of the materia as ultimate substratum or subject: Is it mere *potentiality* or is it a kind of *positive entity?* Suárez tends toward the second answer. Sections 3 and 4 of the *Disputatio* 13 (Suárez 1861) are representative for his efforts to find a solution to the contrasting intuitions about materia: materia is conceived as mere potentiality (see Aquinas) but, being the last *subject* of predication and lasting *substratum* of substantial changes, as something real as well: 'aliquid rei, aliquid substantiale' (Suárez 1861: D. 13, s. 4). As such, it cannot be nothing ('omnino nihil') nor mere accidence: Suárez defends the assumption of Aristotelian substances as basic/fundamental but stresses, on the other hand, that materia cannot be mere potentiality. Materia has its own existence and is in its kind (species) stable.

Materia is for Suárez the first/ultimate subject of changes and forms ('primum subjectum mutationum vel formarum') (Suárez 1861: D. 13, s.1, 4). Since regress *in infinitum* is to be excluded, there must be such an ultimate or first subject for changes, and this is matter. One has to stop at a last subject of predication, which is no longer subjected to something else ('quod non subjectetur') (Suárez 1861: D. 13, s.1, 10). If there weren't materia as enduring subject, it wouldn't be possible to account for substantial changes, the generation and the destruction of substances. Annihilation and creation out of nothing are excluded from the realm of nature.

Suárez is, however, deeply reluctant to accept a naturalistic ontology: substances are real composita and not identical with what they are made from. He definitely rejects the view that materia is the only constituent of reality. He too points out that a mere atomistic ontology does not allow of any generation or corruption. Changes and movements would consist in 'perpetua agitatione', i.e. in mere shifts and changes of complexes of atoms. Differences between things would be reducible to mere differences in place and complexity of atoms ('in situ et coordinatione atomorum') (Suárez 1861: D. 13, s. 2, 3).

Suárez's main reasons for accepting Aristotelian substances, i.e. the composita of materia and forma, as basic or fundamental, and not matter, are, however, reasons of practical philosophy, e.g. the personal experience that we are *agents* (Suárez 1861: D. 13, s. 3, 9). We have the experience of being *continuants* endowed with various potentialities and powers. We do act and cause changes in the real world and are thus agent-substances. The starting point for the justification for an ontology with substances as basic entities is that we are unitary living beings and agents (Suárez 1861: D. 15, s.1, 6).

Even though Suárez clearly distinguishes between theoretical and practical philosophy and tends to confine ontology to the realm of theoretical philosophy, he does not exclude from it arguments taken from the practical realm, deriving from action theory and legal philosophy: both agent and final causality presuppose Aristotelian substances. In a mere naturalistic ontology devoid of Aristotelian substances there would be neither final nor efficient causality.

Again and again, Suárez stresses the point that ontology requires a wider approach than a mere physical one. One has to take more aspects into account, especially the phenomena of change and action, human action included. Suárez, therefore, supports an ontology of Aristotelian substances. However, I mentioned him because of his sharing the intuition that matter too has a kind of substantiality. He certainly is closer to a naturalistic perspective than Aquinas, materia being for him last substratum. The shift in the understanding of 'materia' favours the intuition that the ultimate ontologically independent substratum is matter. To this relates the common, modern ordinary sense 'what a thing is made of'. Knowledge of stuff or of the last constituents of things, on the other hand, is best achieved by science. The shift in the meaning of 'materia', thus, favours in the long run a naturalistic ontology.

I turn now to one of the fathers of analytic philosophy on the Continent, to Bolzano. His ontology is a typical example of the consequence of both, the shift in the account of the meaning of 'materia' and the analytic striving for the ultimates of analysis as the last basic constituents of reality. Bolzano strongly opposes all kinds of idealism or transcendental philosophy and strives for an ontology of substances, but his conception of substance is not Aristotelian. On the one hand, it is *in nuce* naturalistic, material substances being for him *atoms*, and on the other dualistic, spiritual substances being for him *souls*.

For Bolzano, substance is, as for Aristotle, something real which is not a property of something else, i.e. ontologically independent. Bolzano, however, conceives substance as *simple*, departing from the original Aristotelian conception: x is a substance = x is real & x is not an adherence & x is simple (Bolzano 1994: § 142; 1827: 21f.). Everything with a finite extension is composed of infinitely many simple substances which have no parts. Each material thing is constituted by *atoms*, located at single points in space. Substances, as basic entities of reality, are not living beings in their

primordial unity, but the atoms or the stuff these are made from. Bolzano's ontology does not allow of unitary substances as bearers of physical and intentional properties.

Being a competent psychologist he had to account for human action and intentional or mental phenomena as well. He did this by opting for a dualist ontology: he postulated a special kind of substance, i.e. spiritual substances or souls. These too are simple and are not composed of parts. They are thus immortal. However, being continuants they can change in time. Bearers of perceptions and intentional properties are not humans but souls.

Bolzano's ontology is a typical example of the consequence of the shift in the understanding of 'materia' and 'substantia' favouring, on the one hand, naturalism, on the other, dualism. One is consonant with the striving for the ultimates of analysis as the last constituents of reality, the other with the consequences of the theoretical/practical dichotomy. The primordial unity of the Aristotelian substance, as bearer of bodily and intentional properties, has vanished. Instead of an ontology of Aristotelian substances, we are confronted, on the one hand, with an ontology of stuff, on the other, with an ontology of souls. The dichotomy between theoretical and practical philosophy has an equivalent in the dichotomy between material and spiritual substances.

We have seen that Aristotle's text supports intuitions that stuff too has some substantiality and that the shift in the account of the meaning of 'materia' in the seventeenth and eighteenth century has decisively strengthened these intuitions.

Advantages of Aristotelian substances

In science we do not need Aristotelian substances in order to make predictions and explain the various empirical phenomena associated with the apparent generation, existence and destruction of such entities. Accepting only stuff that constitutes them, along with its various properties, will do just as well. If, on the other hand, we do take into account practical philosophy, we are confronted with responsibility, personal identity through time, agency, subjectivity, indexicality, and so on.

A narrow ontology of stuff will not suffice for an adequate account of such phenomena, unless we complement it with dualism. The best way to avoid, on the one hand, naturalistic monism and, on the other, dualism still is to accept Aristotelian substances as basic entities. Taking them as basic/primitive, we do not have to reduce them to some complexes of ultimates of analysis or construct them out of these ultimates.

Identity through time

In science we can get along without diachronic identity, generation and destruction, ultimately even without change. In science there is no 'fact of

the matter' about identity. Nothing is destroyed when stuff is rearranged. Quine sticks to the naturalistic consequence of his four-dimensional ontology that there is not even change.

But for practical purposes and practical rationality we do rely on identity through time. We assume it whenever we talk of responsibility and of striving for aims. For the purpose of practical rationality we take ourselves to be *endurers* or *continuants*.

Four-dimensional space-time ontology was and is very successful, especially in its application in science. Not only events but things as well, human beings included, are conceived as spread out in time, composed of temporal stages or parts. For practical purposes, however, being acting humans, we assume that we are only three-dimensional beings, spread out spatially but not temporally. If we substitute the naturalistic ontology of stuff with an ontology of Aristotelian substances, we can account much more easily for these assumptions, since Aristotelian substances are continuants, moving through time and being wholly present at each moment of time.

Aristotelian substances have *persistence/endurance conditions*. These and their existence-conditions depend on their *kind* or *species*. It is not a matter of convention what determines their beginning and end in time and the kind of change they can undergo. For every substance, there are facts about what changes it can and cannot survive, and these are determined by what they are. Rea calls them 'modal facts', since they are facts about what is necessary or possible for an object or a substance of a given kind. A man over six feet tall cannot survive being crushed down to the size and shape of a sugar cube (Rea 2000: 112). The potencies, tendencies and capacities of a substance depend on what it is.

In Aristotelian ontology, things are not identical with the material they are made of or with the sum of their parts. The nature of their composition, their dispositional properties, potencies and tendencies, powers and their mode of activity are constitutive for them. When we are asked what a given thing is, we answer with a *sortal expression*. Sortal expressions convey these conditions, the continuity/persistent conditions included, i.e. the conditions of coming to be and passing away. Where there are no persistence conditions, there is no object, but only stuff. Having persistence conditions is part of our concept of a material object.

In substance ontology the sortally determined individuals are *basic/primitive*. They are centres of action, potencies and functions and as such not the result of some conventional construction out of atoms or other ultimates of analysis. The persistence conditions of substances, humans and other organisms included, are not results of conventional posits or linguistic conventions, but *given*.

If we, thus, distinguish between the stuff making up an object and the sortally and functionally determined object itself, between the *materia* and the *compositum*, do we have to assume that two numerically distinct objects can occupy at the same time the same space-region? If the tree and the

aggregate of cellulose molecules that constitute that tree are distinct, it seems that there are two objects in exactly the same region of space-time (Rea 2000: 114f.). Do we have to accept *cohabitation*?

There is, however, not only the alternative either to accept or to reject cohabitation, i.e. either to accept or reject different persistence conditions of objects occupying at the same time the same space region. In substance ontology *objects proper* are only sortally determined entities falling under a natural kind, but not their stuff. Thus, the tree and the aggregate of cellulose molecules that constitute that tree are not two objects on a par. One is only the stuff of the other.[2]

For scientific purposes monistic ontologies – be they ontologies of stuff, of tropes or energy fields – might suffice. For an account of the ontological presuppositions of practical philosophy, however, we need more. If we do not exclude from ontology considerations and assumptions of practical rationality, we need a stronger ontology.

Substance ontology provides us with various advantages. It allows also for agent causality. We do not need it in science, we even have to exclude it from scientific investigations, but we need it for practical philosophy. If we do not limit ourselves to a purely naturalistic ontology we do not have to exclude this form of causal explanation. We are free to consider agent causality as a species of causality.

Agent causality

The idea behind talk of agent causality is that agents *bring about* or *produce* effects; they are somehow the source of what they do. In this sense a human being is said to be the cause of what she brings or has brought about. The causal relation is, thus, said to hold not only between events, but between an agent as an endurer or an Aristotelian substance, on the one hand, and an event/state of affairs or a series of events/states of affairs, on the other.

In the context of a wider ontology with Aristotelian substances we can think of causal notions in terms of 'causal power' or 'causal capacity'. Given this ontological framework, the difference between event causation and agent causation concerns the way in which causal capacities are exercised. In the case of event causation, having in the right circumstances the cluster of properties that ground the capacity directly gives rise to one of the effects within its range. 'By contrast, having the properties that subserve an agent-causal capacity doesn't produce an effect; rather, it enables the agent to determine an effect (within a circumscribed range)' (O'Connor 2000: xiv). When and how such capacity will be exercised is determined by the agent.

A modern version of the Aristotelian approach to causality is due to R. Harré and E.H. Madden (Harré and Madden 1975: 5). When placed in the appropriate circumstances, a particular manifests its causal powers in observable effects. They are based in the object's underlying nature, e.g. its physical, chemical, or genetic constitution and dynamic structure.

However, it was Thomas Reid who emphasized the role of agent causality as a species of causality: if an agent freely and directly brings about an effect, she must have the capacity to represent possible courses of action to herself and have desires and beliefs concerning these alternatives. Her active powers are grounded in a set of properties that in conjunction with appropriate circumstances make possible her producing an effect. 'These choice-enabling properties ground a different type of causal power or capacity – one that in suitable circumstances is freely exercised by the agent himself' (O'Connor 2000: 72).

For purposes of practical philosophy mere event causality will not suffice; we have to assume agent causality and for this we need a suitable ontology. We need endurers/continuants endowed with powers and capacities to act, i.e. Aristotelian substances. They endure through time, wholly existing at each moment of an extended temporal interval, as opposed to things that perdure (in Lewis' terminology). In an Aristotelian context, freely acting agents are biological entities that have mental properties and capacities. Their powers, however, must not be reducible to the powers of their microphysical constituents or stuff.

If we opt for an ontology naturalistically confined to what things are composed of, we have to assume that every macro-level phenomenon – human action included – has arisen through natural micro-physical causal processes. Its existence continues to causally depend on processes of this kind. All macro-level phenomena are not only causally dependent on but also constituted by a whole network of microphysical processes. In contrast, the Aristotelian agency theorist is committed to a special sort of properties, among them intentionality and consciousness, enabling the substance that has them to freely and directly bring about any of a range of effects.

But are the bearers of intentional properties really the same as the bearers of the biological properties? Do they not require a very different kind of substance? Doesn't Aristotelian agency theory presuppose substance dualism? Many who discuss agency theory seem to assume that its adherents are dualists. As a matter of fact, because of the peculiarity of human agent causality and intentionality many plead for a dualistic ontology. Aristotelian ontologists, by contrast, are not committed to substance dualism.

One reality

Aristotelian substances allow for both kinds of properties, bodily and intentional. Substance ontology reconciles two intuitions: agents are part of the causal order of nature and can nonetheless freely bring about various state of affairs or changes in reality. The employment of the concept of agent causality is not at odds with the attempt to give a scientific account of natural processes. However, it does run counter to the general naturalistic programme of reductionist explanation.

It is constitutive for human persons to engage in theoretical *and* in practical reasoning: reasons have their characteristic effects on both, beliefs *and* actions. However, practical reasoning presupposes a subjective perspective, the first-person view and indexicality or egocentricity. Reasoning that instantiates practical arguments in the objective perspective is not what is commonly called practical reasoning (Audi 1989: 102). Practical reasoning is centred on the subject *hic et nunc*, in particular situations, indicated by indexicals.

This, however, does not imply that there are special facts or a special kind of reality besides the one we refer to in the objective language. By using indexical identity-statements like, 'today is the 24th of March', 'This is Milan', 'I am E.R.', we express our belief that what we refer to with indexicals is identical to what we refer to in the objective language. There is no unbridgeable gulf between the subjective/indexical and the objective, between propositions containing indexicals and propositions formulated in the objective language of science.

'Here'-thoughts, 'this'-thoughts and 'I'-thoughts do not presuppose that there is a special reality besides the one we approach by objective thoughts. By using indexical I-identity-statements we claim to be part of the objective spatio-temporal framework. Our subjective thoughts of ourselves presuppose the intelligibility of a link with the world thought of objectively: 'Nothing can be a cognitive map unless it can be used as a map – unless the world as perceived, and the world as mapped, can be identified' (Evans 1982: 212). The subjective mode of thought and its approach to reality are bound up with the objective mode.

Perspectival subjective beliefs are by no means mere matters of opinion with one opinion as good as any other. It is *true* that I am E.R., and any contrary beliefs would be false. However, when I know that I am E.R. and I know it in the first person, nobody else can know this in this way. 'The reason you can't know this is that there's no first-person belief of yours which this first-person fact could make true' (Mellor 1991: 6). For you, the corresponding belief is a second-person or third-person one.

My subjective knowledge is confined to me, i.e. to one single person. This seems to imply that there are subjective facts, facts belonging to a special world besides the objective world. So it seems that we do not live in a single objective world. But our being *agents* endowed with subjective knowledge and the first-person perspective does not make us subjective beings, trapped in our own subjective worlds. No such worlds exist. There is just the one single world of which each of us, every self, every subject, is equally – and wholly – a part.[3]

To reject subjective worlds and subjective facts is by no means to deny that we all have our own first-person knowledge of ourselves and of how things are. The manifest diversity in our first-person knowledge shows that we neither live in different worlds, nor that our subjective knowledge is really mere opinion. This is a false dichotomy.

> Just as there is one world onto which both perception and conception may be directed, despite the different ... sorts of representation, so there is one world onto which indexical and non-indexical thought are directed, despite the essential differences between these styles of thought.
>
> (McGinn 1983: 92)

Conclusion

If, in ontology, we do not restrict our attention to those processes that can be studied by positive science and, thus, do not exclude the presuppositions of practical rationality, it makes sense to assume that the *basic/primitive* entities are *agents*, i.e. *Aristotelian substances*. The basic/primitive entities should not be confused with the *ultimates of analysis*. Within the framework of substance-ontology, diachronic identity, agent causality and indexicality do not look mysterious. We do not need to reduce or to naturalize them. There isn't only one alternative, either naturalism or dualism. An ontology of substances instead of one of stuff has the benefit of allowing for more than one approach to the one reality of human beings. There is only one reality, but more than one approach to it: the objective scientific method, its factual scope being limited, is only one of them.

Notes

1 See O'Connor (2000: 108).
2 See Runggaldier (1998: 366ff).
3 See also Mellor (1991: 15).

Aristotelian ontology versus naturalistic ontology

Comment on Runggaldier's paper

Alessandro Giordani

Introduction

The present paper analyses the correctness of an argument aiming to show that Aristotelian ontology justifies a better interpretation of the world than naturalistic ontology.

The argument can be synthesized as follows. There is only one world and there is an evident distinction between descriptive propositions, expressing a *theoretical approach* to the world, and prescriptive propositions, expressing a *practical approach* to the world. The former is connected with scientific knowledge, based on the third-person perspective, objective, which is to say in principle not depending on a subject, and intersubjectively under control. The latter, instead, is connected with indexical knowledge, subjective, in principle depending on a subject, and not intersubjectively controllable. The study of the world that becomes accessible through the scientific approach does not involve reference to human features expressible in first-person language, while the study of the world that becomes accessible through the indexical approach involves reference to basic human features such as personal identity, intentionality, responsibility and agent-causation. If one adheres to the sole scientific approach – as the several trends of naturalism do – one tends to develop a monistic ontology, i.e. an ontology that recognizes only one type of basic entity, the one studied through scientific knowledge and described in the language of science. If, on the contrary, one does not adhere to the sole scientific approach, one can develop an Aristotelian ontology, where basic entities are matter-form structured substances that can be studied both from an objective point of view, with respect to properties typical of material states of affairs and from a subjective point of view, with respect to properties typical of mental states of affairs. So, Aristotelian ontology enables us to give an account both for the ontological model underlying theoretical, objective knowledge and for the ontological model underlying practical, subjective knowledge. As a consequence, it enables us to develop a better interpretation of the world.

The problems connected with this argument can be reduced to three: (1) the assumption of a scientific approach to the world does not imply the

exclusion of subjectivity or intentionality; (2) the assumption of an ontology of substances of Aristotelian type does not imply the exclusion of ontological models deriving from the scientific approach to the world; and (3) the assumption of an Aristotelian ontology of substances is not a solution but it is linked to the problem of the relation between the objective and the subjective world, involving the negation of causal closure of the objective world. An analysis of these three problems will be presented below, together with a hypothesis of solution to the problem of illusoriness of the subjective dimension, in order to justify the confutation of an extreme naturalistic conception such as eliminativism.

Objective/subjective dichotomy and intentionality

The first problem suggested by the argument presented above derives from the way in which the connection between the objective approach to the world and the third-person perspective is posed. The connection indeed is posed in such a way that the assumption of an objective approach seems to exclude the assumption of a first-person perspective, or an intentional access to the world. Yet, to say that a person has an *access* to the world means that the actual world *appears* to that person. To say that a person has an objective approach or an *objective access* to the world means that the actual world *appears* to that person in such a way that what appears does not depend on that particular person x to whom it appears, but it can appear as it is to everyone who can assume the same position as x or entertain the same relation as x to the world. In other words, what appears is objective and is objectively known when it is *invariant with respect to the substitution* of the person to whom it appears, i.e. when it is invariant with respect to the first-person perspective.[1] Thus, objective perspective is established on the basis of invariance with respect to subjective perspective, not on the basis of exclusion of such perspective: if subjective perspective is denied, the very access to the world is denied. Both theoretical, objective approach and practical, subjective approach are based on the same human intentional access to the world.[2]

Objective/subjective dichotomy and naturalistic ontology

Ontology is the study of substance, intended as a necessary condition for the actual being of a state of affairs.[3] Thus, substance is presented in the answer to the question: what exists properly? If substance is so intended, every philosophical position that does not limit itself to claim the existence of an objective domain of being, but claims the existence of a sole domain, an objective one, or of a domain on which every other domain depends, points out what is classically called substance. Thus, it is legitimate to try to find out a naturalistic conception of substance in so far as naturalism is committed to the answer to the question: what exists properly?

The naturalistic conception of substance is strictly connected to the scientific approach to the world, in particular, with the approach typical of physics, the basic natural science. In fact, the claim shared by every naturalistic conception can be expressed in the following way: naturalism is 'the view that nature is all there is and all basic truths are truths of nature',[4] where 'truths of nature' are the truths accessible through the scientific approach. The proposed definition involves two theses. First, an *exclusion thesis*, according to which what exists in a proper and independent way is only what is natural. Second, an *identification thesis*, according to which what is natural is the object of natural science, especially physics. The outcome of the two statements is that nature is the totality of states of affairs constituted by space-and-time-determined elements knowable by natural science. So, what exists is what is scientifically describable and knowable, where describable means accessible on the basis of experiments and knowable means accessible on the basis of a theory. The ontological conception of naturalism – substance is the object of science – is therefore based on the assumption that the objective access is the sole possible access to what properly exists. This assumption constitutes the problem of sole access, which will be considered below.

The naturalistic view[5] can be further articulated so as to give rise to at least two different conceptions: an eliminativistic one and a non-eliminativistic one. The difference between these conceptions can be better understood if some preliminary definitions are introduced. Let us here introduce three different well-known domains: *world 1*, which includes material states of affairs; *world 2*, which includes mental states of affairs; and *world 3*, which includes the contents of mental states of affairs, abstract objects such as propositions or demonstrations. The two conceptions of naturalism share two specific theses and differ as to the way of intending the world of mental states and consequently the world of mental contents: eliminativism maintains that world 2 is illusory; mental states and contents of mental states do not properly exist; instead, non-eliminativism maintains that world 2 is not illusory, but the states of this world do not constitute an independent and causally efficient world with respect to world 1. The shared theses are: *naturalistic thesis 1*, the existence of world 1 is a necessary condition for the existence of world 2, i.e. if world 1 did not exist, nothing would exist; *naturalistic thesis 2*, world 1 is causally closed, there are no causal influences from world 2 to world 1, i.e. the states and laws of world 2 are supervenient with respect to the states and laws of world 1. Since, in naturalistic ontology, the possibility of causal interactions is considered the typical feature of what properly exists, to say that world 1 is causally closed coincides with saying that world 1 includes all that properly exists. Similarly, to say that the states and laws of world 2 are supervenient with respect to world 1 coincides with saying that the existence and relations of mental states are based on the existence and relations of material states, from which derives that world 1 is causally closed.

Posing the question

The attack on naturalistic ontology under the initial argument is then as follows: given the assumption that world 2 is an existent, not illusory and not supervenient domain, the Aristotelian ontology is better, because it enables us to say that the same substance is the subject of both material, objectively accessible states and mental, subjectively accessible states. Moreover, it enables us to provide justification for the connection between material and mental states through the matter–form structure of substance. Yet, we can point out that such an argument does not represent the solution, but the problem itself. The same person is subject to both material and mental states. If there is interaction between the two types of states – as for instance in action – then it is necessary to deny the causal closure of world 1, i.e. one of the two theses typical of naturalism. Thus the argument clarifies only the opposition between the ontological independence of world 2 and the causal closure of world 1, but does not present a solution, if it is not completed by a further argument concluding for the initial assumption that world 2 is not illusory and not supervenient.[6]

Objective/subjective dichotomy and Aristotelian ontology

Before facing the issue of the ontological status of world 2, one must consider the relationship between scientific approach and Aristotelian ontology: is ontology based on the scientific approach to the world compatible with the Aristotelian conception of substance? The answer to this question induces us to make a preliminary distinction. The Aristotelian conception of substance can be analysed from two points of view: from an intensional point of view, the problem lies in determining the concept 'substance', i.e. what Aristotle means by this concept; from an extensional point of view, the problem lies in determining the set of objects falling under the concept 'substance', i.e. what Aristotle considers to fall under this concept. The compatibility between the scientific approach and the Aristotelian ontology regards the former point of view. Therefore, the question is whether the scientific approach delimits a domain of objects describable as substances from the point of view of Aristotelian ontology. The answer is affirmative.

Aristotle proposes two definitions of substance,[7] a logical one, according to which substance is what is not said of anything else and that of which everything else is said, and an ontological one, according to which substance is what is individual and independent. Substance is subject of a *per essence* predication, concerning predicates describing properties that a certain individual cannot cease possessing during its existence, and of a *per inherence* predication, concerning predicates describing properties that a certain individual can cease possessing during its existence. Since it is not predicated per essence of the other, substance is the individual object of which the universal – the knowable structure shared in principle by several objects – is

said to be, and, since it is not predicated per inherence of the other, substance is the independent object existing regardless of the existence of the other. Moreover, since the structure exists only as a structure of individual objects and the inherent property exists only as property of individual objects, if substance did not exist, nothing would exist. Thus, Aristotelian ontology is based on the existence of primary substance. Now, the substance of physical objects is discovered because it does not cease to be in changing and the distinction between changing properties and unchanging properties determines some basic characteristics of substance. In the first place, the changing object is composite of: (1) a substantial structure, εἶδος, *forma*, *causa formalis*; and (2) a substantial substructure, ὕλη, *materia*, *causa materialis*. The two causes, formal and material, are both necessary for a changing object to exist and both are knowable, ergo determined according to structural properties. In the second place, matter, as substantial substrucure, is not only potentiality with respect to form, but possesses a form of its own: this gold is potentiality, i.e. a possible substructure, with respect to a ring, but is different from silver, sharing a structure that differentiates it from other types of matter. In the third place, matter determines the way in which a changing object can interact with other objects, so that matter coincides with the whole set of objective interaction ways. Finally, if we call basic matter the substance possessing a structure but not a substructure, then such matter is to be considered as a necessary condition for every object to be real in the material world. If basic matter did not exist, there wouldn't be any changing object, because one of its necessary conditions would fail to exist. Thus, according to the proposed definition, Aristotle's matter is the analogue of the basic elements of physical interactions.[8]

Interpreted in this way, the Aristotelian concept of substance can be assumed by the scientific approach: physical science studies the structure of matter, intended as the whole of basic physical elements or, as Aristotle would have said, the substance of the physical world. What differentiates Aristotelian ontology from naturalistic ontology is not *the way of intending substance*, but *the way of identifying that which is substance*. Actually, from the point of view of Aristotelian ontology, the description of the structure of basic matter is not the description of the structure of every substance. In other words, Aristotle could accept the *naturalistic thesis 1*: if world 1 did not exist, nothing would exist; but he could not accept the *naturalistic thesis 2*: the superior structures of a substance are not supervenient with respect to the substructures. If this is true, we can conclude that substance ontology is compatible with the naturalistic point of view and that the basis to solve the problem of supervenience or independence of world 2 cannot be given by the acceptance of such ontology. In fact, it is impossible to maintain the causal closure of world 1, the causal independence of world 2 and the bilateral connection between world 1 and world 2. In order to face this problem, there are three possible solutions: (1) a monistic strategy, denying the causal independence of mental states; (2) a dualistic strategy, denying the bilateral

connection between material and mental states; and (3) an Aristotelian strategy, denying the causal closure of material world.

Is there a sole access to the actual world?

A naturalist can put forward the following argument: (1) substance is a unitary structure of ways of interaction; (2) access to substance is given by particular interactions with it; and (3) the interactions through which we have access to substance are determined by natural sciences, because only natural sciences are able to define objective – ergo legitimate – ways of interaction. Thus, only the scientific approach allows access to substance and to the domain of that which properly exists. Yet, by such an argument, premise (3) is no less problematic than its denial, because it centres on the *sole access posit* according to which the sole cognitive access to that which properly exists is the scientific access to that which objectively exists. Is such a posit justified?

The access to the world allowed by the interaction procedures defined inside natural sciences is always a human access to the world. The scientific approach allows the world to appear in its objective aspect, which is invariant with respect to the substitution of the person to whom the world appears. To the person x, there appears an objective state of affairs p and, together with it, there appears that particular mental state of affairs that consists in the appearance of p, be it $\mathbf{A}(x,p)$. The implication between the appearance of p and the appearance of $\mathbf{A}(x,p)$ is necessary, even though not explicitly recognized. Now, as the appearance of p constitutes the access to the objective and actual state of affairs p, so the appearance of $\mathbf{A}(x,p)$ constitutes the access to that particular subjective and actual state of affairs $\mathbf{A}(x, p)$. With respect to $\mathbf{A}(x,p)$ we can say that (1) *it is not objective*, because it is not invariant with respect to the substitution of the person; (2) *it is not illusory*. Let us assume that the appearance of p is illusory. In what sense is it said to be illusory? One is a victim of illusion with respect to a certain state of affairs when it seems to be real while nothing corresponds to it in the real world. Let us assume, then, that nothing corresponds to $\mathbf{A}(x,p)$ in the real world. What can be said of p? How is it possible to say that p is an objective and real state of affairs if $\mathbf{A}(x,p)$, i.e. the objective access to p, is illusory? If we declare that the appearance of the world is an illusion, it is no longer possible to describe an objective world: the illusoriness of the appearing implies the illusoriness of what appears. Therefore, if there are objective and actual states of affairs, then there are subjective and actual states of affairs and that which properly exists does not coincide with that which objectively exists.

This is why it seems possible to confute the extreme naturalistic conception eliminating world 2, even though the more liberal conception, which confines itself to holding that world 2 is supervenient with respect to world 1, has not been confuted.

Notes

1 The foundation procedures adopted by sciences tend to guarantee this invariance: the apparent dimensions of an object, for instance, vary as positions of the observers vary, since they can find themselves at different distances from the object; on the contrary, the measured dimensions do not vary, if the observers adopt the same measuring procedures.
2 The connection between scientific access to the world and intentionality or appearing is essential and will be used below to reject the illusoriness of subjectivity.
3 The definition of substance is the classical one. See Aristotle, *Metaphysics*, Book Z.
4 Audi (1996: 372).
5 For an introduction to contemporary naturalism, see Papineau (1993). We assume as naturalistic every conception sharing the two theses here presented.
6 The indication of the dichotomy between scientific language and a subjective one is, by itself, not sufficient to sustain the thesis that world 2 is not illusory or supervenient with respect to world 1.
7 For a global presentation of Aristotle's ontological thought, see Irwin (1988). For a discussion of the concept of substance, see Frede and Patzig (1988) and Loux (1991). In the present discussion, by substance, we mean what Aristotle indicates as primary substance.
8 As happens with basic matter, the basic elements of physical theories, such as elementary particles and fields, possess a structure defined by theories themselves through mathematical laws. However, the ontological definition of matter does not depend on reference to a particular physical theory.

4 What is naturalism? What is analytical philosophy?

Peter van Inwagen

The title of this conference is 'Analytical Philosophy Without Naturalism?', or, if I read the Italian title right, 'Can Analytical Philosophy Be Other Than Naturalistic?' Now I am an analytical philosopher, and I am not a naturalist – or, at any rate, I have not been a naturalist for at least twenty years. So I have been a practitioner of analytical philosophy without naturalism for twenty years. If I had ever been a naturalist, I certainly ceased to be one when I became a Christian. Before I was a Christian, I was neither an adherent of some other religion nor some sort of 'philosophical' theist. Nor did I hold any philosophical position – dualism, for example – that was inconsistent with naturalism. For all that, I'm not sure I ever was a fully paid-up card-carrying naturalist, but I can say that there was a time when I at least thought that naturalism was *probably* true.

My becoming a Christian had little or nothing to do with my philosophical work or my philosophical training – beyond the fact that I am the sort of person who is influenced by arguments, and all the arguments against Christianity that I knew of were, or so my philosophical training told me, either bad arguments (most of them) or inconclusive arguments (a few of them). But, of course, that all the arguments against some position are bad or inconclusive is hardly a reason for accepting it. (For one thing, at least in my view, to adopt the rule 'Accept any position such that all the known arguments against that position are bad or inconclusive' would lead to contradiction, since most if not all substantive philosophical positions have the following property: the arguments against them are at best inconclusive, and many of these positions are the denials of or are otherwise inconsistent with one another.)

When I became a Christian, I carried on with my philosophical work, just as physicians and cabinet-makers who have been converted to Christianity presumably carry on with theirs. In one sense, at least, my philosophical work did not seem deeply relevant to my religious convictions; no doubt most physicians and cabinet-makers would make a similar judgment. A physician who had been in the practice of providing the means of suicide to selected patients and who became a Christian would have to change his ways, of course, but he might well not have been doing that or anything

else inconsistent with Christian teaching. If I had been a defender of naturalism or of the moral permissibility of euthanasia, or of some other thesis inconsistent with Christian teaching, I too should have had to change my ways when I became a Christian. But, although I believed things inconsistent with Christian teaching before I became a Christian, I had never been interested in defending any of those things in my philosophical work. When, therefore, I ceased to be a naturalist I carried on being an analytical philosopher without noticing any break or discontinuity in my philosophical practice. And so I have done for twenty years. I have never seen any inconsistency in being both an analytical philosopher and a non-naturalist. (I should say that I don't see non-naturalism as a philosophical position as naturalism is a philosophical position. And I don't see opposition to naturalism as the constituting factor of any possible community of intellectual interest. If I learn that a notorious atheist is a Cartesian dualist, I don't say to myself, 'Well, at least he's a non-naturalist.' I say only: 'Well, that makes *two* things we disagree about.') I have never seen any connection between being an analytical philosopher and being a naturalist. Analytical philosophers, after all, disagree about a wide range of things; in my view, the truth or falsity of naturalism is one of them. There have been analytical philosophers who have believed that there were true contradictions, that unrealized possibilities were physical objects, that human persons were tiny material particles inside human brains, and that there were no such things as tables and chairs. If analytical philosophy is so tolerant a hostess as to admit such as these into her house, she will, surely, not gibe at granting entrance to a few non-naturalists.

It is of course true that most analytical philosophers are naturalists. Perhaps it is even true that most analytical philosophers are propagandists for naturalism, team players, proselytizing enthusiasts. (As I certainly was not when I was a naturalist – *if* I was a naturalist.) But the explanation of these facts, if the latter suggestion attains to facthood, must be psychological or sociological or something of that order. It is certainly not logical or philosophical, not even in part. Nelson Goodman once said that the opposition between empiricism and realism about universals (an opposition that in some sense certainly exists) was like the opposition between being a truck driver and being a lover of the ballet. That is, the opposition is real and undeniable, but to be explained otherwise than by appeal to the content of empiricism and realism. I would say the same about the connection between analytical philosophy and naturalism.

If I see matters this way, then, what in my view is there for a conference called 'Analytical Philosophy Without Naturalism?' to be *about*? – and what place is there in it for a philosopher with my views? I shall consider a few possibilities.

Someone might take the title of the conference to imply that a plausible *prima facie* case could be made for the thesis that there was an intimate connection (a logical or philosophical connection) between analytical philosophy and

naturalism. And that person might go on to infer that (because of the presence of the word 'without' in the conference title, and because the conference took place under the auspices of a Catholic university) the purpose of the conference was to provide a forum for philosophers who had something to say against this *prima facie* case. But if this is the purpose of the conference, I cannot contribute to it for the simple reason that I see no *prima facie* case for a philosophical connection between analytical philosophy and naturalism.

Here is a second possibility. Someone might suppose that the title of the conference was chosen in recognition of two facts: that most analytical philosophers are naturalists, and that this fact had no ground in the nature of analytical philosophy or in the content of naturalism. This person might suppose that the purpose of the conference was to underscore the second of these facts; that its purpose was to provide a forum for analytical philosophers who were not naturalists, to provide an occasion on which a series of philosophical papers were presented that were free from naturalistic presuppositions. Might I have something to say at a conference that fitted this description? I might indeed, but it is so easy to have something to say at such a conference that one can be reasonably certain that our second 'someone' has not got the intentions of the organizers of the conference right. Imagine that I had proposed to present a paper to you on one of the following topics:

- An examination of the account of variables given in Quine's 'Variables Explained Away'.
- An argument for the conclusion that agent-causation does not solve what writers on free will call 'the control problem'.
- A discussion of David Lewis's argument for the conclusion that any theory of composition besides universalism must have the consequence that the linguistic theory of vagueness is false.

(I have indeed written on each of these topics.) I can assure you that if I had presented a paper on any of these topics, it would have had no naturalistic presuppositions. And nor would it have had any non-naturalistic presuppositions. In fact, every proposition I should have defended in a paper on any of these topics would have been consistent both with naturalism and with the denial of naturalism. And that, I think, is exactly what would have rendered papers on these topics unsuitable for presentation at this conference. Whatever the purpose of the conference may be, it cannot be simply to afford philosophers an opportunity to present papers on topics that have nothing whatever to do with naturalism.

Here is a third possibility. It represents *my* guess as to the intention of the organizers of the conference: they propose to provide a forum for analytical philosophers who are anti-naturalists, to provide an occasion for the presentation of philosophical papers that presuppose or defend the denial of

naturalism. (If this is right, a better title for the conference would have been 'Analytical Philosophy with Anti-naturalism'.) Have I anything to contribute to a conference of that nature? Well, no, not really. As I have said, most of my philosophical work is simply irrelevant to naturalism. Most of my philosophical convictions, moreover, are regrettably close to those of the typical naturalist. True, I think that there are non-human immaterial intelligences, such as God and St Michael. I believe that Christ was raised from the dead, and that one day we, like him, shall be raised imperishable. I believe that every Sunday I eat and drink – in a mystery, but an entirely non-metaphorical mystery – the body and blood of Christ. These beliefs of mine are obviously inconsistent with naturalism; if I belonged to the Federated Society of People Who Are Naturalists Because They Are So Smart, and if I publicly confessed these beliefs, I should certainly be expelled. But these beliefs don't play any role in my philosophical work. It's true that I have written essays on topics like the Holy Trinity and biblical inspiration and miracles and the problem of evil. But I don't think that these essays contain much that couldn't be accepted by a naturalist. (At most, the occasional sentence.) The purpose of my essays on religious topics has always been apologetical. My purpose in each of them was to examine an attempt to demonstrate the falsity of some Christian doctrine: arguments for the incoherency of the doctrine of the Holy Trinity, for example, or various forms of the argument from evil. My conclusion in each case was that the argument under consideration was a failure. And that a certain argument for the falsity of some Christian doctrine is a failure is a conclusion that a naturalist can consistently, if not happily, accept.

If one turns from the title of the conference to the 'Presentation' that accompanies the conference materials, one finds that the subject-matter of the conference is there said to be those programs that propose to 'naturalize' some field of enquiry or some part of philosophy – to produce a naturalized epistemology, for example – or to provide naturalistic accounts of certain problematical concepts or entities or phenomena (of truth, of concepts, of intentionality). (The present paper is officially part of the 'Ontology' section of this conference, although I confess it contains little in the way of ontology, a fact I am now in the process of excusing.) There is, it seems to me, a certain tension between the title of the conference and the 'Presentation', for naturalization projects do not make up a very large part of analytical philosophy. If one opens a recent volume of any uncontroversially analytical journal to a randomly chosen page, one is not *very* likely to find on that page a part of some attempt to naturalize something. So, it would seem, a great deal of even the most recent analytical philosophy has no connection with the various naturalization projects that some analytical philosophers are engaged in. In any case, I have nothing to say about these projects. I know little of them, and what I do know of them suggests that they are vague, pretentious, amorphous, programmatic, and have not achieved anything of value. I should say, however, that, except on that ground, I am not

particularly opposed to them. (My attitude toward them is like Francis Bacon's attitude toward magic. 'Magicians', Bacon said, 'attain little to greatness and certainty of works.') In principle, I could accept the results of many of them, if they had results, although I'd probably balk at any attempt to naturalize religious belief. Suppose, for example, that someone proposed to naturalize morality. I am not thinking of someone like Michael Ruse, who has attempted to show that morality is an illusion that can be explained in naturalistic terms. I am rather thinking of someone whose project is to defend just that meta-ethical thesis that was called naturalism long before the current fashion for naturalizing things got under way: the thesis that moral properties are to be identified with certain natural properties. Grave difficulties attend this project, but, supposing them to have been overcome, I can see no objection to accepting its results.

So, this suggestion, too, the suggestion that the papers presented at this conference concern themselves with projects that propose to 'naturalize' various things offers me no prospect for contributing a paper to the conference.

What, then, am I to do? I can think of nothing to do but what my title suggests I am going to do: to try to say what naturalism is and what analytical philosophy is. If I cannot contribute directly, as it were, to what I take the purpose of this conference to be, at least I can try to say something that is relevant to that purpose.

I begin with naturalism. The word 'naturalism' obviously has something to do with 'nature'. There can, I think, be no objection to saying that a naturalist is someone who says that there is nothing besides nature, nothing in addition to nature, nothing outside or beyond nature – or no objection besides this one: it's not a very informative thing to say. For what is this 'nature' besides which, if the naturalists are right, there is nothing? The etymologies of the words '*natura*' and '*physis*' are connected with the ideas of birth and growth. And this root meaning is preserved in the usual understanding of the English word 'nature' by most people: as most people use the word 'nature', nature is something that is mostly green. (Indeed, in its primary sense, the English word 'naturalist' applies not to a person who holds a certain philosophical position but to a person who knows all about birds and flowers.) But if naturalism is the thesis that there is nothing besides nature, the word 'nature' must be understood as denoting something that is not mostly green but mostly black. (I owe the 'nature as green'/'nature as black' distinction to C.S. Lewis.) 'Nature', in the sense that interests us, is simply another name for the physical universe or the cosmos. Naturalism is, therefore, the thesis that there is nothing besides the physical universe. The physical universe is the sum total of all physical things. Naturalism is, therefore, the thesis that all things are physical things. Naturalism is, therefore, physicalism. (We should certainly be very puzzled if someone said, 'I'm not a naturalist, but I'm a physicalist' – or 'I'm not a physicalist, but I'm a naturalist.') But this is of very little help, for it

merely replaces the question, What does 'natural' mean? with the question, What does 'physical' mean? – and the latter word has come a very long way indeed from its archaic Greek genesis in the growth of plants. And what *does* 'physical' mean? Well, the word (in the sense that is of present interest) obviously has something to do with the science of physics. If we were satisfied with the current state of physics, of standard working physics, as opposed to speculative physics, we could define 'physical' this way. There are two basic kinds of physical entities: space-time and elementary particles. A physical thing is either space-time or some part of space-time or an elementary particle or something composed of elementary particles. But no one is in fact satisfied with the current state of physics, a composite of the general theory of relativity and the so-called standard theory of elementary particles. One very good reason for this is that the general theory of relativity and the standard theory of elementary particles are inconsistent with each other, although in most situations that physicists are concerned to study, the inconsistency can be walled off and ignored. A second reason is that, while the general theory of relativity is beautifully plausible, the standard theory, for all its effectiveness, contains much unlovely arbitrariness. It is hoped that one day physics will be a unified and beautiful whole, that the science of physics will be coextensive with a single theory that explains both gravity (the present business of general relativity) and everything else (the present business of the standard theory). But it is generally conceded that the ontology of this dreamt-of final theory will almost certainly be very unlike the ontologies of general relativity and the standard theory. Perhaps there will in the end be no space-time, but only particles (or something analogous to particles; maybe they won't be point-like); or perhaps there will be no particles but only space-time (particles being reduced to singularities in space-time); or perhaps there will be neither, but some third kind of thing, a kind of which we do not at present have a clear picture.

However this may be, most people would agree that we cannot plausibly define 'physical' by reference to the ontologies of the two components of present-day physics. What, then, are we to do if we wish to define 'physical'? Most thinkers who have tried to answer this question would say something like this: we must abstract from the physics we know its essential core, something that, because it is essential to physics, will survive any future transformation of physics. We can do this; and when we have done it, we shall not have something so abstract and general as to be philosophically uninteresting. We shall have a really interesting concept, and it will reveal to us the essence of or allow us plausibly to define the physical (and hence the natural and hence naturalism).

Attempts at 'abstracting the essential core' of present-day physics have been of two types: epistemological and ontological. The former finds the essence of physics in an unchanging method. Although (the proponents of the epistemological attempt say) the physics of the future will no doubt

employ concepts very different from those of the present day, physicists will always use the same methods in their investigations, and these methods will one day lead them to a final theory and this final theory will be a *physical* theory (and not something that grew out of physics but is not physics, as experimental psychology grew out of philosophy but is not philosophy) because it is historically continuous with the discipline now called 'physics', its continuity with present-day physics consisting precisely in the continued application of those unchanging methods of investigation. This final theory will postulate entities of certain sorts; it will assert their existence. We may therefore plausibly define physical things as things of the sorts that will be postulated by the final theory. Or, at any rate, we may define physical things as things that are composed of things of those sorts, for no doubt there are special contingent structures – human beings and stars, for example – whose existence will not be a consequence of the theory itself, although their existence will be a consequence of the theory plus certain boundary conditions. Naturalism, or physicalism, may be defined as the thesis that there exist only things of the sorts that will be postulated by the final theory or else things composed of those things.

I don't know what to say about this suggestion, other than to voice some suspicions. I'm not a philosopher of science, after all, and I know little of these matters. My unschooled suspicions are these: I suspect that there may be no such thing as the unchanging methods of investigation in physics that the definition depends on; and I suspect that if there are such methods, human beings, even given world enough and time, may not be capable using them to reach a final theory. (Who knows?: maybe there will be competing candidates for the office 'final theory' such that the unchanging methods of investigation will tell physicists that one could choose between these candidates only by performing an experiment that would require a linear accelerator a thousand light-years long, a device human beings will never be in a position to construct.) I do not know what others would say about this, but, for my part, I should not want the validity of the account of naturalism I endorsed to depend on the assumption that the suspicions I have voiced are ungrounded.

The ontological approach seems to me to be the more promising. It is based on the following idea: to see what is essential to physics, look at the kinds of theoretical entities postulated by physics and ask what is essential to entities of those kinds. (I use 'theoretical entity' as an epistemological, not an ontological term: theoretical entities are entities such that our reasons for believing in them are theoretical.) When we do this, we find that the following things can be said about the theoretical entities of physics with such assurance that it seems reasonable to suppose that any possible future physics will postulate only entities that have these properties. First (whatever the 'new pan-psychists', people like David Chalmers, may try to tell us), nothing in the nature of these entities has anything to do with the mental: nothing in their nature pertains to personality, consciousness, or

intentionality. Human persons and all other entities that are conscious or have intentional states are complex, special, physically contingent structures composed of the fundamental entities postulated by physics. (It seems plausible to suppose that these structures will turn out to be very large relative to the fundamental entities. But this is uncertain: perhaps the fundamental entities are not the particles or tiny loops we are inclined to expect but all-pervasive fields or some such.) The mental properties of these complex structures must supervene on the properties of the entities physics postulates, of course, but those entities themselves do not have the least tincture of the mental about them. Just as the theoretical entities of physics are in no way alive (although all living things are ultimately composed of them), they in no way think or feel. The theoretical entities of physics, moreover, have no teleological properties. Even if certain complex structures entirely composed of these entities have ends or purposes or functions (even if the function of hearts is to pump blood, for example), teleology is in no way present in the theoretical entities that are the ultimate parts of these structures. Finally, these properties are numerically quantifiable. If, for example, the theoretical entities have mass or charge (properties that may not belong to the theoretical entities of future physics), one entity may have twice the charge or one-half the mass of another. And what applies to the properties of these entities applies, *mutatis mutandis*, to the relations that they bear to one another. If, for example, one entity is separated from another by a certain space-time interval (a relation that may not figure in the theories of a future physics) – well, there is nothing mental or teleological about space-time intervals, and they have numerical measures.

Physicalism or naturalism may now be defined: nothing exists but (a) entities having such properties (non-mental, non-teleological, numerically quantifiable properties) and bearing such relations (non-mental, non-teleological, numerically quantifiable relations) to one another; and (b) composite objects that have these entities as their ultimate parts. Some among the mereological sums *may* have mental or teleological (and non-quantifiable) properties; whether they have such properties or not, their properties, *all* their properties, supervene on the properties of the entities that are their ultimate parts and the relations these ultimate parts bear to one another. The distribution of properties in any world depends on the distribution of properties among the 'ultimate' entities. As a matter of metaphysical necessity, two worlds that are alike in respect of the properties of and relations among their ultimate entities are alike in *every* respect. According to naturalism, the truth of what is expressed by any of the following sentences, supposing them to express truths, supervenes on the (non-mental, non-teleological, numerically quantifiable) properties of and the (non-mental, non-teleological, numerically quantifiable) relations that hold among the ultimate entities:

- Tokyo is the capital of Japan.
- Alice's favorite color is blue.

- It is wrong to bear false witness against one's neighbor.
- There are at least half a million species of beetles.
- It's easier for a German to learn Dutch than it is for an Italian.
- Most mathematicians have a low opinion of mathematical logic.

There is one issue that divides naturalists that I will simply note. Some naturalists are unhappy about affirming the real existence of platonic or abstract entities (such as properties or propositions or mathematical objects like numbers or functions or tensors) and some are not. Rather than take sides in this dispute, I will simply insert one additional word into my statement of naturalism: 'Nothing *concrete* exists but . . . ' This is naturalism *simpliciter*. We may then say that there are two schools of naturalism: anti-platonic naturalism, which insists that there are no platonic entities, and the more liberal school, which simply does not worry about platonic entities (which, after all, have no causal powers) and happily allows quantification over them. One might, of course, worry about what the opposed pair of terms 'concrete' and 'abstract' mean, but this is a problem that confronts everyone with ontological interests, and there is no reason to suppose that a naturalist and a non-naturalist (or an anti-platonic and a 'liberal' naturalist) would disagree about the merits of any proposed way of making the abstract/concrete distinction clear.

This, then, is naturalism. I turn now to the topic of analytical philosophy. It used to be, until quite recently I think, that the general public, insofar as they were even vaguely aware of them, referred to analytical philosophers as logical positivists, whether they were logical positivists or not, rather as some people refer to citizens of the United Kingdom of Great Britain and Northern Ireland as 'Englishmen'. And the label 'logical positivist' fits most analytical philosophers about as well as 'Englishman' fits a schoolgirl in the Scottish Highlands. Calling all analytical philosophers logical positivists is, in fact, even more misleading than calling all Britons Englishmen, for while lots of Britons are Englishmen, few if any of the analytical philosophers of the present day are logical positivists. The application of 'logical positivist' to analytical philosophers generally is important for our present concerns, because a commitment to naturalism (in the form of the following thesis, which the logical positivists shared with the nineteenth-century positivists: scientific knowledge is the only knowledge) really is an essential component of logical positivism. The logical positivists were, properly speaking, a school of philosophers that originated in Vienna in the 1920s. The members of this school were murdered or driven from the European continent by the Nazis (many of them were Jews; they were all liberals or socialists). Logical positivism in exile, so to call it, flourished briefly in the English-speaking countries till the late 1940s, and was then abandoned by its former adherents for purely philosophical reasons. (We should honor the logical positivists for this. I believe that logical positivism is the only important philosophical position to have been abandoned by its own inventors because rational argument convinced them that it was false.)

Although the widespread identification of analytical philosophy with logical positivism was an error, it was not an inexplicable error. Logical positivists *were* what came to be called analytical philosophers (just as Englishmen *are* Britons). No doubt a page of philosophy written by a Viennese logical positivist in the early 1930s and a page of philosophy written by any analytical philosopher in the early twenty-first century would seem very much the same sort of thing to those philosophers whose primary twentieth-century texts are Husserl and Heidegger. Analytical philosophy is not in any useful sense of the word a school of philosophy (like Thomism); it is rather a philosophical community, a community of philosophical discourse (like scholasticism). It is because the logical positivists (on the one hand) and present-day analytical philosophers (on the other) belong to the same community of philosophical discourse that their respective writings can seem very much the same sort of thing to members of other communities of philosophical discourse.

What, then, it is that defines this philosophical community to which both present-day analytical philosophers and the logical positivists belong? One obvious answer to this question is that, as the phrase 'analytical philosophy' suggests, this community is defined by the fact that it assigns a central place in philosophy to something called 'analysis'. Let us examine this suggestion.

The post-war popularity of the term 'analytical philosophy' was largely a product of the popularity of a certain answer to a question that troubled anglophone philosophers in the 1950s: What is it that we philosophers *do*, anyway? This was not, of course, the first time in the history of philosophy that this question had been posed. But, like all perennial philosophical questions, it had been answered in different ways in different eras in the history of philosophy. The following answer has had its advocates in every era: philosophy provides a special kind of knowledge. But this answer immediately raises a second question: How is this knowledge related to the knowledge provided by the other sciences (or disciplines, as we should say today)? In the Middle Ages it was said that philosophy was the handmaiden of theology, the queen of the sciences. When physical science had become the queen of the sciences, many said that philosophy was the handmaiden of science (in the modern sense of science): if biology and chemistry and astronomy are the branches of the tree of human knowledge (Descartes said) and physics is its trunk, philosophy comprises its roots. Others said that philosophy was not an indoor but a rather less prestigious outdoor servant of the sciences, the under-laborer of the sciences, whose job was to clear away intellectual litter that might cause science to stumble or otherwise impede its progress. A further view, perhaps not held to be inconsistent with the under-laborer view, was that it was the business of philosophy to understand the sources of human knowledge. And there was always a school that assigned a much grander role to philosophy: Spinoza and Leibniz believed, as Plato had, that philosophy could produce, that it was the

business of philosophy to produce, a kind of knowledge that would exhibit the whole of reality as a rational system. Or was the business of philosophy (as still others said) to show what we, we thinkers and knowers, must be like if we have the kinds of *a priori* and *a posteriori* knowledge that we know we have? Well, you know the story of these conceptions of the purpose or function of philosophy as well as I (in many cases, no doubt, much better). The point is: they eventually came to be seen as wholly unsatisfactory by many who reflected on the nature of philosophy. Just as philosophy came to be seen by many as unable to provide indisputable knowledge of God, freedom, and immortality (not to mention morality, knowledge, and all the other matters of concern philosophy had once claimed to provide indisputable knowledge of), philosophy came to be seen by many as unable to provide indisputable knowledge of (or indeed any halfway satisfactory account of) itself, its own nature and function.

In the 1950s, as I have said, some anglophone philosophers became interested in finding some little thing for philosophy to 'do' that would not run afoul of the claims and accomplishments of the physical and biological and behavioral sciences. (And their interest was not in finding something for philosophy to do such that it would be clear that this was what philosophy had always done. Their interest was in finding something that philosophers could do *from then on*; something for philosophers – that is, members of university faculties of philosophy – to retreat to and occupy themselves with, as one might say.) And this was their solution: The business of philosophers, the *proper* business of philosophers, the only thing philosophers should ever have been concerned with (admittedly, philosophers have had other concerns, much grander and more ambitious concerns, concerns whose disastrous consequences are amply illustrated in the history of philosophy), what philosophers do whenever they produce anything of lasting value, is the analysis of concepts. This was not the first time this thesis had been put forward; it had been a central thesis of logical positivism. The post-positivist philosophers (who were for the most part the same people who had been the positivist philosophers) retained the thesis but did not retain the theory of meaning that had led the logical positivists to maintain that the proper vocation of philosophy was the analysis of concepts. (It had led them by the scruff of the neck: the verification theory of meaning, the central dogma of logical positivism, left no possible vocation for philosophy other than the analysis of concepts, unless it were as a kind of non-cognitive poetry.) The post-positivist account of philosophy had thus no theoretical basis (according to itself, it could have had no basis but an analysis of the concept 'proper vocation of philosophy', since the question, What is the proper vocation of philosophy? is pretty evidently a philosophical question.) And it soon became evident that philosophers, even English-speaking philosophers who were paradigmatically analytical philosophers, were not going to abide by it. There were theoretical attacks on the idea; certainly the attack on the idea of analyticity in Quine's enormously influential

essay 'Two Dogmas of Empiricism' was, among other things, an attack on this conception of philosophy, for, if there is an unproblematical notion of the analysis of concepts, there is an unproblematical notion of analyticity. Much more important to my mind, however, is the fact that analytical philosophers, paradigmatically analytical philosophers, simply refused to restrict their activities to the analysis of concepts. One remembers A.J. Ayer's sad reaction to Alvin Plantinga's Oxford lectures on modality: 'I've lived in vain.' And Plantinga was hardly an isolated case: it could hardly be said that the writings of Chisholm and Kripke and David Lewis contained nothing but the analysis of concepts. In a way, this was a pity: that the proper business of philosophy is the analysis of concepts is a nice theory and a fairly clear theory as philosophical theories go. And if that is what philosophy is, it seems plausible to say that *analytical* philosophy is philosophy that, as a matter of self-conscious methodology, restricts itself to philosophy's proper business, the analysis of concepts. The only trouble with this is that it isn't true. The proper business of philosophy is not coextensive with the analysis of concepts, and many central figures of the analytical movement have demonstrated by their choices of topics and methods that they do not understand philosophy in that way.

If analytical philosophy cannot be defined by reference to its supposed method, the analysis of concepts, might it be defined by reference to its historical roots? Well, perhaps, but those roots are various and tangled. We certainly can't say that analytical philosophy is philosophy that is wholly rooted in the native philosophy of the British Isles. (I say 'the native philosophy' to take account of the fact that for a significant part of the nineteenth and the early twentieth century, British philosophy looked to Germany for its fundamental ideas.) It has for some time now been uncontroversial that analytical philosophy cannot be defined as that philosophy that has its roots in British empiricism and the work of Moore and Russell. Michael Dummett has said that analytical philosophy, which is sometimes called Anglo-American philosophy, would be better called Anglo-Austrian philosophy. There is something to this. The work of Bolzano and Brentano and Meinong is of immense importance to an understanding of the history of analytical philosophy. (The Viennese logical positivists, those of them who *were* Austrians, were not so much Austrian philosophers as philosophers who were Austrians. And, although it is unfashionable to say this, I will say it: I think that this is also true of Wittgenstein.) And of course, we must not forget the contributions of the great Polish philosophers and logicians between the wars to analytical philosophy and the important post-war contributions of Danish and Swedish and Finnish philosophers. And then there is Frege. If Germany, speaking generally, has had little influence on analytical philosophy until very recently, no one is more important to the history of analytical philosophy than Frege – not Hume, not Moore, not Russell, not even Wittgenstein. The history of analytical philosophy is so complex and involves so much historical contingency – from the point of

view of the history of *philosophy*, the Nazi control of the universities in Germany and other countries of central Europe, the Second World War, and the Soviet control of the nations of central Europe, are matters of historical contingency – that it is very hard to base any account of analytical philosophy on its history.

In the end, I think, we can say only that, although we can make some remarks about analytical philosophy, we can give no useful account of its essence – if it has an essence. We can say nothing that is as helpful to someone who wants to know what analytical philosophy is as statements like 'Analytical philosophy is philosophy that, as a matter of self-conscious methodology, restricts itself to the analysis of concepts' or 'Analytical philosophy is philosophy that has its roots in British empiricism and the work of Moore and Russell' would be if only they were true. But we can, as I say, make some remarks:

- Analytical philosophy, as a general rule, aspires to clarity of expression. This does not of course mean that the writings of analytical philosophers are always clear. Wilfrid Sellars, for example, was an analytical philosopher if anyone ever was, and he was notoriously obscure. But he was *trying* to be clear. (Still, how much help is this? Almost everyone admits that a piece of philosophy may be *unavoidably* hard to understand because the things it is about are by their very natures hard to understand. Would Heidegger not tell us that his work is not unnecessarily difficult to read, that its difficulties are necessitated by his subject matter and the Forgetfulness of Being that has pervaded the European consciousness since Plato? I have seen scattered remarks by Sartre that indicate that he believed his own philosophical prose to be extremely clear.) But, as a general rule, analytical philosophers think that philosophical sentences should be as simple as possible and that words should be used in their everyday senses or else explicitly defined.
- Analytical philosophy, insofar as it draws on anything outside philosophy, draws on formal logic and mathematics and the physical and biological sciences (and to some measure from experimental psychology and linguistics). It does not, as a general rule, draw inspiration or material from literature or art or history or the more 'humanistic' parts of the human sciences. (I do not mean to imply that this is a fact that analytical philosophers should be proud of.)
- Analytical philosophy, insofar as it involves the defense of theses, places a high value on explicit argument, on pieces of text that are identified by the author as arguments and whose validity is to be judged by the rules of logic.
- Analytical philosophers have a particularly collegial relationship with the great philosophers of history. I mean this almost literally: analytical philosophers tend to regard Plato and Occam and Descartes as *colleagues*, albeit colleagues who labor under the burden of being dead. Let me

> present the fictional Winifred to you. Winifred, although she is a fiction, is a typical analytical philosopher, and at the moment she is studying Kant's theory of freedom. To understand the relationship of this typical analytical philosopher to the history of philosophy, it is necessary to understand why she is studying Kant's theory of freedom. It is not because she is interested in Kant's philosophy *per se*; it is not because she is interested in the historical development of the idea of freedom; it is not because she wishes to study Hegel's theory of freedom and regards a mastery of Kant's theory of freedom as a prerequisite for that project; it is not because she is interested in contrasting the metaphysical understanding of freedom of a representative philosopher of the Enlightenment with a post-modern anti-metaphysical understanding of freedom. It is for a reason quite unlike any of these reasons. It is because Winifred is possessed of a childlike desire to understand the things to be found in the world, and one of these things is freedom. Her desire is that very desire mentioned in the famous first sentence of Aristotle's *Metaphysics*. She wants to understand freedom, and she is reading Kant because she thinks he may have something to teach her about this freedom. She would like to know whether Kant's theory of freedom is true or partly true or wholly false. If you tell her that her quest for the truth about freedom is indeed childlike, that it is not possible to undertake a project of that sort in today's post-something-or-other world, she will only smile at you. She will smile and ignore you or she will smile and treat what you have said as a thesis that is open to debate and proceed to present arguments against it, for all the world as if it were a thesis about second intentions and the two of you were opponents in a disputation in a medieval university.

I should add that although Winifred is a typical analytical philosopher, she does not represent the way every analytical philosopher approaches the history of philosophy. There are, of course, analytical philosophers who are historians of philosophy, and there are some reputable analytical philosophers who have no interest at all in the history of philosophy. But even those analytical philosophers who are historians have a tendency to approach the history of philosophy in ways that many non-analytical historians of philosophy regard as having the wrong aim, even a perverse aim. One non-analytical historian once directed the following charge at a former colleague of mine, Jonathan Bennett (who had written two books on Kant): 'Bennett is interested in Kant only insofar as he can extract from his writings concepts and arguments that are relevant to the philosophical controversies of the present day; he has no interest in Kant's philosophy itself.' Analytical philosophers (like the fictional Winifred or the real Jonathan Bennett) will reply that they know what it is to be interested in philosophical problems about space and time and knowledge *a priori* and causation, and they will say that they can understand why someone who was interested in these

things might be interested in what Kant had to say about them – but, they will say, they are less clear about what it is to be interested in something called 'Kant's philosophy itself'. And. they will go on to say, if there is such an interest as an interest in Kant's philosophy itself, it is not a philosophical interest. After all, what *Kant* was interested in was philosophical problems about space and time and causation, and so on; *he* never wrote about anything called 'Kant's philosophy itself'. And, although Kant tells us that the philosophy of David Hume awakened him from his dogmatic slumbers, he was not interested in 'Hume's philosophy itself'. He was, rather, interested in what Hume's philosophy was *about*. In this respect, at least, Kant was the reader Hume wrote for. Hume did not write for students of the philosophy of David Hume. He wrote for students of human nature and the human understanding. Analytical philosophers are (in this respect) like Kant and like Hume and like most other pre-Hegelian philosophers. When they read Plato or Occam or Descartes, their primary interest is not (not typically, at any rate) the philosophies of Plato or Occam or Descartes. Their primary interest lies rather in what the philosophies of Plato and Occam and Descartes were *about*.

This, then, is analytical philosophy, or as much as I am going to say about it. What is of central importance for this conference in what I have said about analytical philosophy is that, if what I have said is right, then being an analytical philosopher does not involve commitment to any philosophical doctrine. An analytical philosopher may be a platonic realist or a nominalist, may affirm or deny the freedom of the will, may believe in or deny the existence of an immaterial soul, may make the most dogmatic claims to knowledge or may embrace a thoroughgoing scepticism. An analytical philosopher may regard metaphysics as an illusion or be the most determined and ardent defender and practitioner of metaphysics imaginable. A philosopher may take any position on any philosophical question and still be an analytical philosopher in good standing. And this generalization applies to positions about the reality of a supernatural order. An analytical philosopher may (as I do) recite the Nicene Creed with conviction every Sunday or regard the Nicene Creed as an absurd vestige of a pernicious and dying superstition: nothing in the nature of analytical philosophy lends any support whatever to either of these positions. Analytical philosophy, by its very nature, is neither the friend nor the enemy of supernaturalism. And, by the same token, analytical philosophy can be neither the friend nor the enemy of naturalism.

Naturalism, physicalism, and some notes on 'analytical philosophy'

Comment on van Inwagen's paper

Christian Kanzian

In his lecture, Peter van Inwagen gives answers to the questions of what 'naturalism' and 'analytical philosophy' are. His main point is that there is no inner connection between being an analytical philosopher and being a naturalist. Of course, most analytical philosophers today are naturalists. Likewise, most naturalists are analytical philosophers. This fact, however, is a contingent matter. Analytical philosophy, as such, does not commit itself to naturalism and naturalism does not commit itself to analytical philosophy.

The core thesis of naturalism is that there is nothing in addition to, outside of or beyond nature. Insofar as 'nature' is just another name for the physical universe, naturalism is physicalism: all things are physical things. The project of defining 'physical' by reference to the ontology of the two components of present-day physics, i.e. space-time and elementary particles, is implausible. Therefore, either epistemologically or ontologically considered, physicalists apply to the 'essential core' of present-day physics. The former find the essence of physics in a methodological kernel of physics and the postulate of an ultimate or final physical theory. Thus, physicalists postulate what I like to call a 'physicist's paradise': there is one physical world formula and every truth outside the formula can be deduced from it. The ontological approach looks at the essences of theoretical entities postulated by physics. In both cases, the result is a universe without mental entities, with nothing pertaining to consciousness, with no intentionality, etc. The mental sphere is what Armstrong (1997) calls an 'ontological free lunch'. It supervenes on the physical basis.

Although analytical philosophy has its roots in logical positivism and the philosophy of the 'linguistic turn', it cannot be defined by reference to these roots. Furthermore, it is not simply an Anglo-American matter. Thus, analytical philosophy cannot be defined as the opposite of continental philosophy. As mentioned before, analytical philosophy cannot be explained by a canonical list of philosophical presuppositions, convictions or theories. That does not mean that it is entirely impossible to distinguish between analytical and non-analytical philosophy. According to van Inwagen, analytical philosophy aspires to clarity of expression. That means, philosophical thought should be expressed as simply as possible. Words should be used in

their everyday sense or be explicitly defined. Analytical philosophy draws on formal logic and, outside of philosophy, on mathematical, physical and biological sciences. Analytical philosophy also places a high value on explicit argument, the validity of which may be judged by the rules of logic. And, last but not least, analytical philosophers stand in a 'collegial relationship' with the great philosophers of history: they do not restrict themselves to historical positions *per se*, or to an analysis of the historical genesis of a position. They deal with historical authors insofar as they help them to understand philosophical matters clearer and better, and insofar as they are relevant to contemporary debates.

My problem, as commentator on van Inwagen's paper, is very similar to his own problem when he asks himself: 'What should I say here?' I am at a loss because I really agree with everything that van Inwagen says. Yes, it is one of my own basic philosophical convictions that there is no inner or philosophical connection between analytical philosophy and naturalism. Yes, it is true: naturalists believe in the myth of a physicist's paradise. They believe that every non-physical phenomenon can be reduced to the postulated physical basis and claim (without justification) that everything not reducible to this basis can be declared an 'ontological free lunch'. Yes, analytical philosophy is a forum for people, who try to solve philosophical problems in expressions, which are as clear as possible. Analytical philosophers distinguish between valid and invalid arguments and use logic as an important means of doing so. Analytical philosophers have indeed a special, I would say, systematic way of communicating with the great authors of the history of philosophy.

So, for lack of disagreement, I have nothing to criticize in Peter van Inwagen's paper. Nevertheless I think it makes sense to add some notes to his contribution.

One possible footnote to van Inwagen's contribution could be added to his analysis of 'naturalism'. Van Inwagen's conclusion is that naturalism is (nothing else than) physicalism. I have nothing to say against his argumentation. But, there are authors, like David Armstrong and others, who stress the difference between naturalism and physicalism (Armstrong 1997: 5–6). And I do think (irrespective of Armstrong's concrete line of argumentation) that there are good reasons to distinguish between the two positions. In its original meaning, 'physicalism' stands for the thesis that all meaningful sentences can be translated into sentences of a universal physical language. Carnap and other members of the 'Vienna Circle' explicitly formulated this thesis in an anti-ontological context: ontological sentences are not transferable into physical language and are therefore meaningless. Today, most physicalists combine the thesis of a physical universal-language with ontology. The result is a very special ontological point of view: all things are physical things; to exist is to be the value of bound variables in sentences of the physical universal-language. Presupposed is the thesis of a completed physics, which provides the principles of all other sciences. In my own words: the physicist's paradise.

'Naturalism' may be used in a wider, more liberal sense. Arthur Danto, for instance, explains naturalism as the position that only such things exist, which can be described and explained by the means of *some* (!) positive scientific research (Danto 1967). This scientific research does not have to be physics, but can also be biology or another of the natural sciences. Philosophical explanations of life functions, for example, or – referring to life functions from a biological background – of the diachronic identity of organisms, may be called 'naturalistic' but not 'physicalistic'. A naturalist is not committed to the belief in a completed physical science nor in the universal validity of physical language. Not all meaningful sentences need to be transferable into physical sentences. It is not the case that, in the end, the principles of all sciences are physical principles. I do not maintain that all naturalists are that liberal. I do, however, maintain that naturalism can be so interpreted. The advantage of understanding this distinction is knowing that one can indeed justifiably claim to be a naturalist without being a physicalist. Moreover, with this distinction in place, the confession of being a faithful physicalist can be understood in a more precise and informative sense.

A second note can be added to van Inwagen's analysis of analytical philosophy. As mentioned before, I agree with van Inwagen that the scope of analytical philosophy is much wider than just the philosophy of the linguistic turn. It is wrong to say that analytical philosophy is nothing other than conceptual analysis. However, the claim that analytical philosophy has to do with the philosophical analysis of language, especially semantic analysis, must, according to my understanding, be added to a general characterization of analytical philosophy. Analytical philosophers take semantic considerations into account. Beginning with these considerations, they try to formulate their theories in close connection with them. Analytical philosophers accept the accordance with semantic considerations as one possible criterion of their theories' adequacy. Analytical ethics, in contrast to non-analytical ethics, reflects upon the semantic analysis of terms like 'good', 'bad', 'norm', etc. Analytical philosophers of religion analyze the logic of our speech about God. Analytical ontologists start with considerations about the logic of sortal terms for instance, before they make theories about sorts, about the untranslatability of indexicals, before they theorize about the (ir-)reducibility of the subjective perspective, and so on. That does not mean that analytical philosophers restrict themselves to semantics or that they take semantics as the only means for working out and testing a philosophical theory.

There is a third note, which I suggest should be added to van Inwagen's contribution. It pertains to the understanding of the conference's general topic, 'Analytical Philosophy without Naturalism?'. I think that the question cannot be understood as being merely the question about the (non-existent) inner connection between analytical philosophy and naturalism. It can also be understood as a question about the 'how'; how does analytical

philosophy exist without naturalism? How can one be an analytical philosopher while avoiding the presupposition of naturalistic theorems? And, what can be a genuine analytical criticism of naturalistic philosophy? In reference to what I have said before, I ask myself how to be an analytical, non-physicalistic, or better, anti-physicalistic, philosopher. Indeed, I think that Peter van Inwagen's concept of 'naturalism' is closer to my concept of 'physicalism' than it is to what I called 'naturalism in a wide and liberal sense'. The first typical analytical criticism of physicalism is based on the postulate of avoiding philosophical myths. That this postulate is typically analytical need not be explained. It follows from the general characterization of analytical philosophy given by van Inwagen, just to mention again the high value of explicit argumentation. I think that physicalistic philosophers believe in something, which is nothing else than a myth. It is the myth of the physicist's paradise. One reason for its being a myth is that, while physicalists refer to physics (they believe in physics), there is no serious physicist who is looking for the ultimate world formula, the physical principles of all other sciences, the reducibility of all true sentences to a physical language, etc. Serious physicists are not physicalists. They even deny what physicalists assert. Thus, the physicalist's belief is unjustified; it is a belief in a myth. To point this out is a characteristically analytical contribution to the debate.

A second analytical postulate is the following: no one, who is willing to argue, who is willing to obey logical rules, who is willing to use the technical terminology of analytical philosophy, and so forth, should be excluded from philosophical disputation. 'Do not be esoteric!' is a short formulation of this postulate. I know experts in esoterics, who told me that some esoterics are physicalists. I cannot assess that claim. But what I can assess is that physicalism is in principle an esoteric position. How should somebody, who does not believe in the physicist's paradise, be able to lead a real philosophical debate with a faithful member of the physicalist community?

Finally, I want to mention a third postulate of analytical philosophy: there is a typical openness for all philosophical topics and disciplines. That this is an analytical postulate follows from the denial that analytical philosophy is simply restricted to the philosophy of language as well as the characterization of analytical philosophy as an open forum for argumentative philosophical debate. Physicalists are not open to all philosophical topics and disciplines. Ethics, for instance, cannot be expressed in a universal, physicalistic language. Physicalists normally concede this point. Therefore, ethics must either be eliminated from philosophy or physicalized, which in turn leads to the same end. The same holds for aesthetics, social philosophy, philosophy of politics and so on. A friend of mine teaching at Boston College told me that they once invited Quine to a workshop on topics of philosophy of religion. The only comment that Quine really gave was, 'I have nothing to say about that.' I think that Quine is right. He indeed has nothing to say about that. And, I think that that is a pity. It is not a pity for religion, but

rather for physicalism as a philosophical position. I only speak about 'physicalism' as a rather extreme position. What about naturalism? If naturalists do not accept my liberal conception of their position, then they really are, as Peter van Inwagen says, nothing else than physicalists. My criticism pertains to them too. If the naturalists do accept my conception of their position, then, I confess, I must deal with them in a special way. How would that be? The answer is too much to explain on this occasion.

Part III

Philosophy of religion

5 Contemporary cosmology and the existence of God

William Lane Craig

The fundamental question

In his biography of Ludwig Wittgenstein, Norman Malcolm reports:

> He said that he sometimes had a certain experience which could best be described by saying that 'when I have it, *I wonder at the existence of the world*.' I am then inclined to use such phrases as 'How extraordinary that anything should exist!' or 'How extraordinary that the world should exist!'
> (1958: 56)

This mystery, which, according to Aristotle, lay at the very root of philosophy,[1] is one which thoughtful naturalists cannot avoid. Derek Parfit, for example, agrees that 'No question is more sublime than why there is a Universe: why there is anything rather than nothing' (1998: 24).

As we all know, this question led G.W. Leibniz to posit the existence of a necessary being, which carries within itself the sufficient reason for its existence and which constitutes the sufficient reason for the existence of everything else in the world (Leibniz 1951: 415, 237–239). Leibniz identified this being as God. Naturalists, on the other hand, have typically claimed that the space-time universe is itself at least factually necessary – that is to say, eternal, uncaused, incorruptible, and indestructible[2] – while dismissing the demand for a logically necessary being. Thus, David Hume queried, 'Why may not the material universe be the necessarily existent Being?' (1947: 190). Indeed, 'How can anything, that exists from eternity, have a cause, since that relation implies a priority in time and a beginning of existence?' (ibid.: 190). There is no warrant for going beyond the universe to posit a supernatural ground of its existence. As Bertrand Russell put it so succinctly in his BBC radio debate with Frederick Copleston, 'The universe is just there, and that's all' (Russell and Copleston 1964: 175).

The origin of the universe

It was thus the presumed eternity of the universe that allowed naturalistic minds to rest comfortably in the face of the mystery of existence. This

feeling of *Gemütlichkeit* was first disturbed when, in 1917, Albert Einstein made a cosmological application of his newly discovered gravitational theory, the General Theory of Relativity (GR) (Einstein 1952). In so doing he assumed that the universe is homogeneous and isotropic and that it exists in a steady state, with a constant mean mass density and a constant curvature of space. To his chagrin, however, he found that GR would not permit such a model of the universe unless he introduced into his gravitational field equations a certain 'fudge factor' Λ in order to counterbalance the gravitational effect of matter and so ensure a static universe. But Einstein's universe was balanced on a razor's edge, and the least perturbation would cause the universe either to implode or to expand. By taking this feature of Einstein's model seriously, the Russian mathematician Alexander Friedman and the Belgian astronomer Georges Lemaître were able in the 1920s independently to formulate solutions to the field equations which predicted an expanding universe (Friedman 1922; Lemaître 1927).

The monumental significance of the Friedman–Lemaître model lay in its historicization of the universe. As one commentator has remarked, up to this time the idea of the expansion of the universe 'was absolutely beyond comprehension. Throughout all of human history the universe was regarded as fixed and immutable and the idea that it might actually be changing was inconceivable' (Naber 1988: 126–127). But if the Friedman–Lemaître model were correct, the universe could no longer be adequately treated as a static entity existing, in effect, timelessly. Rather the universe has a history, and time will not be a matter of indifference for our investigation of the cosmos.

In 1929, the astronomer Edwin Hubble showed that the red-shift in the optical spectra of light from distant galaxies was a common feature of all measured galaxies and was proportional to their distance from us (Hubble 1929). This red-shift, first observed by Vesto Slipher in 1926, was taken to be a Doppler effect indicative of the recessional motion of the light source in the line of sight. Incredibly, what Hubble had discovered was the isotropic expansion of the universe predicted by Friedman and Lemaître on the basis of Einstein's GR. It was a veritable turning point in the history of science. 'Of all the great predictions that science has ever made over the centuries,' exclaims John Wheeler, 'was there ever one greater than this, to predict, and predict correctly, and predict against all expectation a phenomenon so fantastic as the expansion of the universe?' (1980: 354).

The standard Big Bang model

According to the Friedman–Lemaître model, as time proceeds, the distances separating material particles become greater. It is important to understand that as a GR-based theory, the model does not describe the expansion of the material content of the universe into a pre-existing, empty, Newtonian space, but rather the expansion of space itself. The ideal particles of the

cosmological fluid constituted by the matter and energy of the universe are conceived to be at rest with respect to space but to recede progressively from one another as space itself expands or stretches, just as buttons glued to the surface of a balloon would recede from one another as the balloon inflates. As the universe expands, it becomes less and less dense. This has the astonishing implication that as one reverses the expansion and extrapolates back in time, the universe becomes progressively denser until one arrives at a state of infinite density at some point in the finite past. This state represents a singularity at which the space-time curvature, along with temperature, pressure, and density, becomes infinite. It therefore constitutes an edge or boundary to space-time itself. P.C.W. Davies comments:

> If we extrapolate this prediction to its extreme, we reach a point when all distances in the universe have shrunk to zero. An initial cosmological singularity therefore forms a past temporal extremity to the universe. We cannot continue physical reasoning, or even the concept of space-time, through such an extremity. For this reason most cosmologists think of the initial singularity as the beginning of the universe. On this view the big bang represents the creation event; the creation not only of all the matter and energy in the universe, but also of spacetime itself. (1978: 78–79)

The term 'Big Bang', originally a derisive expression coined by Fred Hoyle to characterize the beginning of the universe predicted by the Friedman–Lemaître model, is thus potentially misleading, since the expansion cannot be visualized from the outside (there being no 'outside', just as there is no 'before' with respect to the Big Bang).[3]

The standard Big Bang model, as the Friedman–Lemaître model came to be called, thus describes a universe which is not eternal in the past, but which came into being a finite time ago. Moreover – and this deserves underscoring – the origin it posits is an absolute origin *ex nihilo*. For not only all matter and energy, but space and time themselves come into being at the initial cosmological singularity. As Barrow and Tipler emphasize, 'At this singularity, space and time came into existence; literally nothing existed before the singularity, so, if the Universe originated at such a singularity, we would truly have a creation *ex nihilo*' (1986: 442). Thus, we may graphically represent space-time as a cone (Figure 5.1).

On such a model the universe originates *ex nihilo* in the sense that at the initial singularity it is true that *There is no earlier space-time point* or it is false that *Something existed prior to the singularity.*

Now such a conclusion is profoundly disturbing for anyone who ponders it. For the question cannot be suppressed: *Why does the universe exist rather than nothing?* Sir Arthur Eddington, contemplating the beginning of the universe, opined that the expansion of the universe was so preposterous and incredible that 'I feel almost an indignation that anyone should believe in

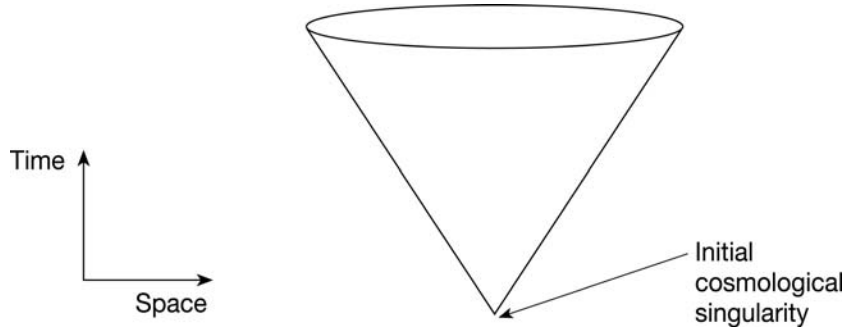

Figure 5.1 Conical representation of Standard Model of space-time. Space and time begin at the initial cosmological singularity, before which literally nothing exists.

it – except myself' (1933: 124). He finally felt forced to conclude, 'The beginning seems to present insuperable difficulties unless we agree to look on it as frankly supernatural' (ibid.: 178).

Standard Big Bang cosmogony thus presents a challenge to scientific naturalism, since, in Quentin Smith's words:

> It belongs analytically to the concept of the cosmological singularity that it is not the effect of prior physical events. The definition of a singularity . . . entails that it is *impossible to extend the spacetime manifold beyond the singularity* . . . This rules out the idea that the singularity is an effect of some prior natural process.
>
> (1993a: 120)

Smith recognizes that the question which then remains is whether the Big Bang might not plausibly be regarded as the result of a supernatural cause. Otherwise, one must say that the universe simply sprang into being uncaused out of absolutely nothing. Thus, in the words of one astrophysical team, 'The problem of the origin involves a certain metaphysical aspect which may be either appealing or revolting' (Reeves *et al*. 1973: 912).

The Steady State Model

Revolted by the stark metaphysical alternatives presented by an absolute beginning of the universe, naturalists have been understandably eager to subvert the Standard Model and restore an eternal universe. Sir Fred Hoyle, for example, could countenance neither an uncaused nor a supernaturally caused origin of the universe. With respect to the singularity, he wrote, 'This most peculiar situation is taken by many astronomers to represent *the origin of the universe.* The universe is supposed to have begun at this particular time. From where? The usual answer, surely an unsatisfactory one, is: from nothing!' (Hoyle 1975a: 165). Equally unsatisfactory was the postulation

of a supernatural cause. Noting that some accept happily the universe's absolute beginning, Hoyle complained:

> To many people this thought process seems highly satisfactory because a 'something' outside physics can then be introduced at $\tau = 0$. By a semantic manoeuvre, the word 'something' is then replaced by 'god,' except that the first letter becomes a capital, God, in order to warn us that we must not carry the enquiry any further.
>
> (1975b: 658)

To Hoyle's credit, he did carry the inquiry further by helping to formulate the first competitor to the Standard Model. In 1948, Hoyle, together with Hermann Bondi and Thomas Gold, broached the Steady State Model of the universe (Bondi and Gold 1948; Hoyle 1948). According to this theory, the universe is in a state of isotropic cosmic expansion, but as the galaxies recede, new matter is drawn into being *ex nihilo* in the interstices of space created by the galactic recession (Figure 5.2).

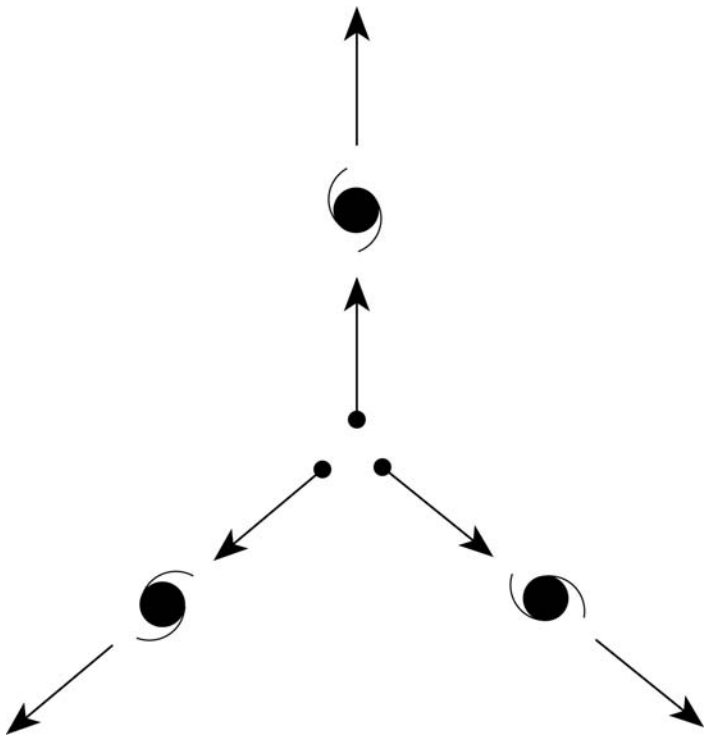

Figure 5.2 Steady State Model. As the galaxies mutually recede, new matter comes into existence to replace them; the universe thus constantly renews itself and so never began to exist.

The expansion of the universe in the Steady State Model can be compared to a rubber sheet with buttons glued to it: as the sheet is stretched and the buttons separate, new buttons come into being in the voids created by the recession of the previously existing buttons. Thus, the condition of the sheet remains constant over time, and no beginning of the process need be posited. If one extrapolates the expansion of the universe back in time, the density of the universe never increases because the matter and energy simply vanish as the galaxies mutually approach!

The Steady State theory never secured a single piece of experimental verification; its appeal was purely metaphysical.[4] The discovery of progressively more radio galaxies at ever greater distances undermined the theory by showing that in the past the universe was significantly different than it is today, thus contradicting the notion of a steady state of the universe. Instead it became increasingly evident that the universe had an evolutionary history. But the decisive refutation of the Steady State Model came with two discoveries which constituted, in addition to the galactic red-shift, the most significant evidence for the Big Bang theory: the cosmogonic nucleosynthesis of the light elements and the microwave background radiation. Although the heavy elements were synthesized in the stellar furnaces, stellar nucleosynthesis could not manufacture the abundant light elements such as helium and deuterium. These could only have been created in the extreme conditions present in the first moment of the Big Bang. In 1965 a serendipitous discovery revealed the existence of a cosmic background radiation predicted in the 1940s by George Gamow on the basis of the Standard Model. This radiation, now shifted into the microwave region of the spectrum, consists of photons emitted during a very hot and dense phase of the universe. In the minds of almost all cosmologists, the cosmic background radiation decisively discredited the Steady State Model.

Oscillating models

The Standard Model was based on the assumptions of homogeneity and isotropy. In the 1960s and the 1970s some cosmologists suggested that by denying homogeneity and isotropy, one might be able to craft an Oscillating Model of the universe and thus avert the absolute beginning predicted by the Standard Model (Lifschitz and Khalatnikov 1963: 207). If the internal gravitational pull of the mass of the universe were able to overcome the force of its expansion, then the expansion could be reversed into a cosmic contraction, a Big Crunch. If the universe were not homogeneous and isotropic, then the collapsing universe might not coalesce at a point, but the material contents of the universe might pass by one another, so that the universe would appear to bounce back from the contraction into a new expansion phase. If this process could be repeated indefinitely, then an absolute beginning of the universe might be avoided (Figure 5.3).

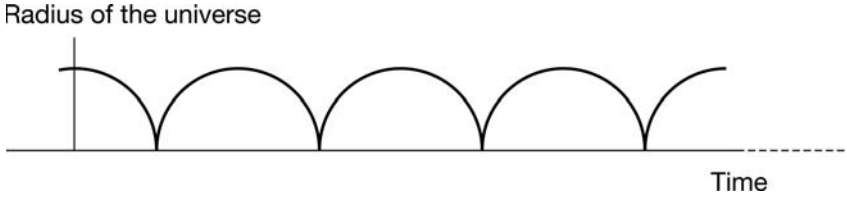

Figure 5.3 Oscillating Model. Each expansion phase is preceded and succeeded by a contraction phase, so that the universe in concertina-like fashion exists beginninglessly and endlessly.

Such a theory is extraordinarily speculative, but again there were metaphysical motivations for adopting this model.[5] The prospects of the Oscillating Model were severely dimmed in 1970, however, by Roger Penrose and Stephen Hawking's formulation of the Singularity Theorems which bear their names (Penrose 1965: 57–59; Hawking and Penrose 1973: 266). The theorems disclosed that under very generalized conditions an initial cosmological singularity is inevitable, even for inhomogeneous and non-isotropic universes. Reflecting on the impact of this discovery, Hawking notes that the Hawking–Penrose Singularity Theorems 'led to the abandonment of attempts (mainly by the Russians) to argue that there was a previous contracting phase and a non-singular bounce into expansion. Instead, almost everyone now believes that the universe, and time itself, had a beginning at the big bang' (Hawking and Penrose 1996: 20).

Despite the fact that no space-time trajectory can be extended through a singularity, the Oscillating Model exhibited a stubborn persistence. Three further strikes were lodged against it. First, there are no known physics which would cause a collapsing universe to bounce back to a new expansion. If, in defiance of the Hawking–Penrose Singularity Theorems, the universe rebounds, this is predicated upon a physics which is completely unknown. Physics predicts that a universe in a state of gravitational self-collapse will not rebound like a basketball dropped to the floor, but rather land like a lump of clay (Guth and Sher 1983: 505–506; Bludman 1984: 319–322). Second, the observational evidence indicates that the mean mass density of the universe is insufficient to generate enough gravitational attraction to halt and reverse the expansion. Tests employing a variety of techniques for measuring the density of the universe and the deceleration of the expansion continue to point to a density below the critical value. In January 1998 astronomical teams from Princeton, Yale, the Lawrence Berkeley National Laboratory, and the Harvard–Smithsonian Astrophysics Institute reported at the American Astronomical Society meeting that their various tests all show that 'the universe will expand forever'.[6] A spokesman for the Harvard–Smithsonian team stated that they were now at least 95 per cent certain that 'the density of matter is insufficient to halt the expansion of the universe'.[7] This effectively rules out an oscillating universe. Third, the thermodynamic properties of an Oscillating Model imply the very beginning its proponents

sought to avoid. Entropy increases from cycle to cycle in such a model, which has the effect of generating larger and longer oscillations with each successive cycle (Figure 5.4).

Thus, as one traces the oscillations back in time, they become progressively smaller until one reaches a first and smallest oscillation. Novikov and Zeldovich conclude, 'The multicycle model has an infinite future, but only a finite past' (1973: 401–402). In fact, astronomer Joseph Silk estimates on the basis of current entropy levels that the universe cannot have gone through more than 100 previous oscillations (1989: 311–312). Even if this difficulty were avoided (Hochberg *et al*. 1999), a universe oscillating from eternity past would require an infinitely precise tuning of initial conditions in order to perdure through an infinite number of successive bounces. A universe rebounding from a single, infinitely long contraction is, if entropy increases during the contracting phase, thermodynamically untenable and incompatible with the initial low entropy condition of our expanding phase; postulating an entropy decrease during the contracting phase in order to escape this problem would require us to postulate inexplicably special low entropy conditions at the time of the bounce in the life of an infinitely evolving universe. Such a low entropy condition at the beginning of the expansion is more plausibly accounted for by the presence of a singularity or some sort of quantum creation event.

Although these difficulties were well known, proponents of the Oscillating Model tenaciously clung to it until a new alternative to the Standard Model emerged during the 1970s.[8] Looking back, quantum cosmologist Christopher Isham muses:

> Perhaps the best argument in favor of the thesis that the Big Bang supports theism is the obvious unease with which it is greeted by some atheist physicists. At times this has led to scientific ideas, such as continuous creation or an oscillating universe, being advanced with a tenacity which so exceeds their intrinsic worth that one can only suspect the operation of psychological forces lying very much deeper than the usual academic desire of a theorist to support his/her theory.
>
> (1988: 378)

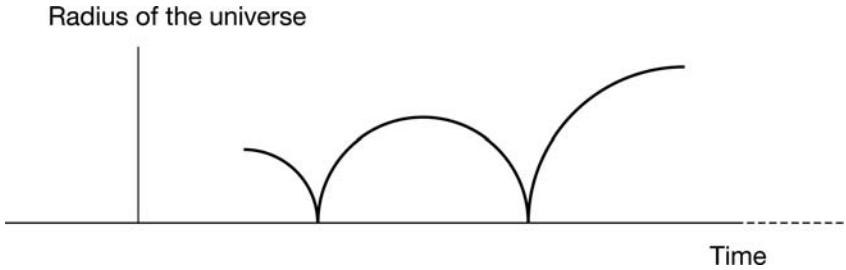

Figure 5.4 Oscillating Model with entropy increase. Due to the conservation of entropy each successive oscillation has a larger radius and longer expansion time.

The Oscillating Model drew its life from its avoidance of an absolute beginning of the universe; but once other models became available claiming to offer the same benefit, the Oscillating Model sank into oblivion under the weight of its own deficiencies.

Vacuum fluctuation models

It was realized that a physical description of the universe prior to the Planck time (10^{-43} second after the Big Bang singularity) would require the introduction of quantum physics in addition to GR. On the quantum level so-called virtual particles are thought to arise due to fluctuations in the energy locked up in the vacuum, particles which the Heisenberg Indeterminacy Principle allows to exist for a fleeting moment before dissolving back into the vacuum. In 1973, Edward Tryon speculated whether the universe might not be a long-lived virtual particle, whose total energy is zero, born out of the primordial vacuum (1973: 396–397). This seemingly bizarre speculation gave rise to a new generation of cosmogonic theories which we may call Vacuum Fluctuation Models. These models were closely related to an adjustment to the Standard Model known as Inflation. In an attempt to explain – or explain away, depending on one's viewpoint – the astonishing large-scale homogeneity and isotropy of the universe, certain theorists proposed that between 10^{-35} and 10^{-33} sec after the Big Bang singularity, the universe underwent a phase of super-rapid, or inflationary, expansion which served to push the inhomogeneities out beyond our event horizon (Guth 1981). Prior to the Inflationary Era the universe was merely empty space, or a vacuum, and the material universe was born when the vacuum energy was converted into matter via a quantum mechanical phase transition. In most inflationary models, as one extrapolates backward in time, beyond the Planck time, the universe continues to shrink down to the initial singularity. But in Vacuum Fluctuation Models, it is hypothesized that prior to inflation the Universe-as-a-whole was not expanding. This Universe-as-a-whole is a primordial vacuum which exists eternally in a steady state. Throughout this vacuum sub-atomic energy fluctuations constantly occur, by means of which matter is created and mini-universes are born (Figure 5.5).

Our expanding universe is but one of an indefinite number of mini-universes conceived within the womb of the greater Universe-as-a-whole. Thus, the beginning of our universe does not represent an absolute beginning, but merely a change in the eternal, uncaused Universe-as-a-whole.

Though still bandied about in the popular press, Vacuum Fluctuation Models did not outlive the decade of the 1980s. Not only were there theoretical problems with the production mechanisms of matter, but these models faced a deep internal incoherence (Isham 1988: 385–387). According to such models, it is impossible to specify precisely when and where a fluctuation will occur in the primordial vacuum which will then grow into a universe. Within any finite interval of time there is a positive probability

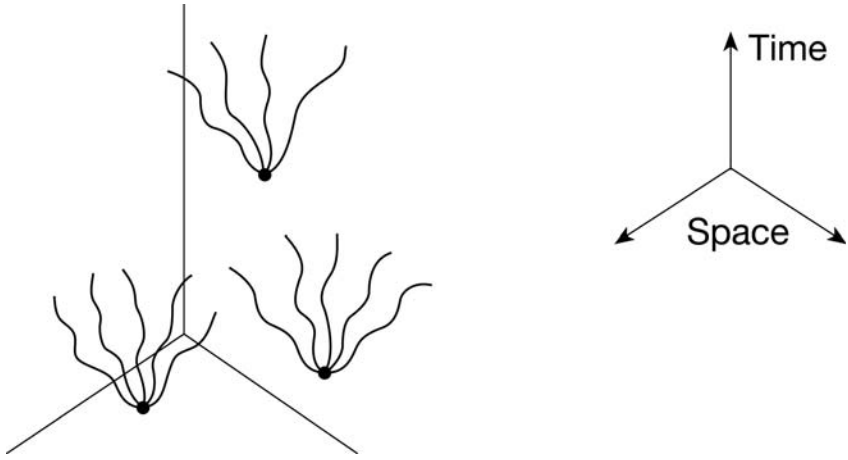

Figure 5.5 Vacuum Fluctuation Models. Within the vacuum of the wider universe, fluctuations occur that grow into mini-universes. Ours is but one of these, and its relative beginning does not imply a beginning for the Universe-as-a-whole.

of such a fluctuation occurring at any point in space. Thus, given infinite past time, universes will eventually be spawned at *every* point in the primordial vacuum, and, as they expand, they will begin to collide and coalesce with one another. Thus, given infinite past time, we should by now be observing an infinitely old universe, not a relatively young one. One theorist tries to avoid this problem by stipulating that fluctuations in the primordial vacuum only occur infinitely far apart, so that each mini-universe has infinite room in which to expand (Gott 1982: 304–307). Not only is such a scenario unacceptably *ad hoc*, but it does not even solve the problem. For given infinite past time, each of the infinite regions of the vacuum will have spawned an open universe which by now will have entirely filled that region, with the result that all of the individual mini-universes would have coalesced.

Isham has called this problem 'fairly lethal' to Vacuum Fluctuation Models and says that therefore they 'have not found wide acceptance' (Isham 1990). About the only way to avert the problem would be to postulate an expansion of the primordial vacuum itself; but then we are right back to the absolute origin implied by the Standard Model. According to Isham, these models were therefore 'jettisoned twenty years ago' and 'nothing much' has been done with them since (Isham 1994).

Chaotic inflationary model

Inflation also forms the context for the next alternative we shall consider: the Chaotic Inflationary Model. Inflationary theory has not only been criticized as unduly 'metaphysical', but has also been crippled by various physical

problems (such as getting inflation to transition to the current expansion). We have seen the Old Inflationary Theory and the New Inflationary Theory, both of which are now dead. One of the most fertile of the inflation theorists has been the Russian cosmologist Andrei Linde, who currently champions his Chaotic Inflationary Model (Linde 1984, 1983). According to cosmologist Robert Brandenberger, 'Linde's chaotic inflation scenario is ... the only viable inflationary model in the sense that it is not plagued with internal inconsistencies (as "old inflation" and "new inflation" are)'.[9] In Linde's model, inflation *never* ends: each inflating domain of the universe when it reaches a certain volume gives rise via inflation to another domain, and so on, *ad infinitum* (Figure 5.6).

Linde's model thus has an infinite future. But Linde is troubled at the prospect of an absolute beginning. He writes, 'The most difficult aspect of this problem is not the existence of the singularity itself, but the question of what was *before* the singularity ... This problem lies somewhere at the boundary between physics and metaphysics' (Linde 1984: 976). Linde therefore proposes that chaotic inflation is not only endless, but beginningless. Every domain in the universe is the product of inflation in another domain, so that the singularity is averted and with it as well the question of what came before (or, more accurately, what caused it).

In 1994, however, Arvind Borde and Alexander Vilenkin showed that a universe eternally inflating toward the future cannot be geodesically complete in the past, that is to say, there must have existed at some point in the indefinite past an initial singularity. They write:

> A model in which the inflationary phase has no end ... naturally leads to this question: Can this model also be extended to the infinite past, avoiding in this way the problem of the initial singularity?

Figure 5.6 Chaotic Inflationary Model. The wider universe produces via inflation separate domains that continue to recede from one another as the wider space expands.

> . . . this is in fact not possible in future-eternal inflationary spacetimes as long as they obey some reasonable physical conditions: such models must necessarily possess initial singularities.
>
> . . . the fact that inflationary spacetimes are past incomplete forces one to address the question of what, if anything, came before.
>
> (Borde and Vilenkin 1994: 3305, 3307)

In his response, Linde concurs with the conclusion of Borde and Vilenkin: there must have been a Big Bang singularity at some point in the past (Linde *et al*. 1994). Therefore, inflationary models, like their predecessors, failed to avert the beginning predicted by the Standard Model.

Quantum Gravity Models

At the close of their analysis of Linde's Chaotic Inflationary Model, Borde and Vilenkin say with respect to Linde's metaphysical question, 'The most promising way to deal with this problem is probably to treat the Universe quantum mechanically and describe it by a wave function rather than by a classical spacetime' (Borde and Vilenkin 1994: 3307). They thereby allude to the next class of models which we shall discuss, namely, Quantum Gravity Models. Vilenkin and, more famously, James Hartle and Stephen Hawking have proposed models of the universe which Vilenkin candidly calls exercises in 'metaphysical cosmology'.[10] In his best-selling popularization of his theory, Hawking even reveals an explicitly theological orientation. He concedes that on the Standard Model one could legitimately identify the Big Bang singularity as the instant at which God created the universe (Hawking 1988: 9). Indeed, he thinks that a number of attempts to avoid the Big Bang were probably motivated by the feeling that a beginning of time 'smacks of divine intervention' (ibid.: 46). He sees his own model as preferable to the Standard Model because there would be no edge of space-time at which one 'would have to appeal to God or some new law' (ibid.: 136). As we shall see, he is not at all reluctant to draw theological conclusions on the basis of his model.

Both the Hartle–Hawking and the Vilenkin models eliminate the initial singularity by transforming the conical hyper-surface of classical space-time into a smooth, curved hyper-surface having no edge (Figure 5.7).

This is accomplished by the introduction of imaginary numbers for the time variable in Einstein's gravitational equations, which effectively eliminates the singularity. Hawking sees profound theological implications in the model:

> The idea that space and time may form a closed surface without boundary . . . has profound implications for the role of God in the affairs of the universe . . . So long as the universe had a beginning, we

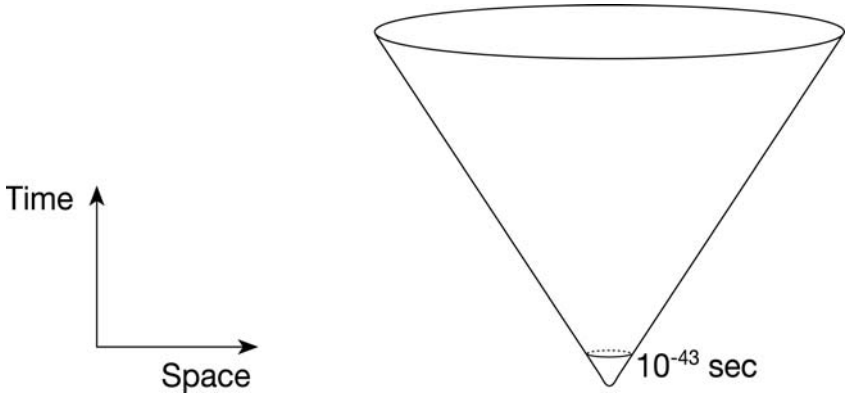

Figure 5.7 Quantum Gravity Model. In the Hartle–Hawking version, space-time is 'rounded off' prior to the Planck time, so that although the past is finite, there is no edge or beginning point.

> could suppose it had a creator. But if the universe is really completely self-contained, having no boundary or edge, it would have neither beginning nor end. What place, then, for a creator?
>
> (Hawking 1988: 140–141)

Hawking does not deny the existence of God, but he does think his model eliminates the need for a Creator of the universe.

The key to assessing this theological claim is the physical interpretation of Quantum Gravity Models. By positing a finite (imaginary) time on a closed surface prior the Planck time rather than an infinite time on an open surface, such models actually seem to support, rather than undercut, the idea that time had a beginning. Such theories, if successful, enable us to model the beginning of the universe without an initial singularity involving infinite density, temperature, pressure, and so on. As Barrow points out, 'This type of quantum universe has not always existed; it comes into being just as the classical cosmologies could, but it does not start at a Big Bang where physical quantities are infinite' (1991: 68). Barrow points out that such models are 'often described as giving a picture of "creation out of nothing"', the only caveat being that in this case 'there is no definite . . . point of creation' (ibid.: 67–68). Hartle–Hawking themselves construe their model as giving 'the amplitude for the Universe to appear from nothing', and Hawking has asserted that according to the model the universe 'would quite literally be created out of nothing: not just out of the vacuum, but out of absolutely nothing at all, because there is nothing outside the universe' (Hartle and Hawking 1983: 2961; Hawking and Penrose 1996: 85). Similarly, Vilenkin claims that his model describes the creation of the universe 'from literally *nothing*' (Vilenkin 1982: 26). Taken at face value, these statements entail the beginning of the universe. Hence, Hawking means to

include himself when he asserts that 'almost everyone now believes that the universe, and time itself, had a beginning at the Big Bang' (Hawking and Penrose 1996: 20). Hawking's statement quoted above concerning the theological implications of his model must therefore be understood to mean that on such models there are no beginning or ending *points.* But having a beginning does not entail having a beginning point. Even in the Standard Model, theorists sometimes 'cut out' the initial singular point without thinking that therefore space-time no longer begins to exist and the problem of the origin of the universe is thereby resolved. Time begins to exist just in case for any finite temporal interval, there are only a finite number of equal temporal intervals earlier than it. That condition is fulfilled for Quantum Gravity Models as well as for the Standard Model. Nor should we think that by giving the amplitude for the universe to appear from nothing quantum cosmologists have eliminated the need for a Creator, for that probability is conditional upon several choices which only the Creator could make (such as selecting the wave function of the universe) and is dubiously applied to absolute nothingness.[11] Thus, Quantum Gravity Models, like the Standard Model, imply the beginning of the universe.

Perhaps it will be said that such an interpretation of Quantum Gravity Models fails to take seriously the notion of 'imaginary time'. Introducing imaginary numbers for the time variable in Einstein's equation has the peculiar effect of making the time dimension indistinguishable from space. But in that case, the imaginary time regime prior to the Planck time is not a space-time at all, but a Euclidean four-dimensional space. Construed realistically, such a four-space would be evacuated of all temporal becoming and would simply exist timelessly. Thus, Vilenkin characterizes this regime as a 'state in which all our basic notions of space, time, energy, entropy, etc. lose their meaning' (1983: 2851). Hawking describes it as 'completely self-contained and not affected by anything outside itself. It would be neither created nor destroyed. It would just BE' (Hawking 1988: 136).

The question which arises for this construal of the model is whether such an interpretation is meant to be taken realistically or instrumentally. On this score, there can be little doubt that the use of imaginary quantities for time is a mere mathematical device without ontological significance. For, first, there is no intelligible physical interpretation of imaginary time on offer. What, for example, would it mean to speak of the lapse of an imaginary second or of a physical object's enduring through two imaginary minutes? Second, time is metaphysically distinct from space, its moments being ordered by an *earlier than* relation which does not similarly order points in space. But this essential difference is obscured by imaginary time. Thus, 'imaginary time' is most plausibly construed as a mathematical *Hilfsmittel*. Barrow observes:

> physicists have often carried out this 'change time into space' procedure as a useful trick for doing certain problems in ordinary quantum

> mechanics, although they did not imagine that time was *really* like space. At the end of the calculation, they just swop back into the usual interpretation of there being one dimension of time and three ... dimensions of ... space.
>
> (1991: 66–67)

In his model, Hawking simply declines to re-convert to real numbers. If we do, then the singularity re-appears. Hawking admits, 'Only if we could picture the universe in terms of imaginary time would there be no singularities ... When one goes back to the real time in which we live, however, there will still appear to be singularities' (Hawking 1988: 138–139). Hawking's model is thus a way of re-describing a universe with a singular beginning point in such a way that that singularity is transformed away; but such a re-description is not realist in character.

Remarkably, Hawking has more recently stated explicitly that he interprets the Hartle–Hawking Model non-realistically. He confesses, 'I'm a positivist ... I don't demand that a theory correspond to reality because I don't know what it is' (Hawking and Penrose 1996: 121). Still more extreme, 'I take the positivist view point that a physical theory is just a mathematical model and that it is meaningless to ask whether it corresponds to reality'.[12] In assessing the worth of a theory, 'All I'm concerned with is that the theory should predict the results of measurements'.[13] The clearest example of Hawking's instrumentalism is his combination of an electron quantum tunneling in Euclidean space (with time being imaginary) and an electron/positron pair accelerating away from each other in Minkowski space-time (ibid.: 53–55). This analysis is directly analogous to the Hartle–Hawking cosmological model; and yet no one would construe particle pair creation as literally the result of an electron transitioning out of a timelessly existing four-space into our classical space-time. It is just an alternative description employing imaginary numbers rather than real numbers.

Significantly, the use of imaginary quantities for time is an inherent feature of *all* Quantum Gravity Models.[14] This precludes their being construed realistically as accounts of the origin of the space-time universe in a timelessly existing four-space. Rather they are ways of modeling the real beginning of the universe *ex nihilo* in such a way as not to involve a singularity. What brought the universe into being remains unexplained on such accounts.

Moreover, we are not without positive reasons for affirming the reality of the singular origin of space-time postulated by the Standard Model. John Barrow has rightly cautioned that:

> one should be wary of the fact that many of the studies of quantum cosmology are motivated by the desire to avoid an initial singularity of infinite density, so they tend to focus on quantum cosmologies that avoid a singularity at the expense of those that might contain one.
>
> (Barrow 1994: 113)

Noting the same tendency, Roger Penrose states, 'I have gradually come around to the view that it is actually misguided to ask that the space-time singularities of classical relativity should disappear when standard techniques of quantum (field) theory are applied to them' (1982: 4). For if the initial cosmological singularity is removed, then 'we should have lost what seems to me to be the best chance we have of explaining the mystery of the second law of thermodynamics' (ibid.: 5). What Penrose has in mind is the remarkable fact that as one goes back in time the entropy of the universe steadily decreases. Just how unusual this is can be demonstrated by means of the Bekenstein–Hawking formula for the entropy of a stationary black hole. The total observed entropy of the universe is 10^{88}. Since there are around 10^{80} baryons in the universe, the observed entropy per baryon must be regarded as extremely small. By contrast, in a collapsing universe the entropy would be 10^{123} near the end. Comparison of these two numbers reveals how absurdly small 10^{88} is compared to what it might have been. Thus, the structure of the Big Bang must have been severely constrained in order that thermodynamics as we know it should have arisen. So how is this special initial condition to be explained? According to Penrose, we need the initial cosmological singularity, conjoined with the Weyl Curvature Hypothesis, according to which initial singularities (as opposed to final singularities) must have vanishing Weyl curvature.[15] In standard models, the Big Bang does possess vanishing Weyl curvature. The geometrical constraints on the initial geometry have the effect of producing a state of very low entropy. So the entropy in the gravitational field starts at zero at the Big Bang and gradually increases through gravitational clumping. The Weyl Curvature Hypothesis thus has the time asymmetric character necessary to explain the second law. By contrast, the Hartle–Hawking model 'is very far from being an explanation of the fact that past singularities have small Weyl curvature whereas future singularities have large Weyl curvature' (Hawking and Penrose 1996: 129). On Hawking's time symmetrical theory, we should have white holes spewing out material, in contradiction to the Weyl Curvature Hypothesis, the second law of thermodynamics, and probably also observation (Hawking and Penrose 1996: 130). Penrose supplies the following figure (Figure 5.8) to illustrate the difference.

If we remove the initial cosmological singularity, we render the Weyl Curvature Hypothesis irrelevant and 'we should be back where we were in our attempts to understand the origin of the second law' (Penrose 1982: 5). Could the special initial geometry have arisen sheerly by chance in the absence of a cosmic singularity? Penrose's answer is decisive:

> Had there not been any constraining principles (such as the Weyl curvature hypothesis) the Bekenstein–Hawking formula would tell as that the probability of such a 'special' geometry arising by chance is at least as small as about one part in $10^{1000B(3/2)}$ where B is the present baryon number of the universe ($\approx 10^{80}$).
>
> (Penrose 1982: 5)

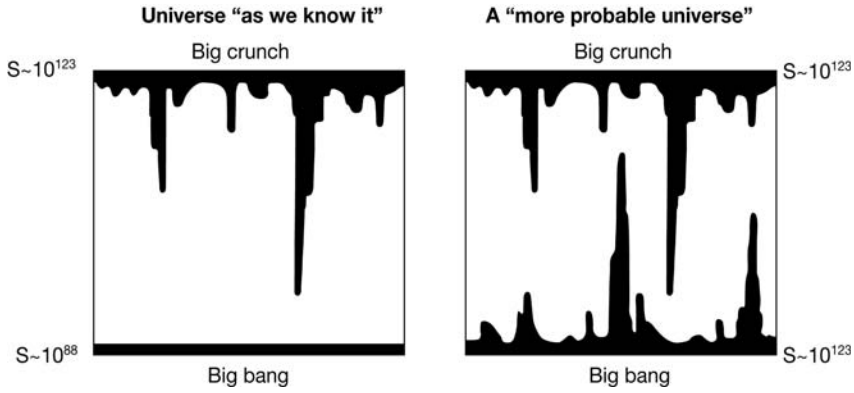

Figure 5.8 Contrast between the universe as we know it (assumed for convenience to be closed) with a more probable universe. In both cases, the Big Crunch is a high entropy ($\approx 10^{123}$), complicated, unconstrained singularity. For the left-hand picture the Big Bang is a low entropy ($<10^{88}$), highly constrained, initial singularity, while for the right-hand picture it is an unconstrained, much more probable Big Bang. The 'stalactites' represent singularities of black holes, while the 'stalagmites' represent singularities of white holes.

Thus Penrose calculates that, aiming at a phase space whose regions represent the likelihood of various possible configurations of the universe, 'the accuracy of the Creator's aim' would have to have been one part in $10^{10(123)}$ in order for our universe to exist (Penrose 1981: 249; Hawking and Penrose 1996: 34–35). He comments, 'I cannot even recall seeing anything else in physics whose accuracy is known to approach, even remotely, a figure like one part in $10^{10\ (123)}$' (Penrose 1981: 249).

Thus, the initial cosmological singularity may be a virtual thermodynamical necessity. But whether it was at a singular point or not, the fact that the universe began to exist remains a prediction of any realistic interpretation of Quantum Gravity models.

Ekpyrotic models

We come finally to the extreme edge of cosmological speculation: string cosmology. These models are based on an alternative to the standard quark model of elementary particle physics. So-called string theory (or M-theory) conceives of the fundamental building blocks of matter to be, not particles like quarks, but tiny vibrating strings of energy. String theory is so complicated and embryonic in its development that all its equations have not yet even been stated, much less solved. But that has not deterred some cosmologists from trying to craft cosmological models based on concepts of string theory to try to avert the beginning predicted by standard Big Bang cosmology.

The most celebrated of these scenarios in the popular press has been the so-called ekpyrotic scenario championed by Paul Steinhardt.[16] In the most recent revision, the cyclic ekpyrotic model, we are asked to envision two three-dimensional membranes (or 'branes' for short) existing in a five-dimensional space-time (Figure 5.9). One of these branes is our universe. These two branes are said to be in an eternal cycle in which they approach each other, collide, and retreat again from each other. It is the collision of the other brane with ours that causes the expansion of our universe. With each collision, the expansion is renewed. Thus, even though our three-dimensional universe is expanding, it never had a beginning.

Now apart from its speculative nature the ekpyrotic scenario is plagued with problems.[17] For example, the Horava–Witten version of string theory on which the scenario is based requires that the brane on which we live has a positive tension. But in the ekpyrotic scenario it has a negative tension in contradiction to the theory. Attempts to rectify this have been unsuccessful. Second, the model requires an extraordinary amount of *ad hoc* fine tuning. For example, the two branes have to be so perfectly aligned that even at a distance of 10^{30} greater than the space between them, they cannot deviate from being parallel by more than 10^{-60}. There is no explanation at all for this extraordinary set-up. Third, the collapsing and retreating branes are the equivalent of a 4-D universe which goes through an eternal cycle of contractions and expansions. In this sense, the cyclic ekpyrotic model is just the old oscillating model writ large in five dimensions. As such, it faces exactly the same problem as the original: there is no way for the universe to pass through a singularity at the end of each cycle to begin a new cycle and no

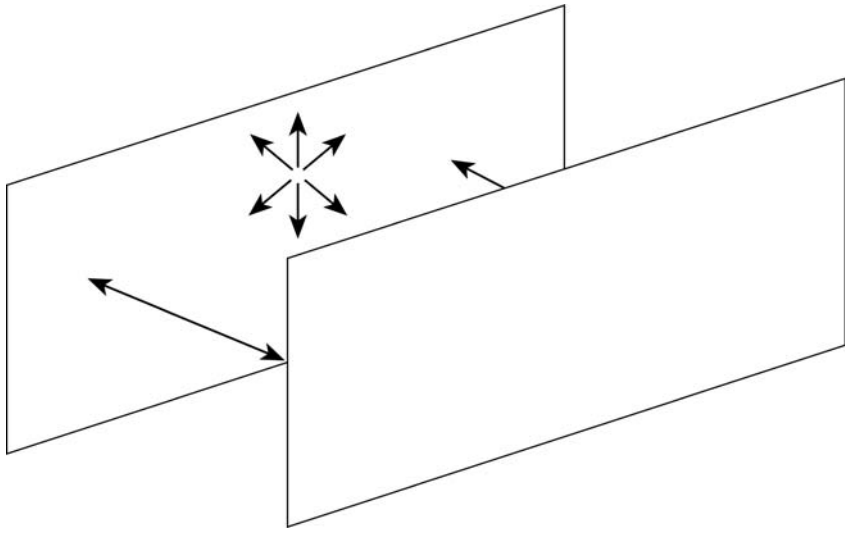

Figure 5.9 Two three-dimensional membranes in an eternal cycle of approach, collision, and retreat. With each collision the expansion of our universe is re-invigorated.

physics to cause a non-singular bounce. Finally, even if the branes could bounce back, there is no means of the physical information in one cycle being carried through to the next cycle, so that the ekpyrotic scenario has been unable to deliver on its promises to explain the large-scale structure of the observable universe. These are just some of the problems afflicting the model. It is no wonder that Andrei Linde has recently complained that while the cyclic ekpyrotic scenario is 'very popular among journalists', it has remained 'rather unpopular among scientists' (Linde 2002: 8).

But let all that pass. Perhaps all these problems can be somehow solved. The more important point is that it turns out that, like the chaotic inflationary model, the cyclic ekpyrotic scenario cannot be eternal in the past. In September of 2001 Borde and Vilenkin, in cooperation with Alan Guth, were able to generalize their earlier results on inflationary models in such a way to extend their conclusion to other models. Specifically, they note, 'Our argument can be straightforwardly extended to cosmology in higher dimensions', specifically brane-cosmology.[18] According to Vilenkin, 'It follows from our theorem that the cyclic universe is past-incomplete',[19] that is to say, the need for an initial singularity has not been eliminated. Therefore, such a universe cannot be past-eternal.

Summary

With each successive failure of alternative cosmogonic theories to avoid the absolute beginning of the universe predicted by the Standard Model, that prediction has been corroborated. This beginning of the universe, of space and time themselves, reveals the contingency of the universe. The universe is evidently not necessarily existent, as Hume suggested, since it is not eternal, and therefore its existence does cry out for explanation. It is no longer sufficient to dismiss this problem with a shrug and a slogan, 'The universe is just there, and that's all.'

Of course, in view of the metaphysical issues raised by the prospect of a beginning of the universe, we may be confident that the quest to avert such a beginning will continue unabated.[20] Such efforts are to be encouraged, and we have no reason to think that such attempts at falsification will result in anything other than further corroboration of the prediction of a beginning. In the meantime, the beginning cannot be wished away. Given its origin *ex nihilo*, the demand why the universe exists rather than nothing presses insistently upon us.

Beyond the Big Bang

The alternatives before us

The discovery that the universe is not eternal but had a beginning implies that the universe is not necessary in its existence and therefore has its

ground in a transcendent being. The only way of avoiding this conclusion would be to deny Leibniz's conviction that whatever exists has a reason for its existence, either in the necessity of its own nature or else in an external ground, and to claim that the universe simply sprang into being uncaused out of nothing. Reflecting upon the current situation, P.C.W. Davies muses,

> 'What caused the big bang?' ... One might consider some supernatural force, some agency beyond space and time as being responsible for the big bang, or one might prefer to regard the big bang as an event without a cause. It seems to me that we don't have too much choice. Either ... something outside of the physical world ... or ... an event without a cause.
>
> (1995: 8–9)

J. Richard Gott and Li-Xin Li seek to break this dilemma by defending the extraordinary hypothesis that *the universe created itself*. Observing that 'The question of first-cause has been troubling to philosophers and scientists alike for over two thousand years,' they note that modern scientists have, like Aristotle, found models of the universe attractive which involve the universe's sempiternal existence, since in this way 'one would not have to ask what caused it to come into being' (Gott and Li-Xin 1998: 023501-1). 'Now that it is apparent that our universe began in a big bang explosion', however, 'models with a finite beginning have taken precedence' and 'the question of what happened before the big bang arises' (ibid.: 023501-1). They observe that inflation seemed to be 'a very promising answer, but as Borde and Vilenkin have shown, the inflationary state preceding the big bang could not have been infinite in duration – it must have had a beginning also. Where did it come from? Ultimately, the difficult question seems to be how to make something out of nothing' (Gott and Li-Xin 1998: 023501-1). Gott and Li-Xin, however, suggest instead that we should ask whether anything in the laws of physics would prevent the universe from creating itself.

Noting that general relativity allows for the possibility of closed time-like curves, they hypothesize that as we trace the history of the universe back through an original inflationary state, we encounter a region of closed time-like curves prior to inflation. According to one possible scenario, a metastable vacuum inflates, producing an infinite number of (Big Bang type) bubble universes. In many of these a number of bubbles of metastable vacuum are created at late times by high energy events. These bubbles usually collapse and form black holes, but occasionally one will tunnel to create an expanding, metastable vacuum or baby universe. One of these expanding, metastable vacuum baby universes 'turns out to be the original inflating metastable vacuum we began with' (Figure 5.10).

Gott and Li-Xin conclude that 'the laws of physics may allow the universe to be its own mother' (ibid.: 023501-1).

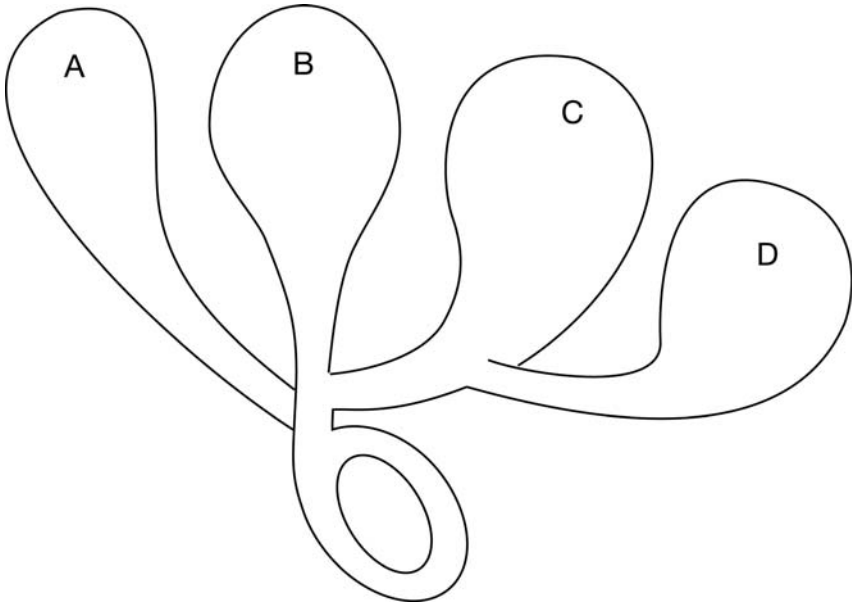

Figure 5.10 A self-creating universe. Four inflating baby universes are shown. Universes A and D have not created any baby universes. Universe C has created universe D. Universe B has created three universes: A, C, and itself, B. The torus-shaped region at the bottom is a region of closed time-like curves. Such a universe neither arose from a singularity nor tunneled from nothing, but it created itself.

Now we may leave it to the physicists to assess Gott and Li-Xin's claim that the laws of physics permit such a scenario, as well as the question of whether there are non-lawlike physical facts which contradict it. For the Gott–Li-Xin hypothesis raises fundamental metaphysical issues about the nature of time which, I think, render their hypothesis either metaphysically impossible or else superfluous.

Philosophers of time have distinguished two different views about the nature of time, which have been called the A- and the B-theories of time respectively.[21] According to the A-theory, temporal moments may be classed as past, present, and future, and only that moment which is present exists. Past moments and the things or events which occupy them have passed away and no longer exist; future moments, things, and events have not yet come to be and so do not yet exist. On the A-theory of time things come into and go out of being, and thus temporal becoming is a real and objective feature of reality.

By contrast, on the B-theory of time the distinction between past, present, and future is a subjective illusion of human consciousness. All things or events in time are equally real and existent, and moments, things, and events merely stand to one another in tenseless relations of *earlier than*,

simultaneous with, or *later than*. Nothing ever comes into or goes out of being, and temporal becoming is an illusion.

Now all instances of causal influence over the past – whether we are talking about closed time-like curves, time travel, retro-causation, tachyonic anti-telephones, or whatever – presuppose the truth of the B-theory of time.[22] For clearly on the A-theory of time, at the time at which the effect is present, the cause is future and therefore literally non-existent. Thus, the effect just comes into being from nothing. Not only does this scenario seem absurd, but it also reduces to the first horn of Davies' dilemma with respect to the origin of the universe. The universe just came uncaused from nothing.

Thus the Gott–Li-Xin hypothesis presupposes the B-theory of time. But if one presupposes such a view of time, then Gott and Li-Xin's hypothesis becomes superfluous. For on a B-theory of time the universe never truly comes into being at all.[23] The whole four-dimensional space-time manifold just exists tenselessly, and the universe has a beginning only in the sense that a meter-stick has a beginning prior to the first centimeter. Although the space-time manifold is intrinsically temporal in that one of its four dimensions is time, nonetheless it is extrinsically timeless, in that it does not exist in an embedding hyper-time but exists tenselessly, neither coming into nor going out of being. The four-dimensional space-time manifold is in this latter sense eternal. Thus, there is no need for the device of causal loops or closed time-like curves at the beginning to explain how it came into being.

Now space does not permit me to review the arguments for and against the A- and B-theories of time. I have explored this fascinating debate for the last dozen years and report my findings elsewhere.[24] Here I can only outline my reasons for affirming an A-theory of time as the most plausible view of the matter:

1 Arguments for the A-theory
   (a) Linguistic tense, which is ineliminable and irreducible, mirrors the tensed facts which are characteristic of reality.[25]
   (b) The experience of temporal becoming, like our experience of the external world, should be regarded as veridical.[26]
2 Refutation of arguments against the A-theory
   (a) McTaggart's Paradox is based upon the illicit assumption that there should exist a unique tenseless description of reality, as well as the illicit conflation of A-theoretic becoming with a B-theoretic ontology.[27]
   (b) The passage of time is not a myth, but a metaphor for the objectivity of temporal becoming, a notion which can be consistently explicated on a presentist metaphysic.[28]
3 Refutation of arguments for the B-theory
   (a) Temporal becoming is wholly compatible with the mathematical core of Relativity Theory, even if its affirmation requires a different physical interpretation than the received view.[29]

(b) Time, as it plays a role in physics, is an abstraction of a richer metaphysical reality, omitting indexical elements such as the 'here' and the 'now' in the interest of universalizing the formulations of natural laws.[30]

4 Arguments against the B-theory
   (a) In the absence of objective distinctions between past, present, and future, the relations ordering events on the B-theory are only gratuitously regarded as genuinely temporal relations of *earlier/later than*.[31]
   (b) The subjective illusion of temporal becoming involves itself an objective temporal becoming of contents of consciousness.[32]
   (c) The B-theory entails perdurantism, the view that objects have spatio-temporal parts, a doctrine which is metaphysically counter-intuitive, incompatible with moral accountability, and entails the bizarre counterpart theory of transworld identity.[33]

Given the truth of the A-theory of time, the idea that the universe is self-created, that is to say, brought itself into being via closed time-like curves, is either metaphysically impossible or else reduces to the notion that the universe sprang into existence uncaused out of nothing. Thus, I think that we are stuck with Davies' dilemma: the beginning of the universe is either an event without a cause or it is the result of a supernatural agency.

The supernaturalist alternative

Suppose we go the route of postulating some causal agency beyond space and time as being responsible for the origin of the universe. A conceptual analysis of what properties must be possessed by such an ultra-mundane cause enables us to recover a striking number of the traditional divine attributes. For as the cause of space and time, this entity must transcend space and time and therefore exist atemporally and non-spatially, at least *sans* the universe. This transcendent cause must therefore be changeless and immaterial, since timelessness entails changelessness, and changelessness implies immateriality. Such a cause must be beginningless and uncaused, at least in the sense of lacking any antecedent causal conditions. Occam's Razor will shave away further causes, since we should not multiply causes beyond necessity. This entity must be unimaginably powerful, since it created the universe out of nothing.

Finally, and most strikingly, such a transcendent cause is plausibly to be regarded as personal. As Swinburne points out, there are two types of causal explanation: scientific explanations in terms of laws and initial conditions and personal explanations in terms of agents and their volitions (1991: 32–48). A first state of the universe cannot have a scientific explanation, since there is nothing before it, and therefore it can be accounted for only in terms of a personal explanation. Moreover, the personhood of the cause of

the universe is implied by its timelessness and immateriality, since the only entities we know of which can possess such properties are either minds or abstract objects, and abstract objects do not stand in causal relations. Therefore the transcendent cause of the origin of the universe must be of the order of mind. This same conclusion is also implied by the origin of a temporal effect from a timeless cause. For if the cause of the universe were an impersonal set of necessary and sufficient conditions, it could not exist without its effect. The only way for the cause to be timeless and changeless but for its effect to originate *de novo* a finite time ago is for the cause to be a personal agent who freely chooses to bring about an effect without antecedent determining conditions. Thus, we are brought, not merely to a transcendent cause of the universe, but to its personal creator.

Naturalistic objections

The naturalist, of course, will be reluctant to take on board such metaphysical baggage. But think of the alternative: that the universe came into being uncaused out of nothing. That seems metaphysically absurd. The naturalist philosopher of science Bernulf Kanitscheider remonstrates, 'If taken seriously, the initial singularity is in head-on collision with the most successful ontological commitment that was a guiding line of research since Epicurus and Lucretius,' namely, *out of nothing nothing comes,* which Kanitscheider calls 'a metaphysical hypothesis which has proved so fruitful in every corner of science that we are surely well-advised to try as hard as we can to eschew processes of absolute origin' (1990: 344). Mario Bunge thinks that an absolute origin of the universe 'would be unscientific, for science abides by the principles that nothing comes out of nothing or turns into nothingness ... and that everything happens according to law rather than miracles' (1985: 238–239). On the basis of the first principle Bunge, like Kanitscheider, rejects the view that the universe came into being uncaused out of nothing. On the basis of the second principle he thinks to reject theism. But while the principle that *out of nothing nothing comes* is a first principle of metaphysics as well as science, there is no incompatibility between being a theist metaphysically and a methodological naturalist scientifically; moreover, even methodological naturalism is far from unchallengeable.[34] It is difficult to see how any sensible person, particularly the naturalist, can think that the universe just sprang into existence uncaused out of nothing.

It has therefore been remarkable to observe in recent years the number of naturalists who, under the force of the evidence for an absolute beginning of the universe, have embraced the view that the universe is a surd contingent, something which popped into existence uncaused out of nothing. Quentin Smith declares, 'The fact of the matter is that the most reasonable belief is that we came from nothing, by nothing and for nothing.'[35] Rather than posit a cause of the origin of the universe, Smith advises, 'We should instead

acknowledge our foundation in nothingness and feel awe at the marvelous fact that we have a chance to participate briefly in this incredible sunburst that interrupts without reason the reign of non-being' (1993a: 135).

Sometimes attempts are made to render this remarkable hypothesis more plausible, but these are usually not very impressive. Consider, for example, Peter Atkins's account of the origin of the universe:

> Now we go back in time beyond the moment of creation, to when there was no time, and to where there was no space ... In the beginning there was nothing ... By chance there was a fluctuation, and a set of points, emerging from nothing, ... defined a time ... From absolute nothing, absolutely without intervention, there came into being rudimentary existence ... Yet the line of time collapsed, and the incipient universe evaporated, for time alone is not rich enough for existence. Time and space emerged elsewhere, but they too crumbled back into their own dust, the coalescence of opposites, or simply nothing. Patterns emerged again, and again, and again. Each time the pattern formed a time, and through their patterning into time, the points induced their own existence ... Sometimes chance patterned points into a space as well as a time ... Then, by chance, there came about our fluctuation. Points came into existence by constituting time but, this time, in this pattern time was accompanied by three dimensions of space ... with them comes stability, later elements, and still later elephants.
>
> (Atkins 1992: 129, 149–151)

This account is so obviously incoherent in postulating time before time and so confused in its reification of mathematical entities that we may rightly dismiss it as the pseudo-scientific drivel that it is.[36]

Or again, when John Gribbin asserts that the origin of the universe from nothing presents no problem, since the positive energy associated with mass is precisely offset by the negative energy associated with gravitation, so that in the case of the origin of the universe we got 'Not something for nothing, after all, but *nothing* for nothing' (Gribbin 1986: 374), he commits himself to the absurd position that nothing exists (not even he himself!). At the very best, the fact that the universe contains counter-balancing amounts of positive and negative energy could show that the universe need not have a material cause; but it does nothing to obviate the need for an efficient cause. As Isham puts it, there is still the 'need for ontic seeding' to produce the positive and negative energy, even if on balance it is naught (1994: 8). That is why the quantum vacuum was needed as a substratum in cosmogonic theories postulating such a process.

More often naturalistic thinkers have sought to commend their view either by attacking the causal principle *whatever begins to exist has a cause* or else by arguing for the implausibility or incoherence of the existence of a

cause of the universe. Attacks on the causal principle are usually based on an appeal to quantum indeterminacy. For example, virtual particles are sometimes said to constitute a counter-example to the principle because they spring uncaused from the quantum mechanical vacuum. Wholly apart from the disputed question as to whether virtual particles really exist at all,[37] the central point to be understood here is that the quantum vacuum on which they depend for their existence is not nothing. It is for that reason that the statements frequently made with respect to Vacuum Fluctuation Models that 'the universe quantum tunneled into being out of nothing', or that 'nothingness is unstable' to fluctuations which grow into universes, or that 'the universe is a free lunch' because in this case 'we got something for nothing' cannot be taken seriously, for they treat nothing as though it were something, a sort of substance possessing properties and governed by the laws of quantum physics. In fact, such statements turn out to be just rhetorical flourishes which no informed scientist takes literally. The quantum vacuum, which underlies all of space-time reality, is a fluctuating sea of energy. Because the vacuum is a physical entity existing in space and time, Vacuum Fluctuation Models did not envision a genuine origin of the universe out of nothing, as Kanitscheider emphasizes:

> The violent microstructure of the vacuum has been used in attempts to explain the origin of the universe as a long-lived vacuum fluctuation. But some authors have connected with this legitimate speculations [*sic*] far-reaching metaphysical claims, or at most they couched their mathematics in a highly misleading language, when they maintained 'the creation of the universe out of nothing' . . .
>
> From the philosophical point of view it is essential to note that the foregoing is far from being a spontaneous generation of everything from naught, but the origin of that embryonic bubble is really a causal process leading from a primordial substratum with a rich physical structure to a materialized substratum of the vacuum. Admittedly this process is not deterministic, it includes that weak kind of causal dependence peculiar to every quantum mechanical process.
>
> (1990: 346–347)

Thus, quantum physics does not serve to rebut the principle that whatever begins to exist has a cause.

It is not surprising that naturalists should attack the notion of a cause of the universe, since they reject supra-natural realities independently of their motivation to justify an uncaused origin of the universe from nothing. Sometimes these critiques may be easily dismissed. For example, metaphysician John Post obviously begs the question when he claims that there cannot be a cause of the origin of the universe, since 'by definition the universe contains everything there is or ever was or will be' (Post 1991: 85). Again it is an obvious *non-sequitur* when he infers that because 'the singularity

cannot be caused by some earlier *natural* event or process,' therefore 'the universe has an uncaused beginning' and 'it seems contemporary physical cosmology cannot be cited in support of the idea of a *divine* cause or creator of the universe' (ibid.: 87).

On the other hand, Smith realizes that the metaphysician must take seriously the 'more difficult question' of 'whether or not the singularity or the Big Bang probably is an effect of a supernatural cause, God' (Smith 1993a: 120). What problems, then, are there with a supernaturalist perspective? Adolf Grünbaum has argued vigorously against what he styles 'the New Creation Argument' for a supernatural cause of the origin of the universe.[38] His basic *Ansatz* is based on the assumption that causal priority implies temporal priority. Since there were no instants of time prior to the Big Bang, it follows that the Big Bang cannot have a cause.[39]

It seems to me that the supernaturalist has a number of options for dealing with this objection, one of which is to hold that the transcendent cause of the universe is causally, but not temporally, prior to the Big Bang event, such that His act of causing the universe to begin to exist is simultaneous, or co-incident, with its beginning to exist. Grünbaum provides no justification for his assumption that causal priority implies temporal priority. Discussions of causal directionality deal routinely with cases in which cause and effect are simultaneous. A supernaturalist could hold that the Creator *sans* the universe exists changelessly and, hence, timelessly and at the Big Bang singularity created the universe along with time and space. For the Creator *sans* the universe, there simply is no time because there are no events of any sort; time begins with the first event, at the moment of creation.

The time of the first event would be not only the first time at which the universe exists, but also, technically, the first time at which God exists, since *sans* the universe God exists timelessly.[40] The moment of creation is, as it were, the moment at which God enters time. His act of creation is thus simultaneous with the origination of the universe.

In response to this suggestion, Grünbaum has opposed the following argument:[41]

1 The proponent of simultaneous, asymmetric causation must furnish a generally accepted criterion for distinguishing one of two causally connected simultaneous events as the cause of the other, if simultaneous, asymmetric causation is possible.
2 There is no generally accepted account of causal directionality.
3 Therefore, there can be no simultaneous, asymmetric cause of the Big Bang.

The argument, if successful, would eliminate all purported instances of simultaneous, asymmetric causation, not just a cause of the Big Bang.

The argument, however, is, I think, unsound because (1) is so obviously false. For (i) Why must the proponent of simultaneous, asymmetric causation

furnish a *generally accepted* criterion of causal directionality in order for such causation to be possible? Is this not an extravagant demand? Grünbaum fails to appreciate that there is no generally accepted account of causal directionality *überhaupt*, including accounts which appeal to temporal priority as a condition of causal priority. Indeed, I should dare to say that there is no generally accepted account of causation at all today. But should we therefore infer that causation is impossible or non-existent? Compare the situation in contemporary epistemology. There is today no generally accepted account of justification or rational warrant with respect to beliefs we hold to be true; but should we therefore infer that knowledge is impossible? Deconstructionists and other post-modernists may think so, but I doubt that Grünbaum would be ready to follow in their train. There is no reason to think that the possibility of simultaneous causation depends upon our being able to come up with an uncontroversial criterion of causal directionality. (ii) Indeed, what reason is there to think that the possibility of simultaneous, asymmetric causation depends upon my being able to come up with any kind of criterion of causal directionality at all? My enunciation of a criterion for distinguishing a cause from its effect is an epistemic affair; the existence of simultaneous causation is a matter of ontology. A criterion helps us to *discern* simultaneous, asymmetric causes in the world; but to suggest that said criterion somehow *constitutes* such causal relations in reality is verificationism at its most implausible. Grünbaum has not suggested any incoherence or difficulty in simultaneous, asymmetric causation; if there are such causes in the world, they do not have to wait around for us to discover some criterion for distinguishing them. (iii) There is no reason to think that in order for specific cases of simultaneous, asymmetric causation to be possible or discernible, one must be able to furnish a general criterion broad enough to cover all such alleged cases. All one needs is a way of distinguishing cause from effect in the specific case. Now in the case of the hypothesis of theological creationism, we have a logically airtight means of distinguishing cause from effect, namely, it is *metaphysically impossible* for God to be caused by the world, since if God exists, His nature is such that He exists necessarily, whereas the world's existence is metaphysically contingent (as is evident from its beginning to exist). That entails that there is *no possible world* in which God is caused by the Big Bang. Hence, it is easy for the theist to explain in what sense God is causally prior to the universe or the Big Bang: God and the universe are causally related, and if the universe were not to exist, God would nevertheless exist, whereas there is no possible world in which the universe exists without God. Thus, it seems to me that Grünbaum's objection to a supernatural cause of the origin of the universe is unsuccessful.

The naturalist will perhaps raise a metaphysical objection to the scenario I have sketched of the Creator's status *sans* the universe. For it requires that we conceive of a timeless, personal agent, and some philosophers have argued that such a notion is self-contradictory.[42] For it is a necessary condition of

personhood that an individual be capable of remembering, anticipating, reflecting, deliberating, deciding, and so forth. But these are inherently temporal activities. Therefore, there can be no atemporal persons.

The fallacy of this reasoning is that it conflates *common* properties of persons with *essential* properties of persons. The sorts of activities delineated above are certainly common properties of temporal persons. But that does not imply that such properties are essential to personhood. Arguably, what is necessary and sufficient for personhood is self-consciousness and free volition, and these are not inherently temporal. In his study of divine timelessness, John Yates writes:

> The classical theist may immediately grant that concepts such as reflection, memory, and anticipation could not apply to a timeless being (nor to any omniscient being), but this is not to admit that the key concepts of consciousness and knowledge are inapplicable to such a deity . . . there does not seem to be any essential temporal element in words like . . . 'understand', to 'be aware', to 'know', and so on . . . an atemporal deity could possess maximal understanding, awareness, and knowledge in a single, all-embracing vision of himself and the sum of reality.
>
> (1990: 173)

Similarly, God could possess a free, changeless intention of the will to create a universe with a temporal beginning. Thus, neither self-consciousness nor free volition entail temporality. But since these are plausibly sufficient for personhood, there is no incoherence in the notion of a timeless, personal Creator of the universe.

More recently, Smith has argued that 'the thesis that the universe has an originating divine cause is logically inconsistent with all extant definitions of causality and with a logical requirement upon these and all possible valid definitions or theories of causality' (1996: 169–170). Smith shows that the typical analyses of the causal relation in terms of temporal priority, spatio-temporal contiguity, and nomological relatedness are inapplicable to the event of God's willing that the Big Bang occur and the event of the occurrence of the Big Bang. Therefore, these two events cannot, on the customary analyses, be regarded as cause and effect. Counterfactual analyses of causation, such as David Lewis's, according to which c causes e if (i) c and e are both events which occur and (ii) if c had not occurred, e would not have occurred, fare no better in Smith's view. For if c is God's willing and e is the Big Bang, it is true that if e had not occurred, then c would not have occurred. But this implies that the Big Bang is the cause of God's willing, which is false. Lewis avoids the problem of spurious reverse causal dependence by stipulating that if e had not occurred, then c would have occurred but failed to cause e. But since God is omnipotent and His willing necessarily effective, such a stipulation cannot be made in the present case. Thus, under no extant analysis of causality can God be said to cause the Big Bang.

Smith's argument may be formulated as follows:

4 If the claim that God caused the Big Bang cannot be analyzed in terms of extant definitions of causality, then God cannot have caused the Big Bang.
5 The claim that God caused the Big Bang cannot be analyzed in terms of extant definitions of causality.
6 Therefore, God cannot have caused the Big Bang.

Is this argument sound and persuasive? I think not.

Consider premise (4). I see no reason to think that this premise is true. In general, arguments to the effect that some intuitively intelligible notion cannot be analyzed in terms of current philosophical theories ought to make us suspect the adequacy of those theories rather than reject the commonsense notion. The idea that God caused the universe is intuitively intelligible. A cause is, loosely speaking, something which produces something else and in terms of which the thing that is produced can be explained. That notion certainly applies to God's causing the universe. Indeed, God's causing certain effects is analogous to our acting as agents to bring about certain effects. We certainly conceive of ourselves as causes, and, intuitively, God should count as a cause as well. But Smith's argument, if successful, could be generalized to prove that God is not a cause of anything whatsoever. If God's acting as a cause cannot be analyzed in terms of current philosophical definitions of causation, then so much the worse for those definitions! That only shows that the definitions need to be revised. Indeed, the standard procedure in terms of which proposed definitions of causality are assessed is to postulate counter-examples of intuitively obvious cases of causation and then show how the definition fails to accommodate these examples. In the same way, if God's being a cause cannot be accommodated by some philosophical definition of causality, then that plausibly constitutes a counter-example to the definition which shows its inadequacy as a general metaphysical analysis of the causal relation, however adequate it might be for scientific purposes.[43]

Moreover, there is no reason to believe that we have arrived at the final and correct analysis of causation. In fact, there is good reason to believe the opposite. The definitions discussed by Smith are exclusively concerned with natural, even physical, causes. They were not intended to cover such recondite cases as divine causation of the origin of the universe. It is hardly surprising, therefore, that these analyses should fail to capture this notion. Smith simply ignores analyses of causation which are not currently fashionable but which were crafted in the context of a theistic metaphysic and are consonant with God's being the cause of the origin of the universe, for example, the account of efficient causation and creation given by Francisco Suárez in his monumental *Disputationes metaphysicae* (2002). In his lengthy Introduction to his translation of questions 20–22 of Suárez's work, Freddoso

argues that Suárez's account of causality not only enables one to construe God's creation of the universe as an instance of causation but also contrasts favorably with empiricist accounts of causality offered by contemporary philosophers such as Mackie, Lewis, van Fraassen, and Tooley.

Finally, Smith just assumes that an analysis of the causal relation can be given. But it could be held that such a relation is conceptually primitive, in which case we should not expect a successful reductive analysis to exist which will cover all cases. The plethora of competing extant analyses and the recognized deficiencies of all of them lend credibility to this viewpoint.

What about premise (5)? It seems to me that there are analyses of causation, however inadequate, which can accommodate God's causing the Big Bang. Consider Lewis's analysis of causation. According to Lewis, *c* causes *e* if and only if *c* and *e* are both events which occur, and if *c* had not occurred, *e* would not have occurred. Now God's willing the Big Bang clearly satisfies this definition: God's willing and the Big Bang are both events which occur, and if God's willing had not occurred, the Big Bang would not have occurred. But Smith rejoins, 'But if the Big Bang had not occurred, God's willing would not have occurred. So is the Big Bang the cause of God's willing?' Obviously not; but what this calls into question is the *adequacy* of Lewis's analysis, not whether divine causation satisfies it. Lewis remedies the problem by stipulating that if *e* had not occurred, *c* would still have occurred but failed to cause *e*, a remedy which will not work for divine causation. Actually Lewis's remedy will not work for many natural causes either, since in some cases the counterfactual, 'If *e* had not occurred, *c* would not have occurred' is true. So what Lewis's definition gives is not an analysis of '*c* causes *e*' but rather an analysis of '*c* and *e* are causally related', and it fails to specify the *direction* of causation. But the theist faces no problem there: for, as we have said, it is metaphysically impossible for God's willing to have an external cause. There is no possible world in which the Big Bang causes God's volition. Therefore, given Lewis's analysis of '*c* and *e* are causally related' and the impossibility of the Big Bang's causing God's willing, it follows that God's willing causes the Big Bang. Thus, divine causation satisfies Lewis's definition of causality.

Again, there are analyses of agent causation which are even more relevant in the case of divine causation than the analyses surveyed by Smith. Smith considers exclusively event causation, but it may be disputed whether this is the correct conception to apply to God's case. Smith contends that considerations of agent causation are not germane to the discussion because we are not concerned with the relation between God (the agent) and His act of willing (the effect), but with the relation between His act of willing (an event) and the Big Bang (an event). But not all proponents of agent causation construe agent causation as a relation between an agent and his volitions. Some proponents of agent causation hold that an agent does not cause his volitions, but that by freely willing he brings about some intended event (Lowe 2002: 205–210). In the case at hand God brings about the Big

Bang by an exercise of His active power. The expression 'God's willing that the Big Bang occur' properly describes an action, not an event. The event in this case is the Big Bang, and the cause of that event is God, who, by willing, brought about the Big Bang. Thus, it is simply wrong-headed to think of the Big Bang as caused by the event of God's willing rather than by God Himself.[44]

Thus, neither (4) nor (5) commends itself to us as more plausibly true than its contradictory. Smith recognizes these deficiencies of his argument, but he falls back to what he considers an impregnable position: '*c* is a cause of *e*' entails '*c* is not a logically sufficient condition of *e*' (Smith 1996: 176). This entailment precludes God's being the cause of the Big Bang because God's willing that the Big Bang occur is a logically sufficient condition of the Big Bang. This is because God is omnipotent, and thus necessarily His will is successful. There is no possible world in which God wills the Big Bang and yet the Big Bang fails to occur. Therefore, God cannot be the cause of the Big Bang.

This argument seems quite fanciful. If successful, it can be generalized to show that God cannot cause anything. Thus, precisely *because* He is omnipotent, God is utterly impotent – a curious inference! If being omnipotent entails inability to cause anything, then we are using 'cause' in a highly technical sense which is not incompatible with God's bringing about the Big Bang, which is, after all, the issue. Whether or not God 'causes' the Big Bang, it is still up to Him whether it occurs or not, and it occurs only in virtue of His willing that it occurs. If it seems that bringing about the Big Bang does involve a causal relation, then we shall simply reject Smith's entailment principle. Only someone who is already a naturalist would be tempted to think that that principle is true. Thus, Smith's argument is either question-begging or not incompatible with God's bringing about the Big Bang.

Smith considers such a response and insists that it is the theist who begs the question, since in every other case of causation causes are not logically sufficient conditions of their effects. There is, he says, no justification for exempting God's alleged acts of causation from this principle. We need to have some independent reason for thinking that the relation between God and the Big Bang is a causal relation. Three things may be said about this response: (i) Since only God is omnipotent, it is hardly surprising that His case should be the sole exception to the principle that causes are not logically sufficient for their effects. God is so exceptional a being that He will in general not fit into our customary schemata. For example, it is a general principle that '*S* believes *p*' is not a logically sufficient condition of '*p*'. But since God is essentially omniscient, in God's case His believing *p* is a logically sufficient condition of *p*. Should we therefore conclude that God has no beliefs? In the same way, because God is omnipotent, are we to think that His will has no effects? (ii) There are other plausible counter-examples to Smith's principle. For example, change is plausibly a cause of the existence of time, at least on a relational view of time. The occurrence of events

actually brings time into existence. If there were an absolutely quiescent state, then time would not exist. But if a change occurs, time is immediately produced. Such a relation is plausibly causal; it is certainly not like the purely logical relation between, say, a two-dimensional figure's having three sides and its having three angles. Time is something altogether distinct from change, since time can go on, even most relationalists agree, even though change should cease (Shoemaker 1969). Thus, change, should it occur, would seem to cause time to exist. Yet change necessarily causes time: there is no possible world in which change is going on without time. Change is thus logically sufficient for the existence of time, but is also plausibly a cause of time's existence. (iii) The reason that the relation between God and the Big Bang – or any other event He brings about – is causal is the close resemblance between God and ourselves as agents. Doubtless our deepest intuitions about causality are rooted in our own ability to bring about effects by an intentional exertion of our power. But God is a personal agent like us. The difference between Him and us is that His power is so great that He is infallible in bringing about His undertakings. Is His status as a cause now to be doubted because He is infallible? Hardly! In short, I do not think that Smith's objection poses a serious obstacle to thinking that the Big Bang has a supernatural or divine cause.

All of the above objections have been considered as attempted justification of the apparently incredible position that the universe sprang into being uncaused out of nothing. But I, for one, find the premises of those objections far less perspicuous than the proposition that *whatever begins to exist has a cause.* It is far more plausible to deny one of those premises than to affirm what Hume called the 'absurd Proposition' that something might arise without a cause,[45] that the universe, in this case, should pop into existence uncaused out of nothing.

Conclusion

We can summarize our argument as follows:

7 Whatever begins to exist has a cause of its existence.
8 The universe began to exist.
9 Therefore, the universe has a cause of its existence.

Premise (7) is an intuitively grasped, metaphysical first principle. Premise (8) is supported by the inductive evidence of contemporary cosmology and enjoys greater plausibility in light of that evidence than its contradictory. An analysis of what it is to be cause of the universe reveals that:

10 If the universe has a cause of its existence, then an uncaused, personal Creator of the universe exists, who *sans* the universe is beginningless, changeless, immaterial, timeless, spaceless, and enormously powerful.

From (9) and (10), it follows that:

11 Therefore, an uncaused, personal Creator of the universe exists, who *sans* the universe is beginningless, changeless, immaterial, timeless, spaceless, and enormously powerful.

And this, as Thomas Aquinas laconically remarked, is what everyone means by 'God'.[46]

Notes

1 Aristotle wrote:

> For it is owing to their wonder that men both now begin and at first began to philosophize; they wondered originally at the obvious difficulties, then advanced little by little and stated difficulties about the greater matters, e.g. about the phenomena of the moon and those of the sun and the stars, and about the genesis of the universe.
>
> (*Metaphysics* A, 2, $982^{b}10$–15)

2 For this analysis of so-called factual necessity, see Hick (1960: 733–734).

3 As Gott *et al.* write:

> the universe began from a state of infinite density about one Hubble time ago. Space and time were created in that event and so was all the matter in the universe. It is not meaningful to ask what happened before the big bang; it is somewhat like asking what is north of the North Pole. Similarly, it is not sensible to ask where the big bang took place. The point-universe was not an object isolated in space; it was the entire universe, and so the only answer can be that the big bang happened everywhere.
>
> (1976: 65)

The Hubble time is the time since the singularity if the rate of expansion has been constant. The singularity is a point only in the sense that the distance between any two points in the singularity is zero. Anyone who thinks there must be a place in the universe where the Big Bang occurred still has not grasped that it is space itself which is expanding; it is the two-dimensional *surface* of an inflating balloon which is analogous to three-dimensional space. The spherical surface has no center and so no location where the expansion begins. The analogy of the North Pole with the beginning of time should not be pressed, since the North Pole is not an edge to the surface of the globe; the beginning of time is more like the apex of a cone. But the idea is that just as one cannot go further north than the North Pole, so one cannot go earlier than the initial singularity.

4 As Jaki points out, Hoyle and his colleagues were inspired by 'openly anti-theological, or rather anti-Christian motivations' (Jaki 1974: 347). Martin Rees recalls his mentor Dennis Sciama's dogged commitment to the Steady State Model:

> For him, as for its inventors, it had a deep philosophical appeal – the universe existed, from everlasting to everlasting, in a uniquely self-consistent state. When conflicting evidence emerged, Sciama therefore sought a loophole (even an unlikely seeming one) rather as a defense lawyer clutches at any argument to rebut the prosecution case.
>
> (Rees 1997: 41)

The phrase 'from everlasting to everlasting' is the Psalmist's description of God (Ps. 90:2). Rees gives a good account of the discoveries leading to the demise of the Steady State Model.

5 As is evident from the sentiments expressed by John Gribbin:

> The biggest problem with the Big Bang theory of the origin of the universe is philosophical – perhaps even theological – what was there before the bang? This problem alone was sufficient to give a great initial impetus to the Steady State theory; but with that theory now sadly in conflict with the observations, the best way round this initial difficulty is provided by a model in which the universe expands from a singularity, collapses back again, and repeats the cycle indefinitely. (1976: 15)

Scientists not infrequently misexpress the difficulty posed by the beginning of the universe as the question of what existed before the Big Bang (which invites the easy response that there was no 'before'). The real question concerns the causal conditions of this event, why the universe exists rather than nothing.

6 Associated Press News Release, 9 January 1998 (available on-line to subscribers at http://www.sciencenews.org/20010331/fob1.asp).

7 Ibid. Indeed, the supernovae evidence actually suggested that the universe is *accelerating*, in which case there must be effectually a positive cosmological constant. More recent studies carried out on the microwave background radiation by the Wilkinson Microwave Anisotropy Probe confirm this finding. In this case even a high density universe will expand forever, so that the future of the universe is independent of density considerations. See R. Scranton, *et al*., 'Physical Evidence for Dark Energy', http://arXiv.org/abs/astro-ph/0307335 (accessed 20 July 2003).

8 One thinks, for example, of the late Carl Sagan on his *Cosmos* television series propounding this model and reading from Hindu scriptures about cyclical Brahman years in order to illustrate the oscillating universe, but with nary a hint to his viewers about the difficulties attending this model.

9 R. Brandenberger, personal communication.

10 Vilenkin (1983: 28–54). See Hartle and Hawking (1983); Vilenkin (1982: 25–28).

11 See my Craig (1997: 291–295). With respect to determining the wave function of the universe Bryce DeWitt says, 'Here the physicist must play God' (DeWitt 1983: 120).

12 Hawking and Penrose (1996: 3–4). See his comment: 'I . . . am a positivist who believes that physical theories are just mathematical models we construct, and that it is meaningless to ask if they correspond to reality, just whether they predict observations', S. Hawking, 'The Objections of an Unashamed Positivist', in Penrose (1997: 169).

13 Hawking and Penrose (1996: 121); see p. 4.

14 As pointed out by Isham (1993–1996: 56).

15 Weyl curvature is the curvature of space-time which is not due to the presence of matter and is described by the Weyl tensor. Space-time curvature due to matter is described by the Ricci tensor. Together they make up the Riemann tensor giving the metric for space-time.

16 See http://feynman.princeton.edu/~steinh/.

17 See especially G. Felder, A. Frolov, L. Kofman, and A. Linde, 'Cosmology with Negative Potentials', http://arXiv:hep-th/0202017v2 (16 February 2002) and the therein cited literature, particularly the studies by D. Lyth. See http://eprints.osti.gov/cgi-bin/dexpldcgi?qry2062780997;4.

18 A. Borde, A. Guth, and A. Vilenkin, 'Inflation Is Not Past-Eternal', http://arXiv:gr-qc/0110012v1 (1 Oct 2001): 4. The article has now been updated as of January 2003. See http://arXiv.org/abs/gr-qc/0110012.

19 A. Vilenkin, personal communication.

20 Some recent efforts have been made to describe a pre-Big Bang universe in terms of super-string or M-theory (Gasperini 1999, 2000) on the assumption of a 'duality-symmetry' which associates a geometric mirror image with the familiar post-Big Bang expanding space-time geometry. Apart from the problems that there seems to be no way to join the pre- and post-Big Bang eras together nor any way to smooth the transition to a matter-dominated universe, the scenario is based on a non-existent theory and so cannot even begin to be a plausible alternative.
21 For a helpful introduction to these two competing perspectives, see Gale (1968a: 65–85).
22 See discussion in Craig (1990: 150–156).
23 This is the salient point of Grünbaum's critique of the inference to a First Cause of the origin of the universe (Grünbaum 2000). As a B-theorist, Grünbaum does not believe that the universe ever came into being, even if it had a first temporal interval. As he elsewhere writes, 'coming *into* being (or "becoming") is *not* a property of *physical* events themselves but only of human or conscious awareness of these events' (1967: 153). What Grünbaum fails to see, however, is that the claim that an absolute beginning of the universe entails that the universe came into being is rooted, not in the presupposition of the so-called Spontaneity of Nothingness, but in an A-theory of time.
24 See my companion volumes Craig (2000a, 2000b).
25 For an outstanding defense of this point, see Smith (1991). See also Craig (1996a, 1996b, 2000c).
26 One of the most eloquent spokesmen for this point of view has been Schlesinger (1980: 34–39, 138–139). See also Craig (1999a, 1999b, 2001).
27 The most helpful here are still Broad (1938) and Dummett (1960: 497–504). See also Craig (1998a: 122–127).
28 This point needs further work, but see Prior (1968: 1–14) and Loizou (1986: 44–45). See also Craig (2003, 2000d).
29 See Smith (1991, Chapter 7). See also Craig (2002).
30 See remarks by Black (1962).
31 For adumbrations of this argument, see Gale (1968b: 90–97) and Mellor (1981: 140). For a fuller development see Craig (forthcoming).
32 Again, this point needs to be better developed, but see Geach (1972: 306), and McGilvray (1979: 275–299).
33 See the excellent study by Merricks (1994); see further Hinchliff (1994), Lewis (1986), and van Inwagen (1990).
34 See the very interesting recent discussions about the warrant for methodological naturalism in science, e.g. de Vries (1986); Plantinga *et al*. (1991); Hasker (1992); Plantinga (1992, 1996a); Moreland (1994); Moreland *et al*. (1994).
35 Smith (1993a: 135). Elsewhere he has written:

> [This world] exists non necessarily, improbably, and causelessly. It exists *for absolutely no reason at all* . . . The impact of this captivated realization upon me is overwhelming. I am completely stunned. I take a few dazed steps in the dark meadow, and fall among the flowers. I lie stupefied, whirling without comprehension in this world through numberless worlds other than this.
>
> (Smith 1986: 300–301)

In *Theism, Atheism, and Big Bang Cosmology,* Smith claimed that the universe came into being uncaused out of nothing at the Planck time; but he has since recanted that position under the realization that the whole field of quantum cosmology is then studying a complete fiction!
36 John Leslie asks incredulously, 'How could such nonsense have been churned out by the author of *Physical Chemistry,* a superb textbook?' (Leslie 1993: 3). For a good critique of Atkins, see Ward (1996*,* Chapter 1).
37 See Weingard (1982: 235–242).

38 Grünbaum (1989). For a response, see Craig (1992).
39 Grünbaum (1991). For a response, see Craig (1994a).
40 Brian Leftow puts this nicely when he writes:

> If God existed in time once time existed and time had a first moment, then God would have a first moment of existence: there would be a moment before which He did not exist, because there was no 'before' that moment . . . Yet even if He . . . had a first moment of existence, one could still call God's existence unlimited were it understood that He would have existed even if time did not. For as long as this is true, we cannot infer from God's having had a first moment of existence that God *came into* existence or would not have existed save if time did.
>
> (1991: 269; see 201)

Senor has dubbed such a model of divine eternity 'accidental temporalism' (Senor 1993: 88). See further, Craig (1996c).
41 Grünbaum (1994). For a response, see Craig (1994b).
42 See discussion and references in Craig (1998b).
43 In Smith (1993b), Smith actually arrives at this conclusion himself. He states:

> extant definitions of causality are incorrect since they do not cohere in the proper way with the concept of a cause of the universe . . .
>
> This entails that either there is some other (as yet undiscovered) definition of a cause that is correct or that a cause is indefinable. In the latter case, the concept of a cause would be primitive and the causal relation a simple relation known only by ostension (as is arguably the case with such relations as *being in contact with* or *being earlier than*). I know of no means of discovering or formulating a correct definition of a cause and know of no reason to think that there is such a definition. Accordingly, I think it is reasonable to conclude that the causal relation is indefinable.
>
> One way to avoid this conclusion would be to reject the assumption that the various examples of causes of the big bang . . . are genuine examples of causes . . . I would say that claims that God's creation of the big bang singularity and other examples given . . . are not cases of possible causation are counterintuitive and are *ad hoc* attempts to retain a counterexampled theory. It is more plausible to think that a cause cannot be defined than to think that a mind's creation of a big bang singularity could not be a causal act. (Smith 1993b: 1, 24)

Smith came to think God's relation to the Big Bang is not causal because no cause is logically sufficient for its effect. But Smith does not justify why the actions of an omnipotent being would not be exceptions to this rule.
44 See Moreland (1998). I am indebted to my colleague for several interesting discussions pertinent to agency and creation.
45 David Hume to John Stewart, February 1754, in Greig (1932, 1: 187).
46 Thomas Aquinas, *Summa theologiae*, 1a.2.3.

How to deal with singularities

Comment on Craig's paper

Winfried Löffler

Introduction

William Craig's rich paper contains a mass of material, and as a non-physicist I will not try to enter into a discussion of the physical issues dealt with in the first part of his paper. As far as I can judge, however, Craig's account of recent cosmological theories is adequate and complete (see, for example, Kanitscheider 2004). And I take it that it can be summarized by saying that all alternatives so far put forward to the standard Big Bang Model have one or more serious flaws: they are merely speculative and have little or no evidence in favour of them; some of them even have substantial evidence against them; some of them are internally contradictory or involve serious conceptual implausibilities; and some of them, at a rather remote level, require a sort of temporal beginning or an initial singularity themselves – although the original motivation to establish those models was to avoid such a beginning or singularity. Craig takes these flaws as an argument for a temporal beginning of the universe. In the second part he sketches an argument which (by pleading for a wider account of causation and by invoking a version of the metaphysical principle of causality) shows that there is a cause of the universe which displays the traditional theistic properties and can hence be identified with God.

Out of broad sympathy with Craig's views, my comments fall into four parts. After a short remark from the perspective of the philosophy of science, I will try to clarify which kinds of naturalisms and anti-naturalisms are found in the debate here. After that, I will reflect on the sources of (anti-)naturalistic claims, and, finally I will briefly explore the anti-naturalistic strategy behind Craig's paper and estimate its prospects. I shall constrain my comment to the naturalism/anti-naturalism issue; an appropriate discussion of Craig's theistic argument would go beyond the scope of this brief comment.

A note: Is the standard model corroborated by the flaws of its competitors?

Regarding the first part, I have only one little critical remark. It concerns Craig's interpretation of the impact of the flaws of the competing theories

on the standard model. At least twice in his paper (Craig, 'Summary', p. 115), he talks about the increasing 'corroboration' of the standard Big Bang model by the successive failures of its competitors. If 'corroboration' is to be taken in Popper's sense (and Craig's overall realist account of scientific theories strongly supports this Popperian reading), I do not think that this is really appropriate. A theory is corroborated if it 'survives' an attempt to falsify it in an empirical test. But things are quite different here with the standard model and its competitors. Craig himself points out that one big problem of many competing models is their *lack* of evidential backing. They were not proposed as better explanations for odd pieces of evidence, not even as empirically testable alternative theories. They rather appear as cosmological speculations, primarily designed in order to avoid the singularity problem. Hence, if these extravagant speculations fail, this does not mean that the standard model is 'corroborated'. According to Popper, a reasonable theory with considerable evidence on its side is not corroborated by the failure of some less reasonable competitors which have no good evidence in favour of them. So it seems to me that Craig cannot both claim Popperian corroboration of the standard Big Bang model while attacking the competing models on the grounds of lack of evidence.

It might nevertheless seem intuitively rational to be more confident about the Standard Model if the competitors are proven implausible. But as Craig describes it, the situation rather fits within the framework of Bayesian epistemic probabilities (and in his conclusion, Craig indeed talks about 'the inductive evidence of contemporary cosmology'): there is a variety of competing theories with prior probabilities greater than 0; their respective probabilities must sum up to 1; and the standard Big Bang model becomes epistemically more probable as the competing models decrease in probability. Their decrease in epistemic probability can have various reasons, ranging from internal incoherence to conceptual extravagances to empirical failure. In other words, and somewhat more loosely: the claim that a theory is corroborated turns on its empirical success in the *past*; a claim that a theory is probable is a claim about its *future* prospects.

But as I said, this minor point leaves the substance of Craig's argument untouched. My further comment focuses on some methodological aspects, both of Craig's position as well as of his description of naturalism.

Which (anti-)naturalism?

Craig says that 'Naturalists ... have typically claimed that the space-time universe is itself at least factually necessary – that is to say, eternal, uncaused, incorruptible, and indestructible – while dismissing the demand for a logically necessary being' (Craig, p. 97), and in his conference abstract he calls the eternity of the universe a 'perennial naturalistic assumption'. (It goes without saying that the understanding of 'eternity' here has varied as we have moved from simple pre-Einsteinian accounts towards the extravagant cosmologies of the past decades.)

How should we classify this description of naturalism? Against the background of a widespread (and I think useful) distinction between *semantical*, *methodological* and *ontological* naturalism (Löffler 1999), it appears as a quite strong form of ontological naturalism – the attribution of some highly substantial, non-trivial, ontological properties to the universe, and in connection with that, the claim of the nonexistence of a traditional God-like being with similar properties. It should be noted that physicists – *qua physicists*! – would not be inclined towards such a thesis. Such an ontological naturalism is not a part of physics. For physicists, such doctrines as the eternity and indestructibility of the universe are framework assumptions implicit in their normal scientific method as physicists. When they began to appear questionable, given what was emerging regarding the apparent history of the universe, some physicists tried, quite understandably, to rescue them by speculating about various alternative hypotheses to the Big Bang. However, developing a full-blown *ontology* from such alternative hypotheses is not the business of physicists. It is rather a business for physicists doing naturalistic philosophy or for naturalistic philosophers interpreting physical theories.

Moreover, not even every philosophical naturalist would subscribe to Craig's version of naturalism. *Methodological* naturalists are committed only to the claim that philosophical investigations should use those methods, and only those methods, which would also be acceptable in the sciences. Of course physical theories are likely to presuppose what we might call a 'local ontology', containing, for instance, their theoretical entities. But methodological naturalism can co-exist well with a sort of ontological agnosticism beyond that. Thorough methodological naturalists would restrict themselves to the declaration that we cannot reasonably say anything about the status of the universe before the Big Bang, on the ground that a singularity is a singularity, and hence neither our concepts of space and time nor the known laws of physics are applicable. In the eyes of the methodological naturalist, any attempt to theorize into that field, especially any further ontological claims, would end up in dubious claims transgressing the limits of physics proper.

Admittedly, there could be some discussion whether a rule like 'try to extend the limits and explanatory claims of science as far as possible, and try to avoid singularities as far as possible' belongs to the methods of science and is hence also a legitimate claim of philosophical naturalism. From this point of view, we could say that a physicist or philosophical naturalist with a preference for one of the competing models could well be within their methodological rights (we might add, of course, that they should keep an eye on other methodological rules as well, e.g. the rules of simplicity and parsimony). But in this context, some of Craig's observations are particularly interesting: if it is really true that some alternative models to the Big Bang have their roots not only in a certain (and methodologically acceptable) *horror singularitatis*, but in the atheistic or pantheistic background

convictions of their authors, then these models would be clear cases where a sort of ontological naturalism is the fuel not only for methodological naturalism, but for physical theorizing itself. And this would deserve more attention from historians and psychologists of science.

But let us now turn to Craig's position. That Craig ultimately defends a sort of ontological anti-naturalism, namely Christian theism, is only obvious. But this is not the whole story. I read Craig as defending a methodological anti-naturalism as well. In an aside, Craig says that 'moreover, even methodological naturalism is far from unchallengeable', and in fact, what he does in the second part of his paper amounts to an attack on the narrow-minded concepts of explanation and causation which are typical of methodological naturalisms. Craig strives for the rehabilitation of something like the traditional doctrines of the manifold senses of 'cause' and 'explanation', and I think this is quite a promising way of attacking naturalism. In the subsequent paragraphs, I want to sketch out why.

Do cosmological theories alone favour (anti-)naturalism?

One could think that the result of Craig's investigation concerning the standard Big Bang models and its competitors is already a victory for the anti-naturalist, especially the theistically-minded anti-naturalist. Conversely, if the competitors had scored better, this would have been the victory for the naturalist. However, this is not the case. There are at least two worrying facts: first, there are people who find a purported eternity of the universe just as mysterious as a creation 'ex nihilo' in a singularity. For them, the Big Bang and some variant of an eternal universe are equally wrong answers to the question 'Why is there something rather than nothing?' Second, there are people like Quentin Smith who draw atheistic conclusions from the Big Bang model. The main premise of their argument is the lawlessness of the Big Bang singularity. In an old article from 1976, before his move to quantum cosmology, Stephen Hawking writes:

> A singularity is a place where the classical concepts of space and time break down as do all the known laws of physics because they are all formulated on a classical space-time background . . . [T]his breakdown is not merely a result of our ignorance of the correct theory but represents a fundamental limitation to our ability to predict the future [of the singularity], a limitation that is analogous but additional to the limitation imposed by the normal quantum-mechanical uncertainty principle.
>
> (Hawking 1976: 2460)

One could conclude from this lawlessness and unpredictability that the singularity 'would thus emit all [possible] configurations of particles with equal probability' (ibid.), or, as Paul Davies puts it: 'Anything can come out

of a naked singularity – in the case of the big bang the universe came out' (1981: 161). Quentin Smith concludes from this:

> if God caused the universe to begin to exist with the intention that the universe contains intelligent life, he would have created an initial state that certainly or probably evolves in a lawlike manner that is conducive to the existence of intelligent organisms . . . Since big bang cosmology implies that the initial state is instead a lawless singularity, big bang cosmology disconfirms the theistic hypothesis.
>
> (1997: 125)

If Smith is correct, and I think in that point he is, then we either are back to the mysterious fact of a giant cosmic chance that we are here, or it shows us that the Big Bang theist must indeed defend a little additional hypothesis: that God not only caused the initial singularity, but that he also moved it towards a certain development.

The theist is surely not irrational in doing that – a singularity is a singularity, and we are not breaking *physical* laws if we assume some Divine interaction in it. I think we can learn from this that cosmological theories and models are open to several interpretations, to embeddings in several *Weltanschauungen*, of both the naturalistic and anti-naturalistic kind. And indeed, this is a very important result of Craig's paper, a result which is quite uncomfortable for many naturalists: theists are within their rights to interpret Big Bang cosmology non-naturalistically, as some supernatural action of God. Belief in a theistic God should not be seen as a suggestion by the findings of modern cosmology, let alone a consequence of them. Nevertheless, it is at least fully compatible with modern cosmology, and can be regarded as an epistemologically legitimate completion of the physical world picture. Hence, the answer to the question in the heading of this section is negative: it is not a cosmological theory alone which tells in favour of naturalism or anti-naturalism, it is, rather, its embedding in a *Weltanschauung* or certain metaphysical background assumptions.

Beyond methodological naturalism

But is not Craig's paper meant as a positive argument for God's existence from Big Bang cosmology (and not only the 'soft' compatibility argument I made from it so far)? Of course it is. But my foregoing remarks should have made clear that its argumentative power does not rest ultimately on physical cosmology alone, but on the *interpretation* of that cosmology by means of metaphysical principles, especially the principle of causality, the regress principle, etc. Is it rational to appeal to such principles? I think it is, and a proof of their indispensability would yield a serious argument against methodological naturalism.

Craig makes a clever move here by citing naturalists like Kanitscheider and Bunge who invoke principles like 'out of nothing comes nothing' in

their arguments against a Big Bang cosmology with a singularity. This amounts to a sort of *tu quoque*-argument. I doubt whether they can appeal to such principles *qua* methodological naturalists. It is well known that the applicability of the concept of causality in science has been disputed for decades, chiefly for the reason that it is hard to explain a non-circular comprehensive concept of causation. Moreover, modern science is characterized by a move from the search for *causes* to the qualitative description of *processes*, and consequently, the concept of 'cause' does not even appear prominently in modern physics. Kanitscheider is of course aware of all that when he labels the principle a 'metaphysical' one, but he admits that the principle is omnipresent in science. This shows that scientific activities are intertwined with a set of background assumptions which are not parts of the scientific discipline proper, but which nevertheless display rational structures and can be critically discussed. Methodological naturalism tends to underrate such principles, but it is easy to find arguments showing that they are indispensable for science as well as for our everyday life. Even more, there are good arguments showing that our scientific activities depend on the stable functioning of the pre-scientific, everyday practices of our *Lebenswelt*. And in the everyday context, we have few intellectual scruples about such things as explanations which have much in common with Aristotle's manifold senses of 'cause' (as answers to various forms of *why*-questions), or about something like the metaphysical principle of causality, even if modern science is supposed to have discredited them. There is no room to develop this point here in more detail. It should nevertheless be noted that methodological naturalism has awkward and uncongenial consequences: it makes a whole range of our everyday, pre-scientific practices, communication processes, patterns of theorizing and explaining, etc. appear irrational.

If my position is correct, then natural science loses its air of the paradigm and benchmark of all rationality – it rather appears as a highly derivative, dependent form of rational activity. Consequently, methodological naturalism loses its attractiveness right from the beginning. I think that this line of methodological anti-naturalism is fruitful, and I read the second part of Craig's paper as an argument in that direction. By acknowledging this area of pre-scientific reasoning, explaining and problem-solving as a field of rational enterprise we not only cut off the naturalist's most important source of argument, but also rehabilitate a number of metaphysical concepts and patterns of explanation of the kind that are indispensable for any promising form of philosophical theology.

6 The design argument

Between science and metaphysics

Robin Collins

In this paper, I will first briefly discuss the so-called argument for design based on the fine-tuning of the laws and constants of physics for intelligent life. However, I will spend most of my paper developing the design argument from the beauty and elegance of the laws of nature.

Review of basic fine-tuning argument

The fine-tuning of the cosmos for life refers to the fact that many of the fundamental parameters of physics and the initial conditions of the universe are balanced on a razor's edge for intelligent life to occur: if these parameters were slightly different, life of comparable intelligence to our own would not exist. The first major, though still controversial claims along these lines was in 1956 – that the resonance states of carbon and oxygen had to fall within a narrow range for significant quantities of both carbon and oxygen to be produced in stars. Without enough carbon and oxygen, the existence of carbon-based life would be seriously inhibited. Many other instances of cosmic fine-tuning have been brought to light since then, and much work is continuing. Although the physical reasoning behind some of the most prominent cases of fine-tuning is flawed, many other cases hold up to careful scrutiny. To allay skepticism about the evidence for fine-tuning, in a recent paper, I have carefully explicated six solid cases of fine-tuning (Collins 2003). One of the most impressive and most discussed cases of fine-tuning is that of the cosmological constant, a term in Einstein's equation of general relativity that governs the rate at which space expands. For the universe to be hospitable to life, this constant must be fine-tuned to at least one part in 10^{53} – that is, one part in one hundred million, billion, billion, billion, billion, billion – of what physicists consider its natural range of values.

Some physicists and others have taken the position that fine-tuning provides significant evidence that the cosmos is designed – and, furthermore, that one of the purposes of the designer was to create embodied, intelligent beings. Others have questioned this inference by saying that, as far as we know, the values of the fundamental parameters will eventually be explained

by some grand unified theory. Hence, it is argued, we do not need to invoke a designer to explain why these parameters have life-permitting values. As astrophysicists Bernard Carr and Martin Rees note, however, 'even if all apparently anthropic coincidences could be explained [in terms of such a unified theory], it would still be remarkable that the relationships dictated by physical theory happened also to be those propitious for life' (Carr and Rees 1979: 612). For the theist, then, the development of a grand unified theory would not undercut the case for design, but would only serve to deepen our appreciation of the ingenuity of the creator: instead of separately fine-tuning each individual parameter, in this view, the designer simply carefully chose those laws that would yield life-permitting values for each parameter.

Another objection to considering fine-tuning as evidence for design is one that takes us almost into the realm of science fiction: the proposal that there are a very large number of universes, each with different values for the fundamental parameters of physics. If such multiple universes exist, it would be no surprise that the parameters in one of them would have just the right values for the existence of intelligent life – just as in the case where if enough lottery tickets were generated, it would be no surprise that one of them would turn out to be the winning number.

How did these universes come into existence? One answer, which I call the metaphysical many-universe hypothesis, is that they exist on their own without being generated by any physical process, a view which has been popularized among scientists by University of Pennsylvania astrophysicist Max Tegmark.[1] Typically, however, the answer is to postulate some kind of physical process, what I will call a 'universe generator'. Against the naturalistic version of the universe-generator hypothesis, one could argue that the universe generator itself must be 'well designed' to produce even one life-sustaining universe. After all, even a mundane item such as a bread-making machine, which only produces loaves of bread instead of universes, must be well designed as an appliance *and* have just the right ingredients (flour, yeast, gluten, and so on) in just the right amounts to produce decent loaves of bread. Indeed, as I have shown in detail elsewhere (Collins 2002a), if one carefully examines the most popular, and most well-developed universe-generator hypothesis, that arising out of a combination of inflationary cosmology and superstring theory, one finds that it contains at least seven different laws or 'mechanisms' which are essential for its producing life-sustaining universes. Eliminate one of the fields or laws and no life-sustaining universes would be produced. If this is right, then, to some extent, invoking some sort of universe generator as an explanation of fine-tuning only pushes the issue of design up one level to the question of who or what designed it.

Despite these objections and the fact that the multiple-universe hypothesis typically has been advanced by naturalists as an alternative explanation to design, I am not objecting to the notion of many universes itself. I

actually believe that theists should be open to the idea that God created our universe by means of a universe generator. It makes sense that an infinitely creative deity would create other universes, not just our own. Further, the history of science is one in which our conception of Nature keeps increasing in size in terms of both space and time – from believing that the Universe consisted of the Earth and a few crystalline spheres created around 6,000 years ago to positing a 15-billion-year-old universe with more than three hundred billion galaxies. For the theist, the existence of multiple universes would simply support the view that creation reflects the *infinite creativity* of the creator. Finally, I am somewhat inclined to believe the inflationary many-universe hypothesis since so many factors in physics and cosmology in the past twenty years have come together to make it a viable, if not natural, hypothesis.

A final objection to the fine-tuning argument is to simply ask, why do we need an explanation of the fact that a universe exists with life-permitting values for the parameters of physics? Why not simply accept the existence of a life-permitting universe as a brute fact? The proper reply to this question, I think, is that although it is certainly possible that a life-permitting universe exists as a brute fact, the existence of such a universe counts as significant evidence in favour of theism over the brute fact hypothesis. Elsewhere, I have attempted to develop this argument in a more principled and careful way (for example, Collins 2002b), though here I will simply summarize its main features. As I develop it, the 'core version' of the argument essentially involves claiming that the existence of intelligent-life-permitting values for the constants of physics is not surprising under theism, but highly surprising under the non-design, non-multiverse hypothesis – that is, the hypothesis that there is only a single 'universe' and that it exists as a brute fact without any further explanation. Further, the reason it seems highly surprising under the non-design, non-many-universes hypothesis is that, for certain constants of physics, the range of intelligent-life-permitting values is purportedly small compared to some non-arbitrarily defined comparison range – such as the range of force strengths in nature when discussing the fine-tuning of gravity and other forces.

Using what could be called the surprise principle, it follows that the existence of intelligent-life-permitting values for the constants provides evidence in favour of theism over the non-design, non-many-universes hypothesis. According to the surprise principle, if H1 and H2 are two competing non-*ad-hoc* hypotheses[2] and a body of data E is less surprising under one of the hypotheses H1 than under the other, H2, then the data E provides evidence in favour of H_1 over H_2. The best way, I believe, of explicating the notion of surprise used here is in terms of what philosophers call *conditional epistemic probability*, in which case the above principle is a version of the *likelihood principle* or the *principle of relevance*, which is a standard principle of probabilistic confirmation theory. Here the epistemic

probability of a proposition should be understood as the rational degree of belief we should have in the proposition.[3]

Notice that no claim is being made here that theism is the best explanation of the constants being intelligent-life-permitting. To judge that a hypothesis is the best explanation of a body of data involves an overall assessment of the hypothesis, not simply how well it explains the particular data in question. The fact that Johnny's fingerprints are on the murder weapon might significantly support the claim that Johnny committed the murder. Nonetheless, Johnny's committing the murder might not be the best explanation of the fingerprints since we might have strong, countervailing evidence that he did not commit the murder. Perhaps, for instance, five reliable witnesses saw Johnny at a party at the time of the murder. Similarly, all I claim is that the evidence of fine-tuning supports theism over the non-design, non-many-universes hypothesis. However, to judge whether we should infer that theism is the best explanation of the structure of the universe – versus simply accepting the universe as a brute given – involves many factors beyond the evidence of fine-tuning.

Finally, I should mention, one of the key claims in the above argument is that the existence of a universe with intelligent-life-permitting cosmic conditions is not surprising under theism. This claim needs support instead of merely being assumed in an *ad hoc* way. Essentially, the argument is that if God is good – an assumption that is part and parcel of classical theism – then it is *not* extremely improbable that God would create a world with intelligent beings, because the existence of such beings has positive value, at least under the theistic hypothesis.[4]

Design argument from beauty and elegance

Next, I want to look at another, and I believe more powerful, suggestion of design from modern physics, that arising from the 'beauty' of the laws of nature. This suggestion of design completely bypasses the many-universes objection to the design argument, whether that based on the universe-generator or the metaphysical variety of the many-universes hypothesis. First, the idea that the laws of nature are beautiful and elegant is a commonplace in physics, with entire books being devoted to the topic. Indeed, Nobel Prize-winning physicist Steven Weinberg – who is no friend of theism – devotes an entire chapter of his book *Dreams of a Final Theory* to how considerations of beauty and elegance fruitfully guided the development of successful physical theories. Indeed, Weinberg, who is a convinced atheist, even admits that 'sometimes nature seems more beautiful than strictly necessary' (Weinberg 1992: 250).

To develop our argument, however, we need first to address what is meant by beauty. As Weinberg notes, the sort of beauty exemplified by physics is that akin to classical Greek architecture. Arguably, the high point of the classical definition of beauty was that of William Hogarth in his

1753 classic *The Analysis of Beauty*. According to Hogarth, simplicity with variety is the defining feature of beauty or elegance, as illustrated by a line drawn around a cone. Hogarth claimed that simplicity apart from variety, as illustrated by a straight line, is boring, not elegant or beautiful.

The laws of nature seem to manifest just this sort of simplicity with variety: we inhabit a world that could be characterized as a world of fundamental simplicity that gives rise to the enormous complexity needed for embodied, intelligent life. To see this more clearly, we will need to briefly explain the character of physical law, as discovered by modern physics. For example, although the observable phenomena have a great deal of variety and seemingly chaotic, they can be organized via a relatively few simple laws governing postulated unobservable processes and entities. One need only consider, for instance, the simplicity of Newton's law of gravity ($F = GM_1M_2/r^2$) or Einstein's equation of general relativity, which Einstein specifically selected because it was the simplest equation that satisfied certain higher-level constraints that he required for his theory of gravity. What is more impressive, however, is that these simple laws can be organized under a few higher-level principles, as a few examples below will illustrate. These principles in turn fit into a simple and elegant mathematical framework, such as a vector space over the complex numbers in the case of quantum mechanics.

One way of thinking about the way in which the laws fall under these higher-level principles is as a sort of fine-tuning. If one imagines a space of all possible laws, the set of laws and physical phenomena we have are just those that are at the same time simple and that meet the higher-level principles.[5] Of course, by analogy to the case of the fine-tuning of the parameters of physics, there are bound to be other sets of laws that meet some other relatively simple set of higher-level principles. But this does not take away from the fine-tuning of the laws, or the case for design, any more than the fact that there are many possible elegant architectural plans for constructing a house takes away from the evidence for design of a particular house. What is important is that the vast majority of variations of these laws end up causing a violation of one of these higher-level principles. Further, it seems that in the vast majority of such cases, such variations do not result in new, equally simple higher-level principles being satisfied. It follows, therefore, that these variations almost universally lead to a less elegant and simple set of higher-level physical principles being met. Thus, in terms of the simplicity and elegance of the higher-level principles that are satisfied, the laws of nature that we have appear to be a tiny island surrounded by a vast sea of possible law structures that would produce a far less elegant and simple physics. This is why it is appropriate to refer to it as a sort of fine-tuning.

As testimony to the above point, consider what Steven Weinberg and other physicists have called the 'inevitability' of the laws of nature (for

example, see Weinberg 1992: 135–153, 235–237). The inevitability that Weinberg refers to is not the inevitability of logical necessity (ibid.: 235), but rather the contingent requirement that the laws of nature in some specified domain obey certain general principles. The reason Weinberg refers to this as the 'inevitability' of the laws of nature is that the requirement that one choose the simplest equation that meets these principles often severely restricts the possible mathematical forms the laws of nature can take, thus rendering them in some sense 'inevitable'. If we varied the laws by a little bit, they would either not be as simple or these higher-level principles would be violated. Thus this sort of inevitability implies a sort of fine-tuning of the laws for simplicity and elegance.

This inevitability of the laws is particularly evident in Einstein's general theory of relativity. As Weinberg notes, 'once you know the general physical principles adopted by Einstein, you understand that there is no other significantly different theory of gravitation to which Einstein could have been led' (ibid.: 135). As Einstein himself said, 'To modify it [general relativity] without destroying the whole structure seems to be impossible' (quoted in Weinberg 1992: 135).

This inevitability, or near-inevitability, is also illustrated by the gauge principle, the requirement that the dynamical equations expressing the fundamental interactions of nature – gravity, the strong, weak, and electromagnetic forces – be invariant under the appropriate local phase transformation. When combined with the heuristic of choosing the simplest interaction Lagrangian that meets the gauge principle and certain other background constraints, this has served as a powerful guide in constructing the equations governing the forces of nature. Yet, as Ian Aitchison and Anthony Hey point out, there is no compelling logical reason why this principle must hold (1989: 59–60). Rather, they claim, this principle has been almost universally adopted as a fundamental principle in elementary particle physics because it is 'so simple, beautiful and powerful (and apparently successful)' (ibid.: 60). Further, as Alan Guth points out, the original 'construction of these [gauge] theories was motivated mainly by their mathematical elegance' (1997: 124). Thus, the gauge principle provides a good example of a contingent principle of great simplicity and elegance that encompasses a wide range of phenomena, namely the interactions between all the particles in the universe. Other examples of principles that encompass large classes of diverse phenomena are the law of energy conservation, the least action principle, the second law of thermodynamics, and various quantum principles such as the Pauli-exclusion principle. In each of these cases, nature appears to be fine-tuned so that the *simplest laws* that meet the higher-level principles obtain.

Theism offers a natural, non-*ad-hoc* explanation of why the laws of nature can be encompassed by such principles. As we saw above, this sort of simplicity is a key component of the classical conception of beauty. Further, it has been part and parcel of traditional theism that God would be motivated

to bring about an aesthetically pleasing universe. Can a non-theistic, non-design view of reality offer an explanation? And if it cannot, could they legitimately merely say that the laws fall under such higher-level principles?

To answer this question, we begin by noting that it does not seem that one can plausibly think of these principles as in themselves having any causal power to dictate the lower-level phenomena or laws. Thus, the fact that the lower-level laws of physics are fine-tuned to be encompassed by a few simple higher-level principles cannot be explained by appealing to the causal powers of these principles. It is easy to be misled at this point, however. Because we can derive (with a few additional assumptions) the lower-level laws from the higher-level principle, it is easy to think that somehow these higher-level principles *make* the lower-level laws what they are. Rather, the 'causation' or dependence is in the other direction: it's because the laws and phenomena are what they are that these principles universally hold, not the other way around. We could call this the *bottom-up* view of these higher-level principles.

An analogy from architecture might be helpful to illustrate this point: insofar as the placement of windows in a building follows higher-level principles, it is not because the principles somehow in themselves have a special power to make the windows have the right positions. Rather, it is because of the position of the windows that the higher-level principles hold. Further, insofar as the higher-level principles could be said to have a causal efficacy to determine the placement of the windows, it is only via the causal powers of intelligent agents, such as the people who constructed the building. The lower-level laws of nature, which describe the orderly behaviour of natural phenomena, are like the windows in our analogy: they have just the right form to allow them to be jointly encompassed by a few simple higher-level principles.

One reason for claiming that these principles have no intrinsic causal powers is that except for being an intention or thought in some mind, human or transcendent, it is difficult to see how these higher-level principles could be anything over and above merely the patterns into which the laws and phenomena of nature fall. For example, they do not appear to be reducible to the causal powers of actual entities, as some philosophers claim about the laws of nature.[6] Instead, insofar as entities possess causal powers, the principles describe the arrangement of the causal powers of a diverse class of such entities – e.g., the fundamental particles – and therefore cannot be the powers of any given entity.

Even if they did describe the causal powers of some single type of entity, say, a superstring, the above argument would still apply. Although one would not have a diversity of substances with differing casual powers in this case, one would still have a diversity of causal powers, or at least a diversity of actions of a single causal power, instantiated by one substance. Thus, it would still be surprising that the action of those powers could be captured and unified by a few simple higher-level mathematical rules. This would be

analogous to how even if the fine-tuning of the constants of physics for life were to be explained by some grand unified theory, it would still be very surprising that the grand unified theory that happened to exist was one that yielded values for the constants that were intelligent-life-permitting.

One might wonder to what extent appealing to these higher-level physical principles adds to the case for design over and above merely appealing to the simplicity of the laws of nature. One potential worry with merely appealing to the simplicity of the laws of nature is that, as mentioned above, one could somewhat plausibly postulate that the existence of simple laws is a natural consequence of an ontology in which the laws of nature reflect the causal powers of substances, such as that presented by Harré and Madden (1975). In the case of Newton's inverse square law, for instance, one might initially think that the structure of three-dimensional space is responsible for Newton's law having the form of an inverse square law: given that a force is transmitted by quanta of energy from its source, as within our current understanding of forces, we would expect the force to fall off with the square of the distance from the source – that is, with the surface area of sphere around the source. Of course, from Einstein's theory of general relativity we know that this reasoning cannot tell the whole story, since within general relativity, gravity is explained by the way in which mass-energy bends space and time. One cannot derive the equation for this bending – Einstein's equation – simply by appealing the geometrical structure of space. Nonetheless, within a yet-to-be-worked-out quantum understanding of gravity, the approximate validity of Newton's law of gravity could possibly be seen as a result of the geometrical structure of space.

Laying aside these qualifications, the above reasoning does illustrate one way in which one might contend that the simplicity of the laws themselves need not necessarily be surprising. For example, by analogy to the above line of reasoning in the case of Newton's law, one might contend, the very conditions required to instantiate causal powers in substances require that the laws take on simple forms, though we cannot now see why this is the case. It seems much more difficult, though not impossible, to make such an argument in the case of these higher-level principles. One would have to argue, for instance, that some unknown constraints on instantiation of causal powers operate at a global level, requiring that effectively diverse causal powers fall under simple higher-level principles.

Another way in which one might try to explain the 'fine-tuning' of simplicity and elegance discussed above is by appealing to the metaphysical many-universes hypothesis. The problem with such an explanation is that there is no reason to think that intelligent life could only arise in a universe with simple, elegant underlying physical principles. Certainly a somewhat orderly macroscopic world is necessary for intelligent life, but there is no reason to think that this requires a simple and elegant underlying set of physical principles. This is especially clear when one considers how radically different the framework and laws of general relativity and quantum mechanics are

from the world of ordinary experience: although the regularities of the everyday world are probably derived from the underlying laws of quantum mechanics and general relativity, they do not reflect the structure of those laws. Indeed, it is this difference in structure between the classical, macroscopic world and the quantum world that has largely given rise to the interpretive problems of quantum mechanics. Thus, there is little reason based on an observation selection effect to expect the sort of macroscopic order necessary for intelligent life to be present in the underlying, microscopic world.

Finally, the form of argument in this case for design can be cast in the same form as that in the case of the fine-tuning of the constants for intelligent life, except unlike in the case of the fine-tuning for intelligent life, this fine-tuning cannot be explained by a many-universes hypothesis. One way of putting the argument is in terms of the 'surprise principle' we invoked in the argument for the fine-tuning of the constants of intelligent life. Specifically, as applied to this case, one could argue that the fact that the phenomena and laws of physics are fine-tuned for simplicity with variety is highly surprising under the non-design hypothesis, but not highly surprising under theism. Thus, the existence of such fine-tuned laws provides significant evidence for theism over the non-design hypothesis. Another way one could explain this argument is as follows. Atheism seems to offer no explanation for the apparent fine-tuning of the laws of nature for beauty and elegance (or simplicity with variety). Theism, on the other hand, seems to offer such a natural explanation: for example, given the classical theistic conception of God as the greatest possible being, and hence a being with a perfect aesthetic sensibility, it is not surprising that such a God would create a world of great subtlety and beauty at the fundamental level. Given the rule of inference that, everything else being equal, *a natural non-ad-hoc* explanation of a phenomenon x is always better than no explanation at all, it follows that everything else being equal, we should prefer the theistic explanation to the claim that the elegance and beauty of the laws of nature is just a brute fact.

Two objections

Above I have presented an initial case for thinking in fact that the laws of nature are in some sense 'fine-tuned' for beauty and elegance in the form of simplicity with variety. Here I would like to consider two objections, one raised by physicist Steven Weinberg and one raised by philosopher Nick Bostrom.

Weinberg's objection

One major objection to the above argument for design from the apparent beauty and elegance of the mathematical structure of physical reality is

considered in Weinberg (1992). According to this objection, this so-called beauty in nature is purely subjective, simply the result of our reading into nature anthropomorphic patterns in the same way as humans have read various meaningful patterns – such as the Bear or the Big Dipper – into the random pattern of stars in the night sky. One major problem with this explanation is that it does not account for the surprising success of the criterion of beauty in the physical sciences. We would not expect patterns that are merely subjective to serve as a basis for theories that make highly accurate predictions, such as quantum electrodynamics's successful prediction, to nine significant digits, of the quantum correction to the g-factor of the electron. The second problem is that there are significant objective aspects of beauty, at least in the classical sense of beauty, that one can clearly demonstrate in the realm of physics, such as that of simplicity and symmetry.

Bostrom's objection

Another objection to the above argument is what I will call Bostrom's objection, since it was first presented to me by philosopher Nick Bostrom. According to this objection, it is an epistemic rule that we should always choose the simplest theory, and we typically assume that the simplest theory is more likely to be true or at least empirically adequate than its contenders. This shows that we take simpler theories, and simple laws, to have higher epistemic probability than complex ones. Hence, the premise of the above argument that we should somehow be surprised by the simplicity of the laws of nature is flawed, since we normally take simplicity to be more epistemically likely than complexity.

There are several major responses one could give to this objection, two of which I will present here. The first response to Bostrom's objection is that it was belief in some sort of 'design' that drove the belief in a simple underlying order. As Morris Kline, one of the most prominent historians of mathematics, points out:

> From the time of the Pythagoreans, practically all asserted that nature was designed mathematically ... During the time that this doctrine held sway, which was until the latter part of the nineteenth century, the search for mathematical design was identified with the search for truth.
> (1972: 153)

Today, this belief in an underlying simple mathematical order to the world could simply be based on the success of science. Thus, although today we have a strong belief that the phenomena of nature can be captured by some underlying simple theory, that belief was originally motivated by an implicit belief in some sort of guiding intelligence.

The second response is to note that it confuses two different claims regarding epistemic probabilities. The first claim is that it is epistemically

likely that a particular hypothesis H1 is true or empirically adequate, given that it is the simplest available explanation of some set of data E. The second claim is roughly that there exists an elegant, simple underlying mathematical scheme from which we can derive (with appropriate initial conditions) relevant aspects of the observable phenomena (a scheme is elegant and simple if it has enough simplicity and elegance to be usable by human beings, impresses us as particularly simple, etc.). The supposedly high probability of the first claim does not entail the high probability of the second. It could be the case, for instance, that the phenomena in some domain D cannot be captured by any simple and elegant mathematical scheme, and yet it could also still be the case that we should choose the simplest available scheme that explains the relevant phenomena. In other words, we can still be amazed that the phenomenal world has a character such that it can be 'captured', within appropriate approximations, by a simple mathematical scheme.

The other alternative is to attempt to explain the simplicity of the world by some sort of metaphysical principle according to which the world is more likely to be simple than complex. One problem with this view is that there are many, many simpler possible worlds than ours, such as one with a single particle that simply travels in a straight line. The enormous actual complexity of our world thus strongly testifies against this claim. One way of getting around this problem is to combine this metaphysical principle of simplicity with a metaphysical many-universes hypothesis, according to which all possible mathematical structures are instantiated in some universe or another. One such view is suggested by University of Pennsylvania astrophysicist Max Tegmark (2003b: 487–489).[7] Using this hypothesis, one could explain the actual complexity of the universe by claiming that only sufficiently complex universes could contain embodied observers.

Even granting such a metaphysical many-universes hypothesis, however, such a metaphysical principle runs into severe problems. For one thing, simplicity seems to be conceptual framework relative, and thus it is difficult to see how there could be any such metaphysical principle. Any mathematical equation, for instance, can be written in a simple form, given one constructs the right mathematical properties. For example, consider the equation $Y = 2x + 4x^2 + 7.1x^5$. Define F(x) as by the expression on the right-hand side of the equation. Given that the concept of F(x) is part of our mathematical repertory, we can write the above equation as $Y = F(x)$, which is much simpler than our original way of expressing Y. The only way I can see around this problem is to postulate a set of primitive mathematical properties, and then define complexity in terms of the shortest bit string that would specify the mathematical equation using only those primitive properties. Without such a postulate, simplicity will be relative to the repertory of mathematical properties with which one has to work. An example of this viewpoint dependence occurs when Newton's law of gravity is translated into the framework of general relativity, and vice versa. When

this is done, however, their respective simplicity vanishes. As Misner *et al*. point out (1973: 302–303), expressed in the conceptual framework of general relativity, Newton's gravitational law is extremely complex. On the other hand, they point out, if expressed in the Newtonian framework, 'Einstein's field equations (ten of them now!) are horrendously complex' (ibid.: 303). So the respective simplicity of each is dependent on the conceptual framework in which it is written.

Other problems plague this appeal to a principle of simplicity. For example, the universe appears to be infinitely complex when you take into account the complexity of the initial conditions – such as the initial continuous distribution of mass–energy at some chosen surface of constant proper time. Such a distribution would be continuous, and so would take a non-denumerably infinite amount of information to specify. So, one's metaphysical principle of simplicity would have to be much more restrictive, such as dictating that some global feature – such as the way the distribution of mass–energy develops with time – can be described using a simple rule, which makes it even less plausible.

The fact is that simplicity in this context seems to be relative to both our conceptual framework and to our way of breaking up the elements of physical reality (such as in terms of laws and initial conditions). Since according to theism, our minds and the world ultimately owe their origin to God, it makes sense that the universe would at some deep level be reflective of the preferences of the human mind. This partially anthropomorphic nature of simplicity as occurs in physics, however, does not fit well with Tegmark's proposal of a mind-independent principle of simplicity that determines proportions among possible worlds.

Finally, I would like to conclude with a brief comment on the nature of the design argument from physics and cosmology. Overall, I think that the arguments covered in this paper, along with some other arguments which we did not cover, such as the intelligibility of the universe, provide a strong cumulative case argument for theism. Of themselves, however, the arguments I have offered do not prove the truth of theism, or even show that theism is epistemically warranted or the most plausible position to adopt. To show this would require examining all the evidence both for and against theism, along with looking at all the alternatives to theism. They do, however, provide significant evidence for theism, and hence should confirm the faith of those who already believe while providing a significant challenge to the atheist.

Acknowledgements

Because this paper was originally prepared for presentation at the conference 'Analytic Philosophy Without Naturalism?,' Milan, Italy, June 11–13, 2003, some points and arguments raised in this paper overlap previous publications. I would like to thank Sergio Galvan and Antonella Corradini

for a productive conference and for editing the conference proceedings for publication.

Notes

1 See, for instance, the article by Tegmark (2003a). Also see Tegmark (1998, 2003b).
2 That is, they are hypotheses that were not constructed merely to account for the data E in question.
3 Unlike what Mellor (2003) assumes in his objection to Martin Rees's claim that cosmic fine-tuning supports the multi-universe hypothesis, conditional epistemic probability is not a measure of ignorance. Rather, it has to do with relations of support or justification between propositions. As the famous economist John Maynard Keynes stated in his treatise on probability, 'If a knowledge of *h* justifies a rational belief in *a* of degree . . . , we say that there is a *probability-relation* of degree . . . between *a* and *h*' (1921: 4). Although I think Keynes's account needs to be further worked out, I believe his account is on the right track. For a recent discussion of epistemic probability, see Plantinga (1993b, Chapters 8 and 9).
4 One might object to this argument by claiming that because of the existence of evil, we do not know that the existence of embodied, conscious, intelligent beings is a good thing as the argument seems to assume. To address this objection, let E represent the claim that embodied, conscious beings exist and let E' represent the claim that embodied conscious beings exist in the universe *and* that their embodied existence is a positive good (that is, something that increases the overall value of reality). Then, it follows from the above argument that claim E' is not highly improbable under theism. But, E' entails E, and we know that for all types of probability, if one claim R entails another claim S, then the probability of S is greater than or equal to that of R. Hence, if E' is not highly improbable under theism, then it follows that E isn't either.
5 Typically, a wide variety of laws can be constructed that meet the higher-level principles in question. For example, Newton's law of gravity can be formulated in such a way that the general principles behind Einstein's theory of general relativity are met, but within the mathematical framework of general relativity it takes on a 'horrendously complex form' (Misner *et al*. 1973: 303). What one typically cannot find is a variety of simple laws that meet the higher-level principles. This is why physicists can often find the correct equation by requiring that it be the simplest equation that meets the higher-level principles in question, such as the gauge principle in the case of the forces of nature (see below).
6 For example, see Harré and Madden (1975). Under the conception of laws as expressing causal powers, Einstein's Equation of general relativity would be seen as being grounded in the causal powers of matter to bend space-time.
7 The idea along these lines that Tegmark considers is that of giving simpler worlds – or classes of simpler possible worlds – a larger probability measure than complex worlds (2003b: 488).

Metaphysical presuppositions of the argument from design

Comment on Collins's paper

Sergio Galvan

Collins's paper raises numerous issues of interest, in that it concerns the boundary area between physics, on the one hand, and philosophical reflection, on the other. However, I shall not deal with physics in the strict sense. I shall concentrate instead on a number of points regarding the philosophical aspects of the paper. These points concern in particular the logical-epistemological and metaphysical aspects of the argument from design applied to the fine-tuning case.

Observations on the logical-epistemological structure of the argument

Discussion of the argument from design (see Manson 2003: 1–23) comprises essentially two types of interpretation: the Bayesian argument, and the argument from likelihoods. Common to both interpretations is the base concept of conditional epistemic probability as a measure of the degree of belief in something under the truth condition of something else. An important characteristic of this epistemic notion – to which Collins also refers (see note 3 of Collins's paper, where Keynes and Plantinga are cited) – consists of its objective validity. In short, the probability of a with respect to b as expressing the degree of a subjective belief expresses the degree of belief of a rational subject: that is, of a person who in formulating her belief in a on the basis of belief in b takes account of all the means (deductive, statistical, inductive, etc.) available with which to obtain objective information about the world. Of course, this is not to have the epistemic notion coincide with some objectivist notion of probability (like the frequentist or propensional ones). The objective validity of epistemic probability guarantees at the same time that cannot be purely subjectivist.

Moreover, the probability used in the argument must be objectively founded; for otherwise the argument is thwarted. An essential point in all models of interpretation of the argument from design based on fine-tuning consists, in fact, in the very low probability attributed to the fine-tuning itself. It would be senseless to conceive this probability as a purely subjective estimate, because in that case there would be no explanatory basis for

the argument, which would not be able to function. The very low probability of the fine-tuning (on the non-theistic hypothesis) should on the contrary be an objectively verifiable phenomenon. As Collins puts it: 'for certain constants of physics, the range of intelligent-life-permitting values is purportedly small compared to some non-arbitrarily defined comparison range' (Collins, p. 142) (the range of values of some physical constants that allow the appearance of intelligent life is extremely small compared to a non-arbitrarily defined metric).

Now, it has been established that fine-tuning is regarded as highly improbable, given its complexity. However, for a fact (or an event) like fine-tuning to be the object of teleological explanation, its endowment with low probability is not enough. The event must be such to arouse surprise when it occurs; it must be, as Collins says, surprising. But on what does this element of surprise depend? The epistemological difficulties of the argument from design are connected precisely with determination of the qualities of an event that make it surprising. In general, it is said that the event must be significant, but the problem consists in giving satisfactory account of that concept. What is meant by saying that an event must be significant? I offer a clarificatory example taken from D.J. Bartholomew (Bartholomew 1984). Consider the outcomes of the independent spinning of five coins. There are 2 to the power 5 = 32 sequences. Significant among these may be sequences in which, for example, there are five H or five T, i.e. those sequences that may arouse suspicion that some intentional trickery is going on. But the important fact is that all the sequences have exactly the same probability, namely 1/32. If low probability were the only criterion to qualify for teleological explanation, all the sequences would be candidates, and therefore all of them would be explainable in terms of some intentional intervention. But, on the other hand, unsatisfactory, too, is the argument that significant events are those that are improbable under the hypothesis of randomness and more probable under some other hypothesis. In this case, in fact, all the cases in the example could be the object of some intentional design and therefore in that sense significant.

There is in effect something obscure in the notion of significance, and in my view it cannot be clarified unless we accept (with all the relative ifs and buts) that events can be significant from the point of view of their value, so that we may say that value is the root of their significance. Consider the case when a winning card (for one of the players) is drawn from the pack during a card game, for instance, poker. If a large number of apparently random draws always give the same card, the suspicion arises that someone is cheating. Why? Because the winning card serves the purpose of winning the game, and for the winning player this is obviously a value. In a competitive game, each player aims to win. The situation is therefore as follows: the sequence of extractions of winning cards is significant because drawing the winning card is an event embedded in an end-directed (telic) structure relative to one of the players. Of course, this account entails two presuppositions that are

rather hard to accept. First, we have to assume that among all the possible events (or states of affairs) only a few of them are endowed with value, while the majority of the events (states of affairs) are value-neutral. Note that this presupposition explains the surprise effect in entirely natural manner. Events endowed with value cause surprise because they are rare, and therefore because the probability that *some* value-endowed event will happen (disjunction among positive events) is low. By contrast, neutral events are not surprising because there are many of them. Consequently, it is highly likely that *some* neutral event (disjunction among neutral events) will occur. In other words, even if the neutral event that has occurred is endowed with low probability, its occurrence is not surprising because the event has happened as one of the many events that could do so outside the area of positive events. With respect to the value represented by the end of the telic structure in question, the fact that one neutral event rather than another occurs is insignificant.

The first presupposition does not raise particular difficulties. It is a matter of fact, for example, that the outcomes of the distribution of cards can be divided on the basis of the category of being convenient to one of the players. But the naturalness of these examples is tied to the artificiality of the end-directedness established by the game, which is established by the rules of the game and by the will of the player to win. It is instead very difficult to claim that an end-directedness can be discerned in the reality of things, and especially in the evolution of the universe itself. It is here that the demanding requisite of the second presupposition comes into play. In the specific case of the theistic argument from design, we must assume the objectivity of value attribution. Suppose that the value constituted by evolution towards increasingly higher forms of life is tied to our anthropomorphic view of things. It follows that the event that we intend to explain in teleological terms would be equivalent to others that we would be perfectly willing to explain in terms of randomness. The intent to explain it teleologically would therefore be an illusion. In other words, if the value perspective were only subjective, the distinction among random events and intentional events (objects of intention because endowed with value) would be apparent and consequently useless. In his paper, Collins grasps the crux of the problem when he replies to Weinberg on the question of the subjectivity or otherwise of beauty (as a form of simplicity in variety) in the theories of physics. But I ask myself whether it is possible to answer the question if we confine ourselves to consideration of scientific knowledge alone: science, at least if Hume's law is valid, is unable to make value statements.

The previous observation on the axiological meaningfulness of fine-tuning is also crucial from the point of view of the principle of likelihood that, according to Collins, characterizes the argument from design (in contrast to a Bayesian interpretation of it). For Collins (see Collins, p. 142 and Collins 2002b), the argument can be formalized as follows:

1 Premise: the existence of fine-tuning is not improbable under the theistic hypothesis.
2 Premise: the hypothesis of fine-tuning is highly improbable under the atheistic hypothesis of the universe.

Conclusion on the basis of the Confirmation Principle: fine-tuning makes more likely (confirms) the theistic hypothesis compared to the atheistic one based on one single universe.

And yet the plausibility of the first premise derives only from the fact that the author of the design is guided by an axiologically oriented intentionality and that the value of conscious life is part of that axiology. Hence, in order to estimate the probability of fine-tuning on the theistic hypothesis, it is not enough to have empirical information on the initial conditions, nor that such information is enriched by the contribution of cosmological theory. Required instead is a specific form of evaluative knowledge which science is unable to furnish. Satisfactory analysis of the argument from design requires, I believe, adequate treatment of the problems inherent in that form of knowledge.

A third question concerns the non-Bayesian character of the formulation of the argument proposed by Collins. As it is well known, if one wants to see whether the evidence confirms a certain hypothesis H1 or an alternative hypothesis H2 (which, if it is the only one, coincides with non-H1), one can adopt two strategies: (1) try to calculate the probability of E given H1 (i.e. the likelihood of H1 on the basis of E), which divided by the probability of E (i.e. set in ratio to the surprisingness of E) expresses the importance of H1 in explanation of E. If this value is greater than the probability of E given the alternative hypothesis H2, one concludes that the evidential datum E confirms H1 more than H2; and (2) compare the value of the Bayesian confirmation of H1 given E with the value of the confirmation of H2 given E and choose the better one. In this case, however, it is not the likelihoods that are being compared but the confirmation functions, in the determination of which an important role is played by the initial (or *a priori*) probability of H1 and of H2. The confirmation function is higher, the greater the initial probability.

Collins chooses the former strategy, so that it is entirely natural for him to affirm that there is a difference between saying that the same evidential basis E confirms H1 more than H2, and saying that H1 is the best explanation of E. Collins says that acceptance of H1 as the best explanation (i.e. that p(H1,E) > p(H2,E)) entails consideration of numerous factors that lie beyond confirmation by E of H1 (from that fact, that is, that p(E,H1) > p (E,H2)). As said above, these are the factors that help determine the initial probabilities. In the case of the Darwinian hypothesis versus the theistic hypothesis, for example, the fact that the evolution of living species and many other biological phenomena is explained better in Darwinian terms than in teleological ones depends on the fact that the initial probability of

the Darwinian hypothesis is much greater than the probability of the theistic hypothesis. The Darwinian hypothesis has been confirmed by a large body of independent evidence, and it is much less demanding than the theistic one. In the light of this example, therefore, one is already justified in believing that consideration of the initial probabilities is essential. But then the non-Bayesian analysis based only on the principle of likelihood of the argument from design does not seem sufficient.

Replies to the objections against fine-tuning

In the second part of this reply I shall briefly discuss three objections to the argument from design that Collins seeks to answer.

> *Objection under the hypothesis of a single world*: There exists some law (unified theory of physical forces) on the basis of which the fundamental physical constants are what they are.

Reply: Collins argues that this objection fails because it simply shifts the object to be explained by the argument from design to a lower level. The object to be explained by the argument from design is no longer the fine-tuning but the structure of the world described by the unified theory.

Observation: The underlying question remains unanswered, however, namely whether the structure is ontologically necessary or contingent. Only if it is contingent does the problem of its explanation arise, so that it is advisable to address the argument from design at this level; but not in the case in which the structure described by the theory is understood to be something that is ontologically necessary. In fact, explaining means necessitating, so that it makes no sense to explain a necessary state of affairs. Second, as we have seen, one of the presuppositions of the teleological argument is the contingency (low probability) of the explanandum. In his paper Collins, citing Aitchison and Hey, argues that the gauge principle (which is one possible component of the theory) is not logically true, concluding that it is a contingent principle of great simplicity. This strikes me as indicative of the fact that ontological necessity is to be excluded. But then the argument from design must show that physical necessity is ontologically contingent.

Note that the requirement of the exclusion of the ontological necessity of the physical structure depends on the fact that if the structure were necessary, then it would no longer be surprising, and therefore should not be the object of any kind of explanation and therefore not teleological either. In other words, the probability of the necessary structure under the argument from design is always the same as the probability of the necessary structure under any other hypothesis. Note also that this does not imply any confusion between epistemic probability and logical (or ontological) probability, only a correct connection between them. That an event A is epistemically

unlikely signifies that one does not expect A to happen, and therefore if it does happen, one is surprised. However, if epistemic probability is a rational form of probability, the fact that event A is not part of our expectations must depend on some objective reason grounded *in re*. However one understands this reason, it cannot be compatible with the belief that the fact is ontologically necessary. Hypothesizing the ontological necessity of A means removing from A the reason for its surprisingness, so that to the extent that this hypothesis is allowed to hold – in that it cannot be confuted – introducing further hypotheses in order to reduce the degree of surprise is pointless.

> *Objection under the hypothesis of many physically possible worlds*: The physically possible worlds have all been actualized. Among them also exists the world in which the initial fine-tuning has ensured the existence of forms of intelligent life. We live in this world. It is therefore entirely natural and unsurprising that fine-tuning should have come about, given that it is necessary for each of the physically possible worlds to be actualized, and among them the world in which we exist as observers of it.

Reply: Collins counters this objection by asserting the necessity of a generator of many worlds which displays all the features of fine-tuning. The teleological argument is therefore needed to explain why there is a generator. However, one may respond to Collins's conclusion with an argument analogous to the previous one concerning the hypothesis of only one world. The argument from design can be legitimately extended to the existence of the generator of worlds only if that generator is contingent. But once again: what are the grounds for this contingency? If, on the contrary, the generator (with its laws) constitutes the original state, then the theistic hypothesis is no longer confirmed by the existence of that particular example of fine-tuning which is the generator itself. Of course, this is not to rule out that the hypothesis of an original generator of physical order may be impracticable for other reasons. For example, if the generator is a physical generator, it cannot be the matrix of possible worlds in which really emergent phenomena, as some consider the appearance of consciousness to be, would occur.

> *Objection under the hypothesis of many metaphysically possible worlds*. Collins dismisses the hypothesis of a multiplicity of worlds understood in the metaphysical sense as belonging to the realm of science fiction.

Reply: yet this hypothesis too, I believe, should be discussed, and to do so we must begin with an operation of conceptual cleansing. The hypothesis has two distinctive features: (1) a generator does not exist; and (2) instead of the generator there exists the ultimate horizon in which the set of the

various metaphysically possible worlds comes into being. But in what does this ultimate horizon consist, and what are the metaphysically possible worlds? Classically, a metaphysically possible world is a maximal set of states of affairs. By ultimate horizon is instead meant the untranscendable scenario within which the worlds are made present (actualised). On the other hand, precisely because the worlds are maximal sets of states of affairs, two different worlds are incompatible at the same time; they cannot be co-actualized, although they may perhaps be actualized at different moments. The consequence of the incompatability of different worlds is therefore the presence of time in the original horizon where the actualization of the individual worlds takes place: the appearance of the worlds within the ultimate horizon is a temporalized appearance. Second, the ultimate horizon is of eminently formal nature: it is, so to speak, a sort of pure fixed scenario in which the individual worlds occur, but it does not have corresponding ontological support. But is an original situation of this type possible? Classical philosophical analysis – inspired by the thought of Aristotle – has developed important considerations on the matter. A situation such as the one described is incompatible with an ontology founded on the ontological principle of sufficient reason (or non-contradiction). If the original horizon is a mere formal framework devoid of foundation – that is, devoid of a metaphysical generator – then the original category cannot be anything other than possibility. But this is classically impossible. In fact, pure possibility cannot be the ground for its own actualization.

To conclude: only the hypothesis of many metaphysically possible worlds seems to be conclusively confutable. The hypothesis of many physically possible worlds – i.e. the hypothesis of an original physical generator of those worlds – is only confutable if there is evidence to show the ontological contingency of that generator. The hypothesis is also confutable if evolution attests to emergent (and therefore not physically explainable) forms of being, as some consider consciousness to be.

Part IV

Philosophy of mind

7 Rational selves and freedom of action

E. Jonathan Lowe

Since the beginnings of the modern analytic movement in philosophy, with the work of Russell, Moore and Wittgenstein, analytic philosophers have tended to regard mathematics and the physical sciences as paradigms of progress in human knowledge, and as a consequence have tended to favour ontologies which are hospitable only to physical entities and such abstract entities as may be required to make sense of mathematical truths. But these are only tendencies, not requirements, of analytic philosophy. The central ideals of analytic philosophy – clarity, logical rigour, and the unbiased pursuit of truth – are consistent with richer ontological commitments, provided that such commitments can be rationally justified. In this paper, I shall try to justify belief in a richer ontology where mind and action are concerned.

Introduction

Is a naturalistic account of rational human action possible? Obviously, we can't answer this question without being told what the questioner means by a 'naturalistic account' of some phenomenon. For most philosophers, however, 'naturalistic' just means 'physicalistic'. For many of these philosophers, a naturalistic account of rational human action would be one which represented human actions as being wholly physical events with wholly physical causes and the rationality of an action could only have something to do with how and by what it was caused. For example, it might be held that every human action is simply a bodily movement of some kind and that such an action qualifies as rational just in case it was caused by the onsets of certain psychological states of the agent whose contents represented such an action as serving the agent's interests in the circumstances in which the agent found himself. For such an account to qualify as wholly naturalistic in the sense now under consideration, it would have to incorporate a physicalistic account of mental representation and of an agent's interests. Perhaps such an account of mental representation could be provided in causal or teleofunctional terms and perhaps such an account of an agent's interests could be provided in terms of evolutionary adaptation. I shall not, however, be

pursuing questions about mental representation any further in the present paper, because I think that a physicalist account of rational human action along the lines just proposed inevitably falls at an earlier hurdle, in virtue of endeavouring to explain the occurrence of such actions in wholly causal terms. I shall give my reasons for saying this shortly. Before coming to these reasons, though – which are, of course, reasons for *belief* – I need to say something about the notion of a reason for *action*. Only at the very end of the paper shall I say something about the connection between reasons for belief and reasons for action.

Acting for a reason

What is an action? What is a reason to act? And what is it to act for a certain reason? An action is something done by an agent, such as opening a door, raising an arm, uttering a sentence, or imagining a situation. Some actions, such as imagining, are mental actions. Others, such as raising an arm, are physical actions. An 'agent', in the sense now intended, is a psychological subject – something that can have thoughts, feelings, desires, intentions, urges, and so forth. Persons or selves are psychological subjects with a capacity for reason and self-reflection. And it is with such agents – that is, with beings like ourselves – that I shall be exclusively concerned in what follows.

On the issue of the nature of reasons for action, I espouse what would commonly be called an 'externalist' view. A reason for an agent *S* to perform some action *a* consists, typically, in some fact concerning the agent's circumstances. For example, if *S* is standing near to a dilapidated building, the fact that a loose slate is sliding from the roof and is about to fall on *S*'s head constitutes a reason for *S* to jump out of the way. Of course, *S* will be in no position to act upon that reason and jump out of the way unless *S* is aware of the impending danger. It seems safe to say, then, that only if *S believes* that the slate is about to fall on her head will *S* jump out of the way for that reason. But this is very far from saying either that *S*'s *belief* that the slate is falling is her reason, or part of her reason, for jumping or that this belief is a *cause* of her jumping. *That* an agent has a certain belief can be, but typically isn't, a reason for the agent to act in a certain way – but *a belief* as such cannot ever be a reason for action, because it is not even something in the right ontological category to constitute such a reason. And, although a belief can sometimes cause or help to cause some of an agent's behaviour, a rational action is precisely one that *isn't* caused by an agent's beliefs and desires. Rather, it is one that the agent *chooses* to perform 'in the light of' her beliefs and desires, which is a very different matter.

To expand a little upon these claims, consider first a case in which the fact that an agent has a certain belief constitutes a reason for the agent to act in a certain way. Here I can do no better than adapt an example of Jonathan Dancy's (Dancy 2000: 125). The fact that someone has the paranoid belief

that she is being pursued by alien beings from Mars is a good reason for that person to visit a psychiatrist – although, of course, it is unlikely that such a person will be able to act upon this reason, because it is unlikely that she will be aware that her belief is paranoid. By contrast, in our earlier example of the falling slate, it is not the fact that *S* believes that the slate is falling that constitutes a reason for *S* to jump out of the way: rather, it is the fact that the slate is falling – although, once more, *S* must be aware of this fact and so believe, or at least suspect, that the slate is falling if she is to be in a position to act upon that reason. The mistake that is characteristically made by 'internalists' is to regard such psychological states of the agent as constituting reasons for the agent to act, rather than the *contents* of such states – that is, those facts or states of affairs whose existence is revealed to the agent in virtue of her being in such psychological states. The mistake is twofold in character. First, it represents a reason for action as being a state of the agent, when in fact it is a state of affairs. Second, it represents a reason for action as always being psychological in nature, when in fact this is only the case in exceptional circumstances, such as that of the paranoid believer. Even in the latter sort of case, the relevant reason for action is a psychological *state of affairs* – the fact that the agent has the paranoid belief – rather than the psychological state of the agent, her paranoid belief.

Next let us consider a case in which an agent's beliefs do cause, or help to cause, some behaviour of the agent's. The example of the paranoid believer in persecuting Martians will again serve us well. An agent with such a paranoid belief may well be caused by that belief to act in various seemingly bizarre ways, such as hiding behind bushes when approached by strangers in the park. Now, of course, being pursued by Martians might well constitute a good reason to hide behind bushes and so we are in a better position to understand this agent's seemingly bizarre behaviour once we know that she has the paranoid belief. But we are not yet in a position to judge that the agent is acting rationally, or for a reason, once we know that she has this belief and that its content provides her with a reason for acting in the way that she does. If her paranoid belief is *causing* her so to act, then she is not *choosing* to act in this way in the light of that belief: she is simply being impelled by the belief and as such is not acting rationally. It is already well known that there is a problem of 'deviant causal chains' for those theorists who maintain that to act rationally is to be caused to act in a certain way by beliefs and desires which represent such an action as being in the agent's interests in the circumstances in which she finds herself. To use Donald Davidson's own famous example, a climber may be caused to let go of the rope supporting her companion by her strong desire to save herself and her belief that by letting go she would increase her own chances of survival – and yet the climber's behaviour may be unintentional, because it was only that the onset of this belief and desire so unnerved her that her grip slackened.[1] It has proved notoriously difficult to say, quite generally, *how* beliefs and desires must cause behaviour if the behaviour is to qualify as rational

action performed in the light of those beliefs and desires. What I am now contending, however, is that there *can* be no correct account of rational action in such causal terms, because an action's being *caused* by the agent's beliefs and desires is incompatible with its being an action which the agent *chose* to perform in the light of those beliefs and desires – and no action is rational if the agent does not freely choose to perform it. An action can be *in accordance with reason*, or *reasonable*, without being rational. To be rational, it must be done *for* a reason which the agent freely chooses to act upon – or so I claim. For example, jumping out of the way of a falling slate is a reasonable thing to do, but it is not done for a reason and so rational unless the agent chose to jump out of the way in the light of her belief that the slate was falling.

This may seem to turn on its head a famous argument of Davidson's in defence of the doctrine that 'reasons are causes'.[2] And so it does. The argument points out that it is not enough to explain why an agent acted in the way she did, just to point out that the agent was aware of certain reasons for so acting. For the agent might well have been aware of various different reasons for so acting and, indeed, of various reasons for acting in other ways instead. To explain why the agent acted as she did, we need to know which of these reasons was the reason for which she actually did so act. We are supposed to conclude that the reason for which she actually acted was the one that actually *moved* her so to act, that is, the one that *caused* her so to act – which would have to be a psychological state of the agent, such as a certain combination of belief and desire. However, I have just maintained that being *caused* to act by certain of one's beliefs and desires is in fact incompatible with one's acting rationally, even though it may be compatible with one's acting 'reasonably', or 'in accordance with reason'. And the problem of deviant causal chains is grist to my mill here. But this still leaves me needing to answer the question that Davidson's argument poses: which of those reasons for acting, of which the agent may have been aware at the time of acting, was the reason for which she actually acted, if there was one? But my answer to this is straightforward: the reason for which the agent acted is the reason which the agent *chose* to act upon. Normally, the agent herself will be able to tell us which reason this was. In situations in which an agent is aware of a variety of reasons for acting in a certain way, and other reasons for acting in other ways, the agent is confronted with a choice as to how to act and on which reason to act. By choosing to act in a certain way in the light of a given reason, the agent *makes* that reason the reason for which she acted on that occasion.[3] This is something that the agent herself brings about through her mental act of free choice.

Objections and replies

I am aware, of course, that this talk of a 'mental act of free choice' will ring alarm bells with all right-thinking naturalists. I shall say more about it in

due course. Let me say at once, though, that talk of such mental acts raises no spectre of a vicious infinite regress. There is no danger of our being compelled to say that each mental act of choice must itself be the consequence of a prior act of choice. If choices are to play the role in rational action that I have assigned to them, they must be uncaused and so 'free' in the libertarian sense of 'freedom'. *A fortiori*, they cannot be caused by prior choices. Choices can be causes of subsequent events, but they cannot be effects of prior ones. I shall return to this matter shortly. First, there are some other matters to clear up.

A reason for action, I have claimed, is some fact or state of affairs. More particularly, it is a fact or state of affairs which justifies a particular course of action – as, for example, the fact of the slate's falling justifies the action of jumping out of its way. But two problems are likely to be raised for me here. First, what do I say about an agent who jumps in the *mistaken* belief that a slate is about to fall on her head? Second, how can we say that it is rational, *tout court*, to jump out of the way of a falling slate: mustn't we say, rather, that it is rational to do this *if* one desires not to be hit by it? In which case, mustn't we admit, after all, that psychological states of the agent are at least partly constitutive of an agent's reasons for action?

In answer to the first question, I am inclined to say the following. Whether or not an agent is mistaken in believing that a slate is about to fall on her head, *that* a slate is about to fall on one's head is, by any standard, a good reason to jump out of the way. If an agent is to act rationally in a certain way, she must be aware of a reason for so acting. But 'being aware of a reason' in this sense doesn't require that the reason should actually obtain. Here we might usefully deploy Alvin Plantinga's way of talking about facts and states of affairs.[4] A state of affairs may or may not actually obtain, but if it does, it is a fact and may consequently be represented by a true proposition. Reasons for acting are states of affairs and one must be aware of them if one is to act rationally. If the world is as the agent believes it to be, a reason for acting of which she is aware will be a state of affairs that obtains, and so a fact. If the relevant beliefs of the agent are false, she may still be aware of the same reason for acting and act rationally in the light of that reason, but the reason for which she acts will turn out to be a state of affairs that does not actually obtain.[5] In that case, the agent might well have done better to have acted differently, but she did not act irrationally. Since reasons are not causes, it does not matter that an agent's reason for acting in a certain way may on occasion be a state of affairs which does not obtain and hence which is not a fact. Reasons for action, even in the case of agents with mistaken beliefs, remain perfectly objective and, typically, non-psychological in character. (Here I should concede that I am glossing over the further question of what distinguishes 'good' from 'bad' reasons for action, if indeed such a distinction is a proper one to make. No doubt, for an action to be rational, the reason in the light of which the agent performed that action must at least be one that the agent *judged* to be 'good', in

the sense that she judged the state of affairs in question to justify acting as she did. But this is a further complication which I do not have time to go into at present.)

Perhaps my answer to the first question will not satisfy everyone. For those who are dissatisfied with it, I have a slightly different answer which they may prefer. (For my own part, I am presently undecided as to which is the better answer.) We could say that if an agent jumps out of the way in the light of a *mistaken* belief that a slate is about to fall on her head, then she does not *really* have a reason to act in that way on that occasion. At the same time, however, we can and should insist that such an agent is not acting *irrationally* merely insofar as her belief is false – for if it had been true, she really would have had a reason to act in precisely the way that she does. On this view of the matter, only *facts* – that is, states of affairs which obtain – can actually *be* reasons for action, but an agent can, nonetheless, act 'for a reason' which does not actually obtain and in doing so act rationally in precisely the same sense in which an unmistaken agent can. Most importantly, we do not have to regard the mistaken agent's *false belief* – a psychological state of the agent – as being her 'reason for action', and thus treat mistaken and unmistaken agents asymmetrically.

So let us turn to the second question raised a moment ago, namely, how a mere state of affairs, such as the falling of a slate, could constitute a reason for acting in a certain way save in conjunction with some appropriate desire on the part of the agent, such as a desire not to be hit by the slate. What I am inclined to say about this is similar to what I said about the agent's beliefs earlier: first, that a desire as such is the wrong sort of thing to constitute a reason for action and, second, that the fact that the agent has a certain desire, while it could constitute a reason, or part of a reason, for the agent to act in a certain way, does not normally, much less necessarily, do so. Here I am again simply following Dancy's lead. However, I am happy – as Dancy may or may not be – to say that some reasons for action manifest themselves to the agent's consciousness as the contents of desires while others do so as the contents of beliefs. In the case of the falling slate, the agent's awareness of this reason for jumping manifests itself as her belief, or at least suspicion, that the slate is falling. But jumping in these circumstances is a reasonable thing to do only for an agent in whose interest it is not to be hit by such a slate. Human beings can be seriously injured or killed by falling slates and it is consequently in their interests not to be hit by them. They will, correspondingly, tend to desire not to be hit by falling slates. The full reason for a human being to jump out of the way of a falling slate is that the slate will very probably injure her and it is in her interest not to be injured. However, although her awareness of this interest will manifest itself, at least in part, as a desire not to be injured, her action of jumping out of the way of the slate will only qualify as a free and so rational action if she *chooses* to jump out of the way in the light of this desire. If the desire merely *causes* her to jump out of the way and the power

of choice is not exercised by the agent on this occasion, her behaviour is undoubtedly reasonable, in that it furthers her interests, but it is not an instance of rational action. Or so I want to claim.

Another complication that I should at least touch upon here concerns the distinction between moral and practical reason. I have been talking so far only about reasons for action which consist in the agent's having certain 'interests' and her being in a certain kind of situation – for instance, in her being in danger of being hit by a falling slate and its being in her interest not to be injured by it. But sometimes we consider it rational for an agent to act contrary to her own interests, in the furtherance of some moral consideration, such as the requirement to alleviate the suffering of others. I am perfectly happy to endorse this view and to accommodate it by allowing that some reasons for action consist in moral facts or states of affairs. Of course, many naturalists think that the very notion of a 'moral fact' is at best obscure and at worst unintelligible – witness John Mackie's famous argument from 'queerness' against such facts.[6] Not being a naturalist myself, I have no such problem. This is not, of course, to say that nothing more needs to be said about how and why moral facts can exist. However, morality is not my present concern and I shall say no more about it. Many rational actions are purely prudential and have nothing of a moral character about them – and it is with such actions that I am now primarily concerned. What I now want to discuss is the fundamentally metaphysical question of how it is that 'free choice', which I take to be an essential ingredient in all rational action, is possible – and how, in particular, we can avoid the conclusion that mental acts of 'choice', assuming them to occur at all, would have to be either causally determined by prior events or else merely the random products of chance. For neither of these alternatives is compatible with the view, which I wish to support, that in acting rationally we freely choose between genuine alternatives in the light of the reasons for action of which we are aware.

Choice, causation and free agency

Here is the picture of rational action that I wish to endorse. In certain situations, a rational agent is confronted with two or more alternative courses of action, which are genuine alternatives in the sense that, at the time at which the agent must choose between them, no sufficient cause already exists of one of the prospective outcomes. The agent then deliberates, by weighing the reasons for and against the various different alternatives. Finally, the agent makes a decision and chooses to act in one way rather than another. Her mental act of choice may then, and normally will, contribute causally to a chain of subsequent events issuing in the performance of the chosen action. So, for example, the agent is attending a lecture and is confronted with a choice between raising her arm to ask the speaker a question or refraining from raising her arm and instead listening to other

people's questions. The agent is aware of certain considerations favouring the first course of action and other considerations favouring the second. For instance, in favour of the first may be the fact that her question is an important one and no one else is likely to raise it. In favour of the second may be the fact that, being of a shy and nervous disposition, she is not as good as many other people at articulating questions clearly in public. Having deliberated a while, the agent then decides to ask her question, chooses to raise her arm and, by this act of choice initiates a train of events which results in the rising of her arm. We are to suppose that, prior to the agent's act of choice, no causally sufficient condition of the arm's going up when it did already existed. The first moment at which there existed a causally sufficient condition of this event was the moment at which the agent chose to raise her arm. The agent's act of choice was perfectly free, in the sense that it had no antecedent cause whatsoever. Instead of that act of choice occurring, it was equally possible, given the state of the universe prior to that moment, that a different choice should have been made by the agent – a choice to refrain from raising her arm – in which case her arm would not have risen when it later did.

I suspect that many people will agree that this is how they think of themselves as proceeding when confronted with what appears to be a choice between two or more alternative courses of action. But does this picture really make sense? Naturalistic philosophers will almost certainly say no – and even many anti-naturalistic philosophers may be inclined to agree with them. Some will argue as follows. Suppose the agent chooses to raise her arm and her arm duly goes up. We need to ask *why* the agent chose to raise her arm. There must surely be some explanation as to why the agent chose as she did and the most plausible explanation is that the agent's mental state immediately prior to the act of choice caused it to occur. So, it may be said, at some stage in the agent's deliberations she formed the belief that, all things considered, asking a question was the sensible thing to do and this belief, together with her desire to do what she considered sensible, caused her to choose to act in the way she did, by raising her arm. However, now we are back to the picture of rational action according to which a rational action is one that is not freely chosen in the light of the agent's beliefs and desires but *caused* by the agent's beliefs and desires. And this is precisely the picture that I was at pains to reject earlier on. I think it is a picture that is radically false to the phenomenology of rational action and, worse still, impossible to square with any conception of rationality that we can really make sense of. That is to say, when we take ourselves to be acting rationally it never *seems* to us that we are being caused to act in the ways we do by our beliefs and desires, and once we suppose that, on a given occasion, we were in fact caused to act in a certain way by our beliefs and desires, we find ourselves obliged to withdraw any claim to have acted rationally on that occasion.

Let us go back to the question that precipitated these difficulties. The question was *why* the agent chose to raise her arm. I suggested that what

prompts this question is the thought that there must surely be some *explanation* as to why the agent chose as she did – and I then suggested that, at least for many philosophers, the most plausible explanation is that the agent's mental state immediately prior to the act of choice caused it to occur. But suppose the question is addressed to the agent and she is asked 'Why did you choose to raise your arm?' The agent will no doubt answer that she chose to raise her arm in order to ask a question. If pressed a little further, she will say, perhaps, that she chose to raise her arm because she had an important question to ask which no one else was likely to raise. In other words, she will cite certain reasons that she was aware of for raising her arm. Pressed still further, she may admit that she was also aware of certain reasons for refraining from raising her arm. 'So', we may ask her, 'why did you choose to act on the first set of reasons and not on the second?' She may answer, 'Well, after some deliberation I came to the conclusion that the first set of reasons was better than the second, so I chose to act on the first.' For most ordinary folk, this is where the questioning would come to an end. But philosophers, like little children, sometimes don't know when to stop asking 'why?' They may now ask, 'But why did you choose to act on the reasons that you thought were the better ones?' Here our agent, if she is not a philosopher, is likely to return a blank stare of bewilderment. If she says anything, it is likely to be sarcastic or rhetorical, such as, 'Well, don't *you* usually choose to act on what you consider to be the best reasons?' Our philosophical interrogator may now pounce and say, 'So, you couldn't help choosing as you did, once you came to the conclusion that one set of reasons was better than the other – in other words, your coming to that conclusion *caused* you to choose as you did.' But our agent is likely to resist this suggestion very forcefully and reply, 'No, of course I *could* have chosen to act otherwise and nothing *made* me choose to act as I did – though if I had chosen to act otherwise, it would have been against my better judgement, and sometimes I have done precisely that and later regretted it.'

Choice and chance

At this point, our philosophical interrogator, seeing that she is not going to get the agent to admit that anything *caused* her to choose as she did, may take another tack and try to persuade the agent that if her choice was genuinely uncaused, then it must have been somehow arbitrary – a mere chance event – and as such the very antithesis of anything rational. Moreover, if acts of choice are mere chance events, then how can any agent be said to have an *ability* to choose to act in one way rather than another? For to possess such an ability the agent must surely have some sort of *control* over how she chooses and no one can, almost by definition, have any sort of control over what happens merely by chance.

One way of pursuing this line of argument is through a thought experiment suggested by Peter van Inwagen.[7] We are to suppose that Alice is

faced with a choice between lying and telling the truth on a given occasion and, after much deliberation, chooses to tell the truth. The libertarian will want to represent this choice as being genuinely free, in the sense of being uncaused. This means that, according to the libertarian, the state of the universe prior to Alice's act of choice contained no causally sufficient condition of that act and, up until the moment of choice, there were possible futures in which Alice tells the truth and also possible futures in which she lies. Suppose, then, that all of the circumstances leading up to the moment of Alice's choice were to be replicated or 'replayed' a number of times – say, one thousand times. Since we are supposing that Alice's choice in the original 'play' of these circumstances was causally undetermined by preceding states of affairs, we have to suppose that Alice's choice is likewise causally undetermined by preceding states of affairs in each of the replays. So, it seems, we must suppose that in some of the replays she chooses as she did originally and tells the truth and in the others she chooses differently and tells a lie. It seems that there will be some specific number of replays in which she chooses to tell the truth – say, 513 – and in the remaining 487 replays she chooses to tell a lie. But it was arbitrary to pick one thousand as the number of replays. We can generalize and say that there will be some particular proportion of all the possible replays of Alice's situation in which she chooses to tell the truth. This proportion would have to be less than unity, for to suggest that Alice would have chosen to tell the truth in *every* possible replay of the situation is surely to imply that her choice was, after all, determined by preceding states of affairs. But if the proportion is less than unity – say, it is one half – then what this appears to signify is that Alice's choice in the original situation was simply a chance event whose objective probability of occurrence was 0.5, or 50 per cent.

The first thing I want to say about this argument is that, even if we can make complete sense of the notion of a 'replay' of the circumstances preceding Alice's act of choice and even if we can agree that, given that her choice was causally undetermined, it follows that in some replays she would choose to tell the truth while in others she would choose to tell a lie, it doesn't follow that there must be some determinate proportion of all possible replays in which she chooses to tell the truth. In other words – and avoiding the perhaps questionable notion of a 'replay' of Alice's situation – the fact that Alice might have chosen otherwise than she actually did does not imply that there was a certain numerically precise objective probability of her choosing as she did in the one 'play' that actually did happen. Consequently, it is far from clear that we are entitled to regard her choice as a 'chance event'. Compare Alice's situation with that which occurs in a genuine game of chance, such as a dice game. Let us suppose that on a certain occasion a perfectly fair die is thrown and it lands with the six uppermost. And let us suppose that its so landing is a genuinely indeterministic event, like the spontaneous decay of a radium atom – even though there is reason to doubt that this is really so. If we now contemplate all possible 'replays' of

this situation, we do indeed have some reason to suppose that a definite proportion of them – one-sixth – will result in the die landing with the six uppermost, and hence that there was an objective probability or chance of one in six of this happening when the die was actually thrown. Our reason to think so is obviously based on symmetry considerations concerning the structure of the die. However, nothing remotely like this reason is able to support a similar conclusion in Alice's case. It may be objected that we likewise lack any such reason to assign a numerical degree of chance to a nondeterministic atomic event, such as the spontaneous decay of a radium atom, and yet numerical probabilities are associated with such events. However, this is because large amounts of statistical data enable us to assign 'half-lives' to radioactive isotopes. Nothing similar even to this is actually available in the case of Alice. By being invited to contemplate a large number of 'replays' of Alice's situation, we are being invited to suppose that statistical data of this sort *is* in principle available concerning Alice's situation, albeit data that is not all drawn from the *actual* world, as in the case of the data which enables us to assign half-lives to radioactive isotopes: rather, the 'data' in Alice's case is, as it were, distributed across all the possible worlds compatible with the history of actual events up until the moment of Alice's choice. I suppose that there are some conceptions of possible worlds that would support the idea that such 'data' really exists, but for my own part I see no reason to think that it does. I see no reason to think that even God could assign an objective numerical probability or degree of chance to Alice's choosing to tell the truth.

The next thing I want to say about the Alice thought experiment is this. Alice, we have said, deliberated at some length before deciding how to act and choosing accordingly. The way in which the thought experiment is set up requires us to suppose that in all 'replays' of her situation her deliberations proceed in exactly the same way that they actually did, but that at the end of some of these replays she still chooses to tell the truth while at the end of others she chooses instead to tell a lie. However, if Alice deliberates and chooses rationally, at the end of her deliberations she will have formed a judgement as to whether the reasons in favour of telling the truth are better than those in favour of telling a lie. Now, certainly, if Alice's choice was genuinely undetermined, it was possible for her to have chosen to tell a lie in spite of having formed such a judgement. On the other hand, there is a clear sense in which, in the light of that judgement, Alice would have been acting irrationally in telling a lie. A libertarian should happily accept that our freedom to choose is a freedom even to choose irrationally – and we all recognize, I think, that sometimes we do make irrational choices. So-called 'akratic' action involves precisely this, it would seem, for such an action is one that the agent chooses to perform 'against her better judgement', as we say. At the same time, little sense can be made of the notion of a rational agent who as frequently, or as often as not, chooses to act irrationally as to act rationally.

We have to distinguish clearly between two different conceptions of how Alice might have chosen differently. She might have chosen differently even if the result of her deliberations had been the same, that is, even if she had formed the same judgement about which action had better reasons in favour of it. A libertarian must certainly say this. But another thing that can and should be said by the libertarian is that Alice might have deliberated differently and formed a different judgement as to which action had better reasons in favour of it, in which case, too, she might have chosen to act differently – only this time she would not have chosen 'akratically'. The reason why the libertarian should say this is that deliberation itself should be seen as a process which involves choice. Deliberation isn't simply a matter of the agent's being confronted with ready-made reasons in favour of doing this or that action and then having to weigh those reasons against one another. Deliberation involves the active seeking of reasons for or against carrying out one of a range of alternative possible actions. Consequently, it involves the agent in making choices about what sources of evidence to consult or think about with a view to revealing reasons for or against acting in a certain way. We can choose not only how to move our bodies but also how to direct our thoughts, and the latter kind of choice is essentially involved in all processes of deliberation. But if this is what deliberation really involves and if choice really is causally undetermined in the way the libertarian maintains, then the libertarian should not meekly accept the terms in which the thought experiment involving Alice is described. For the libertarian should reject the idea that one could, even in principle, set up circumstances at any time prior to Alice's act of choice in such a fashion that she was bound to go through the same process of deliberation as she actually did. This is not to say that the libertarian should redescribe Alice's episode of rational deliberation and action simply as a punctuated sequence of discrete indeterministic choices separated by deterministically evolving mental processes – first, choices about how to direct her thoughts in gathering and evaluating reasons for and against truth-telling and, finally, a choice, in the light of those reasons, whether to tell the truth or a lie. Rather, the whole train of deliberation which culminates in the final choice should be seen as a continuous indeterministic process, capable of evolving differently at every moment at which it is going on.[8] Seen in this light there is, I think, no intelligible way of representing what is going on with Alice as having any numerical degree of objective probability or chance associated with it, even setting aside my earlier objections to this notion.

The final thing I want to say about the Alice thought experiment is that the choice that Alice actually made was made in the light of reasons of which she was aware favouring the course of action that she chose – and that if she had chosen differently, whether or not after a different process of deliberation, she would still have chosen as she did in the light of reasons of which she was aware. Choice, by its very nature, is never exercised 'blindly', but is always informed by or responsive to reasons for action. In this respect,

it is utterly unlike any mere chance event, such as the fall of a die or the decay of a radium atom. It doesn't, I believe, make sense to think of an agent as *choosing* between two or more possible courses of action without being aware of any consideration whatever in favour of carrying out any of them. This is not to deny that an agent may find herself faced with a choice between two courses of action for which equally good reasons seem to present themselves – a 'Buridan's ass' situation – so that she can do nothing but make an arbitrary choice in favour of one of them. Nor is it to deny that we sometimes choose to do things 'just because we felt like doing them' – as, for example, when we choose to kick a small stone lying in our path as we stroll along. Even in this sort of case, our choice is made in the light of the consideration that no harm is likely to come from kicking the stone and that we shall get some small degree of pleasure or satisfaction from kicking it.

Choice, agency and control

But what about the notion that rational action requires that the agent be able to exercise *control* over what she does? Is the libertarian notion of choice at odds with this idea? That it is at odds with the idea of control is, of course, one thing that was meant to be shown by the thought experiment involving Alice. Alice's case was supposed to show this by showing that a free choice, on the libertarian conception of what such a choice involves, can only be a mere chance event – and, almost by definition, one cannot have control over a mere chance event. But even if we have succeeded in defending libertarianism against the charge that free choices as it conceives of them must be mere chance events, this doesn't of itself serve to explain how, or in what sense, according to libertarianism, an agent has control over what she chooses to do. However, the idea that what is needed is an account of how the agent has control over how she exercises her power of free choice is perhaps a confused one. According to libertarianism, it is precisely because we have a power of choice which we can exercise freely – that is, one whose exercises are not determined by prior events – that we have control over our actions. If our choices were causally determined for us by prior events, then indeed we would *not* have genuine control over our actions. In fact, our 'choices' would not really be *choices* at all, in any serious sense. Our choices, I have said, are necessarily informed by or responsive to reasons for action, but are none the less necessarily free, in the sense of being undetermined. The notion of free choice exercised in the light of reason provides us with the very paradigm of what it is to be 'in control' of our actions. To demand that we somehow be 'in control' of *our choices* seems either superfluous or incoherent. On the one hand, freely exercising our power of choice precisely *is* being 'in control' and as such needs no *further* exercise of 'control'. On the other hand, what is *in* control cannot, for that very reason, *be* controlled in its exercise of control, if this is understood as involving some further agency. If I say to somebody to whom I assign some special responsibility,

'OK, you are in control now', I cannot consistently then go on to say, 'But remember that I shall be controlling your every move.'

To this it may be objected that so-called 'Frankfurt-style cases' suggest otherwise.[9] Couldn't it, in principle, be the case that some mad scientist is monitoring my brain to see what choices I make, allowing only those choices to take effect which are compatible with what he decides should happen – but that it just so happens that everything I choose to do coincides with his plans? In that case, isn't there a sense in which I am controlling what I do, by exercising my power of choice, even though he has control over me because he would prevent me from doing anything incompatible with his plans? Notice, however, that even if this thought experiment makes sense – which may be questionable – there is no suggestion that the mad scientist has control over my power of choice, only control over its effects. He can make sure that if I choose to do something incompatible with his plans, my choice will be ineffective. But nothing has been said to suggest that he can make sure how I exercise my power of choice – what choices I make. Nor does it make sense that he could do this. For him to have, *per impossibile*, control over how I exercise my power of choice would be for him to *deprive* me of that power, so that this is, logically, not a sort of control that he can have over me. If he had such control over me that I no longer had any power of choice, *he* would now be the only one of us capable of possessing such a power: there would not be two distinct powers of choice, mine and his, with his the dominant power. He would be in control of me, but not in control of my power of choice, for I would have none. By the same token then, it would make no sense for *me* to be in control of my own power of choice, for then I would have to have two such powers, one dominating the other – and yet the supposedly dominated power of choice would simply be *extinguished*, not controlled, by the supposedly dominating power. This, I think, demonstrates that it is incoherent to demand of the libertarian that she provides an account of how we have 'control' over our power of choice. Having a power of choice gives us all the control we could ever have or need.

The pragmatic inconsistency of determinism

The final issue I want to address is this. Having argued that our intuitive self-conception is perfectly coherent in representing ourselves as beings whose rationality in action consists in a power of free choice exercised in the light of reason, the question may still seem to remain as to whether this picture of ourselves is true or false. We may have established that it *could* be true, but could it none the less be the case that it is in fact false and that in painting this picture we are deluding ourselves, perhaps in a way which we find psychologically unavoidable? To those who pose such a question I wish to pose the following question in return. Can we make sense of the thought that we might be confronted either with evidence of an empirical nature, or

arguments of a philosophical or logical character, which would be rationally compelling and speak in favour of the falsehood of the picture that has been painted? I do not believe that we can, because I do not believe that we can separate rationality of belief from rationality in action. To form a rational belief that we are not free to act in the light of reason, we should have to exercise rationality in action at the very least in directing our thoughts towards putative sources of evidence for that belief and in evaluating the deliverances of those sources of evidence, or in considering and evaluating arguments in favour of that belief. If we are not free to act in the light of reason, then we are not free to deploy our judgement in the light of reason in seeking out and assessing evidence and arguments for or against this or that belief. If, lacking freedom of rational action, we were to acquire the belief that we lack that freedom *not* through the free direction of our thoughts and the free use of our power of judgement, but rather as a consequence of prior causes determining the contents of our beliefs, then we would not have acquired that belief rationally and would not be rationally justified in holding it. There is a perfectly good sense, then, in which we simply *cannot* rationally believe that we lack freedom of rational action. If we came to believe that we lack that freedom, it would either be a false belief that we had acquired through a faulty exercise of our rationality, or else it would be a true belief which we did not hold rationally because, in virtue of its truth, we would not be rational beings. What cannot be the case is that we should hold the belief rationally and the belief be true. Nothing, therefore, can rationally commend the belief and reason demands that we dismiss it as false.

Notes

1 See Davidson (1973). The example appears on page 79 of Davidson (1980).
2 See Davidson (1980a).
3 Compare Searle (2001: 16).
4 See Plantinga (1974: 44ff).
5 Compare Dancy (2000: 131ff).
6 See Mackie (1977: 38ff).
7 See van Inwagen (2002).
8 I am grateful to Storrs McCall for insight on this point.
9 See Frankfurt (1969).

I see that Martians persecute me: What should I do, if I want to act rationally?

Comment on Lowe's paper

Lucia Urbani Ulivi

Lowe deals with an ancient problem, which may be traced back to Socrates, whose permanent presence in philosophy textbooks and papers suggests that a significant change is needed in the initial assumptions. In fact, all possible paths moving from these assumptions have been widely explored, but with limited success.

The problem is: when can an action be described as rational? Or, to choose a meaningful question: what distinguishes a rational action from an irrational one (Searle 2001)?

Interlacing reflections and questions provoked by the reading of Lowe's text, I shall express my comments in the following four points.

First, Lowe considers actions as actions, i.e. as acts performed by an agent; the agent under consideration is the human being, here indicated as a *psychological subject with a capacity for reason and self-reflection.* I want to offer some reflections upon such definition. As a matter of fact human beings are complex ontological data, which could be investigated using different cognitive strategies. Depending on the aspect under investigation and on the purpose of the search they are bodies (for the medical doctor), mental subjects (for the psychologist of intelligence), spiritual subjects (for the priest), juridical subjects (for the legislator), political subjects (for the head of a political party, mainly when a poll takes place), and the list could go on and on. In general, none of these strategies claims to exhaust what a human being is, but only to study, to interact with, or to get an idea of, what is the human being. Philosophy is not an exception, as it does not aim to exhaust the knowledge of the human being, although it reflects upon what makes a human being a human being, and not upon specific aspects of the human being. When a philosopher adopts a certain definition of human being, we expect – unless she differently justifies her choice – that, with her definition, she is referring to what makes a human being a human being, and to what, if missing, makes him not a human being. Under the assumption that the definition of Lowe is such, I am asking why he considersit relevant to define the human being as *psychological subject*. I can hypothesise that *psychological* should be intended as *mental, intentional* and maybe also *emotional*. If that is the case, Lowe is telling us that in his opinion the essential aspect of

the human being, which cannot be omitted by a philosophical definition and consideration, is to be a subject with mental, intentional (and emotional) states, with a capacity of reason and self-reflection. At this stage I would like to propose to widen and complete the philosophical definition of human being and I would ask to include the corporeity, as something, from a philosophical point of view, dangerous to forget. Nowadays neurosciences, psychoanalysis and cognitive psychology clearly show the interrelationship among rationality, emotions and physical states, their mutual implications and influences (Damasio 1994, 1999). Therefore I would propose to use in philosophy the following definition of human being: an organized complexity with a body, having mental, intentional and emotional states, with a capacity for reason and self-reflection.

Second, it is extremely important and clear the distinction that Lowe makes between belief as an *act of belief* and the content of such an act, which is *what is believed*. Beliefs, and we may add desires, suspicions, doubts, hopes, are acts performed by the subject, and as such they are never reasons to act, while what they represent is a reason to act. This is the *externalism* position embraced by Lowe. For the externalist, paranoid beliefs constitute a problem. In these beliefs the content of the belief is internal to the subject, so that in this case reasons for action are internal to the subject as well. I would like to propose to Lowe a solution to the problem. This would allow us to keep the externalist position also in the case of paranoid beliefs.

To say that the content of belief is different from the act of belief does not mean that we are engaged in the ontology of the content of belief. This was a frequent misunderstanding in philosophy, and became one of the strongest arguments for the sceptic: we cannot be sure of the content of belief, that can be anything, from dream, to imagination, to fairy tale. And then knowledge as knowledge *of something* is not reliable. I assert that there is no mistake in saying that our beliefs have contents, but that a mistake takes (or can take) place relative to the ontological category in which the content of a belief is to be allocated. Let me use the example of the text: 'I see a slate sliding'. The sliding slate is different from my act of belief, but to make this difference it is not sufficient to place the slate in the correct ontological location. In fact, the slate could be an empirical slate (it was lucky that I jumped out of the way!), it could be a perception error (it is only a newspaper!), it could be a stain in my visual field (I must go to my oculist!), or it could be a hallucination (a nightmare haunts me: many more slates!). In all these cases, with no exception, the externalist view remains valid, although we must go further in the following way: from the ontological standpoint, mental or emotional acts are private acts of the agent; what they represent is something ontologically different from the acts, and their ontology must be checked; the relationship between acts and what they represent is an intentional relationship, whose ontology consists in being a relationship. So explained, Lowe's position could either be called 'externalism' or could be named 'non-identity', as in this position reasons to act are

non-identical to the acts of reasoning that the subject performs when she reasons upon something.

I would like now to better explain the analysis of the paranoid belief, which in Lowe's example is: *Martians persecute me*. Such a paranoid belief is not a case in which the object of the belief is internal to the subject, but rather a case of categorical mistake. In this belief there is, as in all beliefs, an act of belief which is a part of the ontology of the subject, and as such it is not a reason to act; then there is the content *Martians persecute me* which is a state of affairs and, as such, a reason for acting. But facts belong to different ontological domains (both empirical things and dreams exist, but they belong to different domains of being), and the fact *Martians persecute me* could be categorized in the empirical domain, or in the domain of dreams, or in the domain of science fiction, or in the domain of virtual reality, and so on. The mistake the paranoiac makes is to categorize in the empirical domain a fact that belongs to the ontological domain of dreams: this is not to be considered as an exceptional internal fact (every night we dream and have such internal facts), but as an ontological mistake. This mistake could depend on an exceptional circumstance, i.e. that something that we see when we are awake belongs to the dream domain; in fact, what we see when we are awake belongs normally and mostly to the empirical domain. Differently from Lowe, I would say that the paranoiac who runs away because Martians persecute her is acting in a perfectly rational way, and that her behaviour confirms Lowe's externalist view. I assert consequently that the paranoiac needs to go to a psychiatrist in order to perform an ontological reclassification of Martians, who should be moved from the empirical domain to the hallucinations domain, or, for those who prefer an Aristotelian way of speaking, whose being consists in being imagined.

Third, the most complex critical crux that Lowe analyses is the relationship between reason, consciousness and freedom. In Lowe's opinion, the dynamic among them could be summarized as follows: the consciousness allows the exercise of reason, which, in circumstances where there is a possibility to choose, will act freely. If Lowe shares this description, I would recommend an amendment to what, in my opinion, seems to be a potential source of problems. If we assert that we are rational only when we are conscious, we must also assert that when we are not conscious we are not rational. This limitation of reason to consciousness does not allow us to explain several things we do without consciousness – not perceiving them – but that can be recognized as rational. The language of people who do not know grammar and syntax is rational; it is rational to accept a duty – without knowing why – it is rational is to go for a walk – for a reason we could not explain. If we invert the relationship between consciousness and reason – this is my proposal – and assert that consciousness is an outcome of reason, we could be entitled to say that non-conscious actions are also rational actions. The relationship governing reason, consciousness and freedom would then be re-stated as follows: in circumstances where the choice

is possible, there is a rational strategy, which could be adopted by the subject, consciously or unconsciously. I am aware that this proposal, while solving some problems, would create many as well, the most important being the redefinition of the concept of reason. In this view, reason would not only be the ability to follow logical rules – i.e. the concept of reason as introduced by Descartes and dominant till today – but also the ability to set and follow objectives – both theoretical and practical (Devlin 1997). Such ability does not disappear even when partially hidden from the consciousness. More than a proposal, this is a hypothesis to be tested.

Finally, Lowe states, with the support of a vast ethical tradition, that the choice is possible and therefore the action is free, when under certain circumstances there are no sufficient causes for a certain choice, even though there are reasons which allow us to make that choice. These reasons are not determinative, otherwise the action would not be free, but it is the subject that *gives weight* to them, therefore *she deliberates* and *she decides*. The possibility of freedom depends on the correct conceptual definition of *to give weight*, or of its equivalent *to deliberate*, *to decide*. If we intend *to give weight* as a shorthand for 'reasons which could be mentioned, but are not, for the moment', we bring the decision within the realm of reason. If we intend reason, as it usually happens, as the ability to follow rules, it inevitably follows that actions that are both free and rational are impossible. In fact, if actions depend on a mere rational appraisal, and *to give weight* means that there are other reasons which could be mentioned, but for the sake of brevity are omitted, then the choice merely consists in following the strongest reasons – it will be rational, but not free. If, on the contrary, actions come *after* a rational analysis, but they are not caused by it, *to give weight* means to privilege arbitrarily one reason instead of another one, then the choice is free, but irrational. We are confronted with a classic dilemma which, in my opinion, arises from two factors: the first is that we consider reason as the capacity to follow certain rules; the second is that we disregard an essential ingredient in the decision-making process: emotions. Nowadays neurosciences, particularly thanks to Damasio's work, offer philosophers a description of the decision-making process quite different from the classical one. We now know that, when we face a choice, pure reason provides lists of pros and cons, which tend to become endless (which is an indirect confirmation that there are non-binding situations, that there are free situations). We also know that choice is made thanks to the intervention of an emotion, which surgically severs the list of pros and cons and makes us take the decision. It is important to realize that the emotion not only helps the choice from the outside, but it is so deeply involved in the choice that it becomes an integral part of the rational behaviour. This is to such an extent that the lack of emotions generates irrational behaviour, as it prevents us from deciding and when we cannot decide, choose or act, we lose the ability to behave rationally.

I would propose to place *to give weight*, *to decide*, *to deliberate*, or alternatively the *gap* between reasons and choices, in the category of the emotions. This

seems to me the first step towards a philosophical anthropology able to integrate systematically, in the human being, all physical, emotional, rational and social functions. In this new perspective the duty of philosophy is to study the system, to be considered as integrated and complex, without reducing it to its individual functions as separate from one another and without restricting it to its intellectual, or emotional or physical properties.

8 Consciousness and freedom

Uwe Meixner

Two less usual questions regarding consciousness

What is the *meaning* of consciousness? Before offering some speculations regarding the (full) meaning of consciousness, I propose to tackle a somewhat less ambitious question: what is consciousness *good for*? More precisely: what is the advantage that a conscious creature can draw from being conscious? The nature of that advantage must surely be a part of the meaning of consciousness.

It seems undeniable that there must be some good in being conscious to some creature. For if there were no good to any creature in being conscious, why, then, does consciousness exist? The assertion that nothing existing in nature exists in vain is presumably a bit of an exaggeration. But it certainly seems hard to believe that consciousness exists as the rather widespread phenomenon in nature that it is, and at the same time has no positive function for any creature.

The evolutionary advantage of being conscious

It seems to me, on the contrary, that consciousness has at least one very robust positive function for all conscious creatures: consciousness enables them – not always, but more often than not – to survive long enough to contribute their genetic information to the genetic constitution of the next generation of their species. There are no tooth- and talonless cats around hunting mice. Why? Because tooth- and talonless cats could not survive long enough to produce offspring. There are, likewise, no non-conscious cats prowling over the lawn. Why? Because non-conscious cats could not survive long enough to produce offspring.

Not every living creature needs consciousness to ensure survival up to and including successful propagation. A tree does not need consciousness for that. But creatures that are constituted in such a way that they have the ability of wide-ranging self-locomotion and that cannot survive in nature without employing that ability cannot do without consciousness. They need consciousness for finding the food they feed on and for dodging the deadly

dangers their environment is replete with. Being conscious is a necessary condition of their survival, and therefore consciousness has a positive function for them and is a boon for them.

This seems obvious and uncontroversial. But, in fact, it is a bit surprising. For would not a well-balanced intricate network of reactive dispositions, installed in the brain of these creatures and answering in a differentiated life-preserving manner to a huge number of incoming complexes of stimuli, serve the same purposes that consciousness is said to serve? Many philosophers these days are only too happy to answer 'yes' to this question, adding that the findings of neurobiology more or less conclusively show that the possibility envisaged in the question is in fact the case.

Is consciousness an activity of the brain?

If neurobiology is right, then, it seems, we are left with a trilemma: either consciousness does not exist, *or* exists and is just this: the activity, or part of the activity, of a well-balanced intricate network of reactive dispositions installed in the brain, answering in a differentiated life-preserving manner to a huge number of incoming complexes of stimuli, *or* consciousness exists but is of no service to the conscious organism. As is well known, Daniel Dennett is happy to embrace either the first or the second horn of this trilemma (Dennett 1991), whereas David Chalmers is often misunderstood – though not without his own doing – as being content with its third horn.[1]

Though Descartes may have been wrong about many things, perhaps even about his own existence as a *res cogitans*, he was dead right about one thing, which we can also gather from his writings: that there is nothing more rationally certain than the existence of consciousness. I cannot here defend this view if it be thought to stand in need of defending. I hope, therefore, that everyone will agree that denying the existence of consciousness is not a viable option. Consciousness exists, and not only in human beings: it is rather common throughout the animal kingdom. To believe otherwise, to restrict consciousness to human beings only, seems to me a rather uncharitable position.

Embracing the first horn of the trilemma – denying the existence of consciousness – is out of the question. And there is nothing attractive to theory in embracing the third horn: in accepting the existence of consciousness and holding at the same time that it is of no service to conscious organisms. I have already addressed this option and what can be held against it.

We, therefore, seem to be left with the second horn of the trilemma: the identification of consciousness with the activity of a network of reactive dispositions in the brain, in accordance with incoming stimuli, for the benefit of the organism.

But mind–body dualists will hold that consciousness – though associated, in a manner not yet well understood, with cerebral activity, with the exercise

of the physical, electrochemical powers of a physical organ – *is not* identical with that activity or with anything physical. But how can this be more than a statement of blind belief? If neurobiology shows that a network of reactive dispositions, installed in the brain and answering in a differentiated manner to incoming stimuli, serves the very same purposes that consciousness is said to serve, must we not conclude that consciousness *is* simply the activity, or part of the activity, of that dispositional network – given that both the nonexistence and the epiphenomenality of consciousness are out of the question? The mere appeal to intuitions – however fervently upheld – seems insufficient to distinguish non-epiphenomenally-existing consciousness from brain activity.

But, for one thing, even if cerebral activity were the functional equivalent of consciousness, cerebral activity *might not be able to exist* without consciousness – consciousness being, nonetheless, different from it. This is a way of reconciling the *non-epiphenomenal* existence of consciousness with mind–body dualism, on the one hand, and the purported findings of neurobiology, on the other. If nonphysical consciousness were, on nomological grounds, a necessary condition of the cerebral activity which is its functional equivalent, then it could hardly be said to be an epiphenomenon in the sense that is ontologically negative and a reason for philosophical dissatisfaction.

The crucial question, however, is whether a certain network of reactive dispositions in the brain that answers in a differentiated manner to incoming stimuli does indeed serve the very same purposes that consciousness serves. In order to answer this question *negatively*, we need not deny neurobiological findings, we merely need to deny a certain interpretation of these findings. The findings of neurobiology point in the direction of the following conclusion, though they are still far from having conclusively established it:

For every conscious event A, there is a brain event B, such that everything that causes A also causes B, and vice versa, and such that everything that is caused by A is also caused by B, and vice versa.

The relationship between conscious events (events in consciousness) and certain brain events that can be gathered from the preceding thesis is their *causal equivalence*. Thus, the findings of neurobiology point towards the causal equivalence of conscious events and certain brain events, although, as I said, neurobiology is far from having established even that much.

A certain network of reactive dispositions in the brain that answers in a differentiated manner to incoming stimuli can therefore be said to be the *causal equivalent* of consciousness. But we are not forced to conclude from this fact (if it is a fact) that the dispositional network and consciousness serve exactly the same purposes, that they are *functional equivalents*. I am aware that many thinkers identify functional equivalence with causal equivalence, functional role with causal role; but this identification is a mistake, because sometimes there is a functional difference, in a clear sense,

on top of a causal equivalence. And so it is in the case of conscious events and the brain events that are their causal equivalents.

Every conscious event *intrinsically signifies something to someone*; in the overwhelming majority of cases, conscious events *usefully* intrinsically signify something to the subject of consciousness concerned, and hence function to the advantage of that subject. In contrast, no brain event intrinsically signifies anything to anyone. Therefore: although conscious events and certain brain events are – presumably – causal equivalents, they are not functional equivalents.

In this perspective, it is quite clear that the trilemma adduced above is a false one. We can fully accept the findings of neurobiology *and* retain our belief in the existence of consciousness that is *useful* to conscious organisms, and nevertheless there is no need for us to identify consciousness with some existing activity of the brain. In order to see things this way, one merely needs to avoid interpreting scientific data in a metaphysically biased way. These data accord some justification to the assertion that consciousness and a certain type of brain activity are causally equivalent; they accord no justification to the assertion that they are functionally equivalent.

Consciousness as intrinsic signification, and indeterminism

The next questions that must concern us here are the following: (1) how can the invoked concept of *intrinsic signification* be elucidated?; and (2) what is the significance of the fact that conscious events have functional roles that are different from the functional roles of the brain events that are their causal equivalents?

Regarding question (1): an event is intrinsically significant if, and only if, by and in itself it provides immediate information – that is, information which involves neither causation nor translation – to exactly one of its own constituents. Thus a pain event, for example, is intrinsically significant, since by and in itself it provides immediate information to exactly one of its own constituents: to the subject of the pain. Indeed, as I said, every conscious event – whether without an intentional object (as a pain event) or with one (as an event of visual perception) – is an intrinsically significant event, and it also seems to be true that every intrinsically significant event is a conscious event.

There is a further question: what is the nature of the intrinsic addressee of the immediate information provided by an intrinsically significant event? One candidate for the holder of this role that comes to mind is the conscious organism with which the intrinsically significant event is associated. But the organism *is not* a constituent of an intrinsically significant event associated with it (though sometimes it is an intentional object of such an event), and therefore it cannot be the intrinsic addressee of the immediate information provided by that event. A headache that I have at some time – a more than merely unpleasant sensation – is an intrinsically significant

event that is associated with this conscious organism, with my living body; but my body is not a constituent of that event. The true intrinsic addressee of the immediate information that my headache provides by and in itself to exactly one of its own constituents is not my body or any part of it, not even my brain. I am myself that addressee.

My headache has a certain causal equivalent, an electrochemical event in my brain. But the latter event is not intrinsically significant. If it were, it would have to have an intrinsic addressee – one of its own constituents – to whom it provided immediate information; but it has no such constituent. Therefore, the electrochemical event in my brain, though a causal equivalent, is not a functional equivalent of my headache.

My headache is intrinsically significant to me; the corresponding electrochemical event in my brain, though a causal equivalent of the former event, is not intrinsically significant to me. What is the point of this extra function my headache has? This brings us to the second question formulated above: the question of the significance of the fact that conscious events have functional roles that are different from the functional roles of the brain events that are their causal equivalents. What is the point of their intrinsically signifying something to someone, while brain events do not intrinsically signify anything to anyone?

Regarding question (2): one possible answer to this question is to say that there simply is no point to the fact mentioned in it. But this is not a plausible answer. My headache, indeed, may have its extra function uselessly, but this is certainly not true of every pain event.

We get to the heart of the matter if we ask ourselves what would be the point of there being intrinsically significant events – conscious events – if determinism were true. By *determinism*, I mean the doctrine that the laws of nature alone are sufficient to determine the entire history of the world if a complete initial segment of that history is given.[2]

Under determinism, information – immediate or not – cannot be *action-relevant* to anyone, for the simple reason that under determinism there cannot be any actions, where by an *action* I mean the exclusion by an agent at a certain time t of at least one nomologically possible continuation of the history of the world after time t. Clearly, if determinism were true, then no such excluding could be done at any time t, because, under determinism, at every time t there is just one nomologically possible continuation of the history of the world after time t (and that single nomologically possible continuation cannot be excluded because it cannot but be the *actual* continuation of the history of the world after time t).

Thus, if determinism were true, there would be no point in there being intrinsically significant events, no point in there being events which provide immediate information to exactly one of their own constituents. The existence of such events would be utterly otiose – a fairly bad joke of nature. Why? Because intrinsically significant events, conscious events, are evidently geared to providing information that, usually, is maximally action-relevant to an

agent – whereas under determinism there could be no actions and only agents that cannot act. Unless nature has done a very large thing – namely, the bringing forth of widespread consciousness – utterly in vain, determinism must be false.

The function of intrinsic signification that a conscious event has, and that the brain event which is its causal equivalent has not, is to give the agent, which is intrinsic to the conscious event, in the most immediate manner possible information on which to base its actions. That agent – for example, I – is in the service of a certain living organism (which, in its turn, is in the service of the agent); it is nothing other than *the soul of that organism.*

The biological soul both as subject of consciousness and agent

The soul of the organism is the subject of consciousness which is implicit in the conscious events that are associated with the organism, the entity to which the information provided by them is immediately and intrinsically addressed. Normally, the information provided by conscious events fits more or less tightly the task of the soul that is to use this information (i.e., that has evolved to use this information) in acting for the survival and the well-being of the organism of which it is the soul. However, the fit between conscious information and its (so to speak) evolution-intended use is much less tight in the case of modern human souls – because their ancestors have managed to secure, in the course of thousands of years, an environment that, normally, is rather depleted of dangers for human beings and, on the other hand, full of easily accessible resources for them. This historical matter of fact is responsible for the liberty (though not by itself for the capacity) that modern human souls have to pursue interests which can be broadly described as *cultural*. But certainly the generation of culture is, from the evolutionary perspective, only a secondary field of consciousness – as is the generation of pure (i.e., nonfunctional) joy, which plausibly can already be found at the subhuman level[3] – and a secondary task for the souls of organisms. The primary field of consciousness and the primary task for the souls of organisms is survival.

Consciousness and freedom

I have argued that consciousness would be out of place in a deterministic world, since the use of consciousness is to help secure the survival of a living organism by providing its soul, whose appearance in time is an outcome of the evolutionary process, in the right manner with information of the right kind – information on which the soul can base its actions. In a deterministic world there would be no actions, and while consciousness in a deterministic world would still have its function of intrinsic signification, its having that function while its cerebral causal equivalent is lacking it would be a fact

that, contrary to appearance, is without any significance and therefore a fact that is utterly misleading from the metaphysical point of view. It is hard to believe that nature might play such a trick on us (let alone God).

An action is, qua action, a *free* action in the sense that the initial segment of the history of the world that is prior to it does not determine it (on the basis of the laws of nature); otherwise, the nomologically possible continuations of the history of the world that are excluded by it would already have been excluded by the initial segment of the history of the world that is prior to it. Therefore, although some conscious events solicit actions – for example, the pain that ensues upon touching a very hot object – no conscious event determines an action. Hence it is a mistake to assume that actions are (sufficiently) caused by conscious events. If one wants to say that actions are caused by *something*, then one must say that they are caused by the agent, by the soul of the organism. The information that a conscious event provides to that agent is, therefore, nondeterminative; it leaves the ultimate decision what to do with it up to the agent (but certainly the agent-soul is not always able to use the information provided to it beneficially).

That every action is free in the sense just described does not yet mean that its agent *had a choice about it*: that there was an alternative possible action open to the agent at the time. But unless there is some inscrutable determination at work on top of nomological determination, it follows that every action is such that its agent had a choice about it.

The installation of an agent, acting in favor of and through its organism on the basis of immediate nondeterminative information provided to it in conscious events of which it is the subject, that is the installation of a *soul* on top of all the batteries of automatic reaction mechanisms an organism possesses has proved to be a rather successful invention of evolution. One decisive factor of that success is of course that most things that are of vital importance to the organism are not effected by its agent-soul at all, but precisely by the organism's automatic mechanisms. The agent-soul is there for the less common contingencies, and it is usually separated from most other things that vitally concern the organism by not being provided in consciousness with immediate information about them.

Within these limits, however, within the limits set by its state of information and its range of choices (the extent of which range is directly proportional to the richness of its state of information), the power of the agent-soul – especially of the human soul – can be very great, even to the extent of transcending the interests of its organism. This is strikingly illustrated by an old story which German pupils learning Latin in the 1960s and 1970s could still read in their textbooks, but which, presumably, is too awfully heroic for the taste of the present time. I am speaking about the story of Mucius Scaevola. Mucius Scaevola, when captured in the attempt to assassinate King Porsenna who was laying siege to Rome, held his right hand into the fire and allowed it to be consumed by it, thereby dissuading Porsenna from further laying siege to Rome, convincing him that it was full of

hundreds of Mucius Scaevolas fearing neither death nor pain in defending their nation.

Imagine the pain, imagine the soul that withstood it. The story is probably a legend; but comparable things have really happened, as we all know.

The insect objection

It is time to consider the serious objections that can be raised against the views on consciousness I advocate in this paper. One objection is this: insects are conscious animals. They, for example, experience colors. But at the same time they are automata that blindly follow the programs that are activated in them in reaction to outward or inward stimuli. Hence the proposed link between consciousness and freedom of action does not exist.

I respond that the objector is overly impressed by reports on insects that, if encountering some objectively insignificant anomaly in the process of achieving their preset goals, go through their preset rigid routines to achieve these goals an indefinite number of times (as often as one makes them encounter the very same anomaly). These reports are true, of course. But of course they do not show that the entire life of insects consists in rigid routines and reflexes.

If this were the case, if an insect never ever had a choice about anything in any situation of its life, then there would be no point in its being conscious. A set of non-consciously operating mechanisms triggered by non-consciously received stimuli would be quite enough to steer it for a while through the dangers to the resources of the part of the world that is its environment. But while nature is sometimes prodigal, it usually is not, and consciousness is too widespread a phenomenon, even in the kingdom of insects, to be a superfluous excrescence of evolution. This points us to the assumption that even an insect sometimes has a choice, a small choice undoubtedly, and a small soul that makes the choice, while being at the same time the subject of the insect's small consciousness.

There cannot be much deliberation going on when an insect makes a choice, certainly. But, in the first place, the presence of deliberation is not a necessary condition of making choices (since even *we* make choices – and rational ones – without deliberation, and such choices are far too often the right choices for them to be the products of a mere chance generator); and, in the second place, a rudimentary form of deliberation – consisting simply in the naked presentation of alternative possibilities – may well be present even when an insect (its soul) makes a choice. (Even insects seem to be capable of perplexity and bewilderment; if they are indeed capable of these states, rudimentary deliberation should also not be beyond them.)

What is indeed crucial for the making of choices is the presence of a unitary subject of both consciousness and agency which has at least a rudimentary consciousness of itself (and of its 'realm' – the organism – within its environment: a sense of being in the world). But it is sufficient for

rudimentary self-consciousness if there are, for example, pain events associated with the organism: there cannot be a pain of any subject of consciousness (and every pain is a pain of some subject of consciousness) without being in its consciousness *its* pain.

Should biology discover that insects are in fact in every situation and in every respect deterministic automata, then we should reconsider the question whether they are indeed conscious; then we should seriously consider the conclusion that they are not conscious at all (even though they have sensory organs and nervous systems that are remote analogs of ours). Why, for example, should an insect feel pain – and hence have a subject of consciousness (which, properly speaking, feels the pain) – if there is never ever a situation in which the insect – or more properly speaking the insect's agent-soul, which is identical to its subject of consciousness – can effectively decide to do something *or other* about it? For avoiding that a particular damage to the body becomes worse than it is, the insect does not need to feel pain, *if* it is always the case – in any such situation of bodily damage – that there is at most one way of evasion open to it; it does not need, *then*, a subject of consciousness which will *act* as it thinks fit (perform an action in the above-defined sense) on the basis of pain-information and other immediate information provided to it. Likewise, if an insect were a deterministic automaton, why should an insect feel fear or desire or pleasure? There is no point at all, *then*, to its having these emotions – or to its being in any other conscious state.

If an animal is in every situation of its life an automaton that reacts in a deterministic manner to the given combination of inner and outer conditions, then there is no evolutionary advantage whatever in the installation of the *consciousness-agency-apparatus*, having at its center the agent-soul. It is, admittedly, not a logical impossibility that a deterministic automaton is conscious, and it may so have happened that some conscious living beings are deterministic automata. After all, evolution has sometimes produced rather freakish beings. But it is highly unlikely, in view of the considerations that I have offered, that a creature is a deterministic automaton if it is in fact conscious.

The physics-teaches-us objection

Here is another objection to the views on consciousness I advocate in this paper. How could they be true? Doesn't physics teach us (1) that the physical conservation laws are true; (2) that determinism is as good as true (to a very high degree of approximation) in the mesocosmos where conscious beings live; and (3) that every physical event has a physical event as its sufficient cause, if it has any sufficient cause at all?

I respond as follows: Since the agent-soul serves its organism by selecting, in the light of immediate information provided to it in consciousness, from among nomologically possible continuations of the past history of the world

(i.e., from continuations *Y* such that in *the past* + *Y* all the regularities which are the actual laws of nature are preserved), the physical conservation laws are not violated by the activities of the consciousness-agency-apparatus. This takes care of (1).

Concerning (3), which is a principle of physical causal closure and can well be called *simpliciter* 'the Principle of Physical Causal Closure', I would like to point out that it is not something that physics teaches us or could teach us. Rather, it is one of the dogmas of physicalistic metaphysics. Curiously, it is advanced by physicalists as a strong argument in favor of their position. But the correctness of that position was not in question for the physicalists all along; what they are really doing in advancing the Principle of Physical Causal Closure is merely asserting a fairly obvious logical consequence of their own world-view – a world-view that is quite indefeasible and non-negotiable for them.

The metaphysical nature of the Principle of Physical Causal Closure emerges rather strikingly when we consider that the majority of physicists presently believe that some physical events have no sufficient physical cause, the reason for this being ultimately that they have not found any plausible sufficient physical causes for these events even after the most diligent search. Suppose now that it is really true that some physical events do not have any sufficient physical cause. Then – leaving agnosticism aside – we have a metaphysical choice:

(a) We can assume that all of these physical events that have no sufficient physical cause have no sufficient cause at all.
(b) We can assume that all of these physical events that have no sufficient physical cause have – each of them – a *nonphysical* sufficient cause (where I leave it open whether 'nonphysical' means as much as 'entirely nonphysical' or as much as 'not entirely physical').
(c) We can assume that some of these physical events that have no sufficient physical cause have a nonphysical sufficient cause, and that some of them have no sufficient cause at all.

There is *no* – I repeat, *no* – evidence from physics for either (a) or (b) or (c); physics, as the science of physical entities, is entirely neutral between them. The question whether we should adopt (a), or (b), or (c) is a purely *meta*-physical question, a question strictly 'following upon' physics, and no less so if the question is considered and answered by physicists. If we adopt (a), then we can stick to the Principle of Physical Causal Closure, but must deny the Principle of Sufficient Cause, which says that every event has a sufficient cause. If we adopt (b), then we can stick to the Principle of Sufficient Cause, but must deny the Principle of Physical Causal Closure. If we adopt (c), then we must deny both the Principle of Physical Causal Closure and the Principle of Sufficient Cause. Leaving agnosticism aside, *what*, in reason, should we do?

Choosing (c), and therefore the denial of both the Principle of Physical Causal Closure and the Principle of Sufficient Cause, is certainly the rationally least attractive metaphysical option. But there is *nothing* that makes the choice of (a) rationally preferable to the choice of (b); for the Principle of Sufficient Cause, which can be retained if (b) is chosen, is at least as metaphysically attractive as the Principle of Physical Causal Closure, which can be retained if (a) is chosen.

So why should mind–body dualists be impressed if physicalists advance the Principle of Physical Causal Closure against them, claiming for it the authority of physics? It does not in fact fall under that authority, and, from the metaphysical point of view, we are certainly not unreasonable if we consider it *false*.

I have now taken care of (1) and (3) of the above three objections against the views on consciousness I advocate, which objections, taken together, one might term the 'but-physics-teaches-us objection'. There yet remains objection number (2).

Though physicalists are unwilling to deny what the majority of modern physicists believe in – that indeterminism is prevalent in the microworld – physicalists, for understandable reasons, nevertheless maintain that in the mesocosmos determinism rules. They do admit that its rule in the mesocosmos is not guaranteed to be absolute and exceptionless, as was believed in the nineteenth century; but for all practical purposes, physicalists maintain, the rule of determinism in the mesocosmos can be *assumed* to be absolute and exceptionless.

But this is an assumption of physicalism, it is not something that physics teaches us. If it seems to me that I just now freely lifted my right hand, upon deciding to do so, then physics does certainly not teach that this event is, except for a tiny margin of contrary probability, necessitated on the basis of the laws of nature by the complete initial segment of the history of the world that is previous to it. How could physics teach any such thing?

Nor does physics teach another consequence of mesocosmic determinism, namely, that at any point in time before life evolved on this planet the entire history of the human species, which is replete with terrible crimes, was already a more or less inescapable consequence. All compatibilist attempts to reconcile freedom and determinism seem to me just so many attempts to obfuscate the horrible absurdity of such a view of human history.

But physics is entirely innocent of such ideas. The reason for this is simple: it is not a claim of the science of physics that the laws of nature discovered by it are in principle sufficient to explain everything that happens in the world on the basis of initial conditions. The completeness of physics (in the sense exhibited in the preceding sentence) is not a claim of physics, but a claim of physicalistic metaphysics *about physics*. As such, the completeness of physics is a matter of nonscientific, philosophical belief.

Since physics leaves me a choice, I, as metaphysician, rather choose to believe something else; namely, that also in the mesocosmos determinism is false

and not even approximately true. Believing this is all the easier for me in view of the fact that such belief opens up the possibility of giving a satisfactory account of the positive function of consciousness, of what consciousness – consciousness that is *not reduced to* something it is not – is good for in an evolutionary perspective. In a nutshell: consciousness is advantageous from an evolutionary point of view, in the manner I have described; but it can be so only if determinism in the mesocosmos is not even approximately true.

The transcendental objection to physicalism

One may well wonder what makes the metaphysical positions of determinism and physical causal closure so attractive to so many. The explanation I am going to suggest for this phenomenon of the history of ideas will bring me to the issue of the meaning of consciousness, regarding which I promised to offer some speculations at the beginning of this paper.

Physics is a theoretical system that arises out of human consciousness as an attempt – a rather successful one – to make systematic sense of our experiences of the physical world. As such, physics is an interpretation of a region of intentional consciousness, a region shared by the consciousnesses of many. But it so happens in the minds of not a few people that they lose sight of the soil out of which the tree of physics has grown; perhaps they are blinded by its spectacular growth and by the many good, or at least impressive, fruits that it has, in growing, brought forth. For these people, the total intentional object of the region of intentional consciousness that physics is concerned with, the physical world, turns into something that is metaphysically absolute – shown, they believe, to be such by physics itself. The physical world is thought by them to be *everything*, and in consequence physics becomes contaminated in their minds by a massive incursion of metaphysics. Determinism and the Principle of Physical Causal Closure (or some stronger principle than this) are assumed without much hesitation, since they are thought to arise out of physics itself and to be required for its very well-being – principles which, if believed in, make it impossible to understand what physics really is, and also what consciousness really is, as is amply illustrated by the modern philosophy of mind.

The epistemological pathology just described – which lies at the heart of physicalistic naturalism – was pointed out, in effect, as early as Kant's *Critique of Pure Reason* and as late as Husserl's *The Crisis of European Sciences and Transcendental Phenomenology*, and by many other authors, who were inspired by the tradition of German Transcendental Philosophy. Analytic Philosophers whose native tongue is English have largely ignored this tradition, one of the many reasons for this being that they dislike the epistemological idealism that is more or less explicitly advocated by all Transcendental Philosophers. But one does not have to become an epistemological idealist in order to accept the epistemological criticism of physicalistic naturalism that is implicit in Transcendental Philosophy.

Deplorably, the rich notion of consciousness that goes with Transcendental Philosophy (including Transcendental Phenomenology) and the earlier idealistic philosophy – Berkeley's and Hume's idealism foremost – is all but forgotten in the Analytic Philosophy of mind that is prevalent today in the English-speaking world. It is a much needed corrective for this type of philosophy to take cognizance of the fact that there is a notion of consciousness in the history of philosophy according to which some philosophers – e.g., Berkeley, Hume, Kant, Husserl – have believed that consciousness contains (as a construction remaining entirely within its bounds) the entire (knowable) world.[4] To this rich notion of consciousness,[5] I, too, would like to pledge my allegiance – though I am not an idealist. I am, of course, not supposing that the same richness of consciousness can be found at every level of evolutionary development. But I would indeed maintain that certain aspects of consciousness extend all the way down in the ladder of conscious life: the presence (in conscious events) of a subject of consciousness, the presence of phenomenal qualia, the presence of intentionality and hence of intentional objects (though presumably very crude ones in the lower forms of conscious life).[6]

The meaning of consciousness

If we reject the idealistic idea that, in a sense, consciousness is *everything*, what, then, is the meaning of consciousness? Martin Buber beautifully expressed the distance which any attempt to answer this question must straddle: 'The conscious mind [in German: *der Geist*] appears in time as a product, even as a by-product of nature, but nonetheless it is precisely the conscious mind that timelessly envelops her' (Buber 1983: 32, my translation). In this paper, I have offered a sketch of that part of the meaning of consciousness that is given by the fact that consciousness arises as a product of nature[7] (and consciousness can only *seem* to be a mere by-product of nature). But this *natural* (biological) meaning is *only a part* of the entire meaning of consciousness. The other part is given by the astonishing fact that this product of nature, which comes into being at some point in time, seemingly by accident, and maintains itself in existence because it is advantageous in the struggle for survival, nevertheless reveals to us human beings the timeless constitution of nature in her totality. How can this be? We have the two parts of the meaning of consciousness in our hands; what we do not know yet is how they fit together. If we knew how they fitted together, then we would fully comprehend what the meaning of consciousness is.

I do not think that physicalistic naturalism can find a satisfactory answer to the question of how the survival-function and the *theoria-function of consciousness* (as I call it) fit together. The theoria-function of consciousness, and the universal moral consciousness that joins that function and cannot be found without it, certainly cannot be explained as an optimization, brought

about by environmental pressure, of the survival-function of consciousness. Humanity would be the ruler of this Earth even if it had never left the level of conscious intelligence that *homo habilis* had. It was, of course, cultural evolution that initiated the theoria-function of consciousness and brought it to its present height. But what initiated cultural evolution?

Nothing less than a divine spark of enlightenment, I submit. At a certain point in time, humans – they were already *survivors* and in this sense *capax naturae* – became by divine grace *capax Dei*. They became able (in principle) to know God to the extent that He chooses to reveal Himself, and able (in principle) to be like Him to the point of being images of Him as creator. But since the totality of nature – all creation – is the larger part of God's self-revelation and the prototype of His doings, humans became at the same time also able (in principle) to know nature in her totality and to transform her morally responsibly in the light of that knowledge. They, who were already *capax naturae*, became not only *capax Dei* but also *capax naturae secundum imaginem Dei*.

Notes

1 See Chalmers (1996). Chalmers merely upholds the *logical* epiphenomenality of consciousness, not the *nomological* epiphenomenality. But he sometimes speaks as if logical epiphenomenality were epiphenomenality *simpliciter*.
2 This initial segment will have a first moment if the history of the world has a first moment, or be of infinite duration into the past if the history of the world is of infinite duration into the past.
3 I am grateful to Alvin Plantinga for having drawn my attention to this.
4 Concerning Husserl's criticism of naturalism and his comprehensive notion of consciousness, see Meixner (2003).
5 If it had not become common these days to associate with the term 'phenomenal consciousness' the impoverished sense of *purely qualitative consciousness*, it would not be amiss to call the notion of consciousness I adhere to 'phenomenal consciousness'. In order to understand this term in the sense in which I would agree to use it for my conception of consciousness, one must understand it in the way Husserl – the originator of Phenomenology – would have understood it, i.e., in such a way that phenomenality does not preclude either abstractness or structure.
6 I am grateful to Josef Quitterer for his comments on this paper, which made clear to me the need to say more about my concept of consciousness.
7 More on this subject can be found in Meixner (2004).

Which consciousness do we need to have a choice?

Comment on Meixner's paper

Josef Quitterer

Different kinds of consciousness

I think everybody agrees that there is no agreement about what 'consciousness' is. Consciousness is an ambiguous concept which – like most other concepts of common sense psychology – comprises a bundle of meanings. This is what Ned Block probably has in mind when he says that 'consciousness' is a 'hybrid' or a 'mongrel concept' (1995: 227). In his attempt to bring some order into the jungle of meanings of consciousness, Block distinguishes between four basic categories of consciousness: *Phenomenal consciousness*, *access-consciousness*, *monitoring consciousness* and *self-consciousness*. This is not the place to defend or criticize Block's position concerning the different forms of consciousness. What is important here are the different types of phenomena which can be subsumed under the term 'consciousness'. David Chalmers distinguishes between the so called 'hard problems' of phenomenal consciousness and the 'easy problems' of 'awareness' which include forms of consciousness such as the reportability of mental states, the deliberate control of behaviour, etc. (1995: 200–202). Norton Nelkin distinguishes between 'Nagel-consciousness', 'proposition-like awareness' and 'apperception' (1996: 147).

One could enlarge the list of philosophers with other classifications of different meanings of the term 'consciousness'. What can be derived from all of them is a general tendency to classify the different phenomena of consciousness according to their complexity or their position in the hierarchy of mental representations. Following this tendency, we can distinguish three different types of conscious activities:

- phenomenal consciousness (what it is like-consciousness, Nagel-consciousness, hard problems of consciousness, qualia, etc.);
- awareness (monitoring consciousness, proposition-like awareness);
- self-consciousness, self awareness (self-modelling, self-representation).

It might be helpful to apply this classification to Meixner's concept of consciousness in order to better understand the point of his argument. Meixner

does not give an explicit definition of the meaning of the term 'consciousness' which is used in the argument. But if we take a look at the examples in the article, we can get a rough idea of the type of consciousness which is used in Meixner's argument. In his paper he mentions 'pain' (Meixner, p. 186) or 'headache' (Meixner, p. 186). Consciousness is also seen by Meixner as providing 'immediate information ... , information which involves neither causation nor translation' (Meixner, p. 186) or as being an 'intrinsically significant event' (Meixner, p. 186). Meixner describes consciousness in terms of phenomenal consciousness the way it is described by Chalmers and others as non-relational, non-functional, non-representational. The author also presupposes that consciousness is 'rather common throughout the animal kingdom' (Meixner, p. 184). This presupposition makes sense only if the concept 'consciousness' means 'phenomenal consciousness'. Moreover, I take it for granted that Meixner does not claim that animals have reflective consciousness or self-consciousness.

Phenomenal consciousness provides non-determinative information

Now let us turn to Meixner's argument: according to him, being conscious would be meaningless if determinism were true. Meixner establishes a direct link between consciousness and non-deterministic agency. Events of phenomenal consciousness are seen by Meixner as immediate non-determinative information on the basis of which the agent decides to do something (Meixner, p. 186). According to him, the 'proper function of consciousness' (its functional role) is to provide immediate information to the subject (the soul), which is 'maximally action-relevant to an agent' (Meixner, p. 187). A non-relational form of consciousness (intrinsic, immediate, etc.) is seen as the basis for deliberate and free action. The question which I would like to raise in my comment is whether phenomenal consciousness is the best candidate for action-relevant consciousness.

Meixner's argument can be divided into a weaker and a stronger thesis. In the weaker part of his argument, Meixner does not claim that phenomenal consciousness is a sufficient condition for a free decision to act in a certain way. In this part of the argument, he is only saying that phenomenal consciousness provides action-relevant information to the agent in such a way that it does not determine the agent to act in a certain way: 'The information that a conscious event provides to that agent is ... nondeterminative; it leaves the ultimate decision what to do with it up to the agent' (Meixner, p. 189).

In this sense phenomenal consciousness would be only a necessary condition for free agency and it is obvious that in order to perform a free choice on the basis of non-deterministic information provided by (phenomenal) consciousness, the agent must have further cognitive abilities: she must be able to (a) reflect upon the non-deterministic phenomenal information; and (b) form a decision – to intend something and to act according to the

intention. That means that the existence of phenomenal consciousness in itself is not enough to have a choice. In order to have a choice, the subject must be able to use the non-deterministic information in a non-deterministic way – she must be able to pick out the relevant information, evaluate and connect it to her beliefs, desires, preferences, and then form an intention to act.

In this weaker part of the argument we could still assume the existence of individuals who are phenomenally conscious but never had a choice because they do not have the other cognitive abilities required for a free decision: they cannot reflect upon the non-deterministic information, they are not able to form intentions, etc. There is quite a lot of biological evidence that these individuals exist. In phylogenetic and ontogenetic stages of cognitive development there are organisms, whose behaviour compels us to assume that they are phenomenally conscious but that they do not have the ability to metarepresent the phenomenal information provided by conscious experience and to use it for a choice between alternative ways of acting.

It is, however, precisely this conclusion which Meixner would not accept. Here begins the stronger part of his argument. In his reply to the 'insect objection', he claims that 'if an insect never ever had a choice about anything in any situation of its life, then there would be no point in its being conscious ... This points us to the assumption that even an insect sometimes has a choice' (Meixner, p. 190). According to Meixner, it is impossible that evolution produces an individual which is phenomenally conscious of specific information without having a choice concerning the possible reactions to this information. 'Why ... should an insect feel pain – and have a subject of consciousness – if there is never ever a situation in which the insect ... can effectively decide to do something or other about it?' (Meixner, p. 191).

In this stronger part of the argument Meixner seems to assume that phenomenal consciousness – the form of consciousness, which we also find in lower animals, probably even in insects – is a sufficient condition for a free choice. Wherever there is phenomenal consciousness like perception of pain, there also would be the ability for a free choice. Phenomenal consciousness alone would guarantee free choice.

That means that phenomenal consciousness must include in some way the other cognitive aspects required for a free decision-making process. Meixner seems to admit this in the revised version of his paper when he says that 'certain aspects of consciousness extend all the way down in the ladder of conscious life' (Meixner, p. 195). A closer look at the cognitive abilities required in a decision-making process shows that this would be a very problematic thesis.

Objections to the insect objection

The main problem with phenomenal consciousness being self-sufficient in a decision-making process is what Norton Nelkin subsumes under the

'filling-metaphor of phenomenal consciousness' (1996: 129): When a person A is at t_1 in a state Y of phenomenal consciousness (of pain, for example), it is impossible that A is at t_1 in a state X of phenomenal consciousness of the state Y. An event of phenomenal consciousness cannot be represented mentally by another conscious event of phenomenal consciousness, which is 'intrinsically significant', that means, which contains nothing else but phenomenal information. There is no way to come to a choice on the basis of phenomenal information without leaving the domain of purely phenomenal consciousness. In order to meta-represent a state Y of phenomenal consciousness, we need a state W of non-phenomenal consciousness such as a second-order state of a normal decision-making process like a belief, a desire or an intention to act: for example, I know myself to be in pain, therefore I intend to avoid further pains. A choice or decision is made on the basis of second-order representations of conscious events.

Now somebody could object that even these second-order mental representations are forms of phenomenal consciousness because there is a 'what it is like' for a person to be in these mental states; one could easily connect phenomenal qualities with my desire to have a pasta carbonara this evening, it is harder to connect it with beliefs. What would be the phenomenal content of my belief that Milan is 460 km away from Innsbruck? But even if we admit that mental states like desires and beliefs have phenomenal quality, this quality alone does not determine their role in the decision-making process. In a decision-making process where we go through different desires and beliefs it is not the phenomenal quality that enables us to distinguish and evaluate the action-relevant mental states. We distinguish thoughts, beliefs and desires on the basis of their propositional content; a phenomenal quality accompanies our thoughts and beliefs only in a very contingent way (Nelkin 1996: 143).

If phenomenally conscious events play a role in a decision-making process and present – as Meixner suggests – non-deterministic information to the agent, they can play this role only if they are somehow represented by other mental events whose content is not exhaustively determined by their phenomenal quality. As we have seen, precisely those states whose information to the agent coincides with their phenomenal quality are not able to meta-represent other states with a different phenomenal quality.

In the revised version of his paper, Meixner reacts to this critique and pleads for a richer notion of 'phenomenal consciousness'. According to Meixner, even the most primitive form of conscious life must have the following attributes: 'the presence ... of a subject of consciousness, the presence of phenomenal qualia, the presence of intentionality and hence of intentional objects (though presumably very crude ones in the lower forms of conscious life)' (Meixner, p. 195). Even when we assume that phenomenally conscious events like pains, colour-perceptions etc. are intentional mental states with a subject of consciousness and an intentional object, this structure is not enough to guarantee free choice. Let us assume a bee which

is phenomenally conscious of a red flower. In this case, we have an intentional mental state (a colour perception of a red flower), a subject (the perceiving bee) and an intentional object (the red in the flower), and an immediate non-determinative phenomenal information ('red'). Does the colour-perception of the bee in itself imply that the bee can choose between different alternatives how to react to this phenomenal information? Even if the colour-perception provides non-determinative information, the bee would not be able to use it for a choice. In order to have a choice, the bee would need second-order mental states which are transparent for the non-determinative information provided by the first-order mental states of phenomenal consciousness. The bee would need a form of consciousness, whose intentional objects are conscious events themselves; the bee would need a form of monitoring-, introspective- or self-consciousness.

In the same part of his paper, Meixner refers to 'the rich notion of consciousness that goes with Transcendental Philosophy'. According to Meixner, the concept 'phenomenal consciousness' should be interpreted in the terms of Transcendental Philosophy and the earlier idealistic philosophy: 'there is a notion of consciousness in the history of philosophy according to which some philosophers ... have believed that consciousness contains ... the entire (knowable) world. To this rich notion of consciousness, I, too, would like to pledge my allegiance' (Meixner, p. 195). In my view, this richer notion of consciousness does not provide any advantage to Meixner's argument. In the philosophical positions in which consciousness contains the entire knowable world, being conscious does not imply having a choice. Let us take Kant as an example. According to Kant, the entire knowable world is – as Meixner suggests – a world of conscious experience. But this does not imply that this form of consciousness provides non-determinative information for an agent – on the contrary: when consciousness comprises the entire knowable world, it comprises also the causal determinations of the physical world. Thus, according to Kant, conscious experience has to be integrated – according to the forms of intuition and the categories of understanding – into a fully determined manifold of phenomenal appearance. There is no room for free choice on this level of conscious experience, because it contains the causal and deterministic structure of the physical world. As a consequence, free choice is a possibility only outside the world of conscious experience – it is a matter of practical reason beyond the limits of possible (conscious) experience. Therefore, we must assume the existence of conscious beings, which do not have any choice – those who cannot transcend the bounds of conscious phenomenal experience in practical reasoning.

In this case, Meixner would raise the following question: what is then the evolutionary advantage of being conscious? If free choice is not an immediate consequence of being conscious, I would have to develop an alternative evolutionary role for phenomenal consciousness. There are such alternatives. One example is the hypothesis according to which phenomenal consciousness

is required 'to flag the present'. According to Nicholas Humphrey, complex cognitive systems need phenomenal consciousness because they have to compute a lot of different types of information. The cognitive system has to combine the evidence of 'sensory stimulation with contextual information, memory and rules so as to construct a hypothetical model of the external world'. The danger is that, 'if this kind of construction is allowed simply to run free, without being continually tied into present-tense reality, the perceiver may become lost in a world of hypotheticals and counterfactuals'. What the perceiving subject needs is 'the capacity to run some kind of on-line reality check, testing his perceptual model for his currency and relevance'. According to Humphrey, this is the point 'where phenomenal consciousness does prove its value'. Phenomenal consciousness gives 'a here-ness and now-ness and a me-ness to the experience of the world', which pure perception in the absence of phenomenal consciousness does not have (Humphrey 2000: 242–243).

Conclusion

The fact that the action-relevant information is consciously presented to the agent is in itself not sufficient to make a choice between different alternative ways of acting. The agent must display other cognitive faculties in order to perform a choice on the basis of the information provided by phenomenal consciousness: the agent needs the ability to reflect upon the information presented to her in this primary form of consciousness – she needs some form of second-order-consciousness.

Part V

Practical philosophy

9 Naturalism, realism, and objectivity in ethics

Robert N. Audi

Naturalism has become almost an orthodoxy, at least in Anglo-American philosophy. It has long been a force in ethics, reaching its pinnacle in the British tradition extending from Hume to John Stuart Mill; but in the twentieth century it has been developed in virtually every field of philosophy. My aim here is to provide a partial assessment of its success in ethics and the wider theory of practical reason.

Naturalism in ethics

There are many conceptions of naturalism, and I cannot even begin to survey them here. It may be useful, however, to take as a guiding idea (as I have in previous work (Audi 2000)) that naturalism is roughly the view that nature is all there is and the only basic truths are truths of nature. This leaves open what constitutes nature and how we are to determine what truths are basic. It is common for philosophical naturalists to answer these questions on the basis of an interpretation of what sorts of entities scientific inquiry must countenance. These philosophers seek to avoid countenancing the existence of any entities whose existence need not be posited in order to make sense of scientific theorizing. All philosophical naturalists believe (controversially, of course) that this standard excludes countenancing supernatural entities; virtually all of them believe that it rules out positing irreducibly normative properties; most of them tend to think that it limits reliance on abstract entities; and many of them tend to think that it favors, even if it does not require, empiricism.

To see how naturalism is pursued in ethics, we should first consider a great divide in ethical theory. On the historically dominant side is *cognitivism*, roughly the view that moral judgments are truth-valued and hence appropriate objects of such cognitive attitudes as belief and knowledge. On the other side is non-cognitivism: the position (which arguably traces at least as far back as Hume's *Treatise*) that moral judgments are non-cognitive. Non-cognitivists differ markedly in their positive views. Some are prescriptivists, emphasizing the directive character of moral judgment; others are expressivists, emphasizing the way in which moral judgments express

attitudes; and there are various combinations of expressive, prescriptive, and other elements commonly taken to be characteristic of moral judgment.

I do not mean to imply that a non-cognitivist must be an ethical naturalist. Indeed, it is misleading to call non-cognitivists ethical naturalists, given the history of that term. In G.E. Moore's famous critique of ethical naturalism, for instance, that position is the view that goodness is a natural property, such as the property of maximizing pleasure (and by implication this holds for other ethically central properties). This naturalist position, however, is cognitivist. It would be better to compare non-cognitivism in ethics to a kind of eliminativism in the philosophy of mind. Each denies the existence of properties – as opposed to linguistic predicates – of the kind taken to be central in its domain. The difference is this: non-cognitivism (in its plausible forms) is a friendly eliminativism, denying that there are moral properties at all, but taking moral judgments to be capable of both justification and an enduring role in human life, whereas eliminativism in the philosophy of mind, even if it acquiesces to our using mental predicates until we have a better descriptive vocabulary, affirms the possibility of explanations of action and experience that rely only on broadly physical properties.

I suggest, then, that we view non-cognitivism in ethics as in part motivated by philosophical naturalism. It retains moral and other normative predicates but eliminates moral properties. Letting language be as pluralistic as is convenient, it keeps ontology pure.

Non-cognitivism faces serious challenges. One is a dilemma: if the rationality conditions for moral judgments do not presuppose the possibility of evidences of their truth or falsity – or of their objective soundness in a sense difficult to understand without taking them to be cognitive – then a strong relativism would seem to follow. We may *persuade* one another of a moral view, but never *show* that it is true or (it may be plausibly argued) objectively rational. If, on the other hand, evidences of truth or falsity are presupposed by the rationality conditions for moral judgments, it would seem that the judgments are cognitive after all.

A related challenge is the Frege–Geach problem.[1] This concerns the way in which moral judgments can apparently function like truth-valued elements. Consider the schema:

War entails killing innocent people;
If war entails killing innocent people, war is wrong; hence
War is wrong.

This seems valid as it stands. But if it is, its constituents, including the moral judgment that war is wrong, must presumably be taken to have truth value, and the moral element (statement, if you like) in the consequent of the conditional premise must function in a way that seems to presuppose its having truth value. There are attempts to account for the validity status of

the argument without this presupposition, but few would deny that such examples are a challenge to non-cognitivism.

This is not the place to pursue possible logical maneuvers to deal with the Frege–Geach problem. My purpose here is to bring out that there is considerable cost to maintaining ethical non-cognitivism. It may suit an overall naturalism; but I think it is fair to say that it is not philosophically natural.

Let us consider whether a cognitivist naturalism fares better. There are at least two major cases here. First, there are *concept-reductive* views. These embody a naturalistic analysis of moral concepts. Second, there are *property-reductive* views. I take it that if moral concepts are reducible to natural ones, so are moral properties. Take Mill's position in *Utilitarianism* as a possible example of both kinds of reduction. Suppose that the concept of an obligatory act is equivalent to that of an act that maximizes the proportion of pleasure to pain. Then, the property of being obligatory is surely equivalent to (and arguably identical with) that of maximizing that proportion. The converse, however, seems false: property equivalence does not entail conceptual equivalence. Surely the concept of a circle is not equivalent to that of a plane closed curve whose circumference is equal to its diameter times pi, even if the two properties in question are identical (or, assuming this is a different notion, provably equivalent).

What hope is there for a cognitivist naturalism? Before I pursue a plausible form of this view, I want to set aside what may seem to be a version of the position. I refer to the counterpart of what, in the philosophy of mind, is sometimes called non-reductive materialism. This is the view that mental properties supervene on, but are not reducible to, physical ones. Given any plausible conception of supervenience, this is a property dualism which simply provides a naturalistic anchor for mental properties. Perhaps those who have used the term 'non-reductive naturalism' have not taken to heart the point that even the arch non-naturalist, G.E. Moore, took (the property of) goodness to supervene on natural properties, as did W.D. Ross (though Ross, at least, preferred the term 'consequential'). I agree with them on this, and I suggest that 'consequential' is a clearer term to use for the kind of dependence relation that moral and other ethical properties bear to natural ones.

It is, then, reductive naturalism that we should consider in assessing the prospects for a cognitivist naturalism. Here I shall simply consider what I take to be the most promising version to be proposed so far. I refer to the naturalism implicit in what is sometimes called Cornell realism.[2]

In its perhaps most modest form this naturalism rests on the thesis that only natural properties can figure essentially in subsumptively explaining events – such as actions. Cornell realists argue that moral properties can do this and that they are therefore natural. For instance, we may explain an uprising by appeal to the injustice of the government. The injustice might involve police brutality, seizures of property, and arbitrary curfews. Injustice

seems not only to play an explanatory role here; it might also be considered, in a sense, *causal*. In that case, it would clearly be different from any mathematical properties that might be thought essential in explaining events. This causalist realism, then, can be distinguished from a non-naturalist realism that posits moral properties but denies that they are 'natural'.

In earlier work, I have shown how we might grant that such moral explanations as the one just cited can be naturalized (and indeed construed as causal), yet deny that moral *properties* can be naturalized.[3] One crucial point is that the moral attribution central in the explanation – say, the attribution of governmental injustice – presupposes the presence of natural properties on which this moral property is consequential. These base properties include such things as police brutality, seizures of land, press censorship, and arbitrary curfews. But if they do include such elements, then the way is open to hold that the *empirical* explanatory force of the attribution of injustice – its capacity to explain actual events – derives from the ascription of the natural properties in question. I consider this a more plausible interpretation of the data than the property-reductive naturalistic account. In short, we can naturalize 'moral explanations' without naturalizing moral properties.

Suppose my interpretation is more plausible. It does not follow that moral properties cannot be naturalized in some other way. I want to consider two other possibilities, one proceeding via a naturalization of goodness, the other via a naturalization of practical reason itself.

Long ago Plato, and particularly Aristotle, showed that the kind of goodness central in ethics is essentially connected with what it is natural to call human flourishing. Might we reduce the relevant property of being good either to a kind of psychological property or to a kind of functional one – say, that of contributing to the harmonious functioning of a certain kind of human society? Mill and other hedonists have attempted the former; the latter might be pursued along lines indicated by Cornell realism.[4] Let us consider these two strategies in turn.

For reasons brought out by Moore and others, hedonism, conceived as a kind of reductive naturalism, does not succeed, even if pleasure is *among* the good things we should countenance in ethics. Richard Brandt proposed to reconstruct the notion of the good and other normative notions; his proposal is to substitute, for normative notions, naturalistic notions which, though not equivalent to the normative notions, can do all the work we can reasonably ask of them.[5] What is good for us, on his view, would be roughly equivalent to what we want or would want given certain purifying conditions, such as elimination of logical and conceptual error, exposure to relevant facts, and a sound grasp of our own desires – corrected, if necessary, by 'cognitive psychotherapy', as he called the relevant critical procedure. One problem with this as a naturalization strategy is that, in the effort to avoid non-naturalistic assumptions, Brandt relied on a causal conception of the relevance of a fact to judging the rationality of a desire. But the notion of

relevance does not seem causal. For instance, if knowing that a contemplated deed would kill millions would not influence someone made insane by brain manipulation, this does not imply that it is irrelevant to the morality or even the rationality of the deed. On the other hand, the fact that, say, getting cosmetic surgery would make one sound like a certain popular actor might causally influence an agent, yet might not be relevant to the morality or even the rationality of the would-be transformative deed. Brandt's project also encounters other difficulties (which I have discussed in some detail elsewhere).[6]

Let us suppose, however, that the property of being (intrinsically) good could be naturalized. Even if pleasure is not the only good, there are other natural properties, such as engaging a person's intrinsic interest, that may seem to constitute kinds of goodness (this is not the same property as relieving boredom, though that, too, on a different ground, is a candidate for a kind of goodness). One might think, then, that goodness might be a disjunctive natural property or kind of complex natural property. Assume that it is. Even if so, it is far from clear that rightness is a kind of goodness. Indeed, there are apparently deontological requirements on action such that certain kinds of actions, for instance, abstaining from killing, are by their very nature *prima facie* wrong.[7] Thus, even if, as I doubt, a naturalistic reduction on the axiological side of ethics will succeed, this would not imply that reduction will succeed on the deontological side. I doubt that it can.

I turn now to the possibility of a naturalization of practical reason itself. The idea here would be to provide a general naturalistic account of reasons for action and then proceed to show how moral reasons, and hence moral judgments, can be understood in terms of it. The best prospect might be a reconstructed Humean view, combined with the assumption that the right is a subcase of the practically rational. Brandt's view may be considered to be this kind of position, but it will help to put the project in terms that do not depend on his particular version of it. For one thing, there are many versions of a broadly Humean instrumentalism.[8]

Hume held that the function of reason (at least *practical reason* – though he did not to my knowledge use this term) is to serve desire. This is highly vague, and in plausible versions there will be a restriction: on the conative side, only non-instrumental desires count as foundational for rational action, and, on the cognitive side, only rational beliefs count as instrumentally central for rational action. In short, rational actions will be roughly those that optimally serve one's non-instrumental desires in the light of one's rational beliefs. The crucial beliefs will be those concerning the probabilities that one's actions or possible actions will satisfy the relevant desires.[9] Here I will make just three points about this project.

First, as a naturalization project, it is not self-contained. It requires a naturalization of the notion of rational belief. I doubt that this can be accomplished. I am not claiming that no naturalistic epistemology can be sustained; I grant that a good case can be made for naturalizing the notion of

knowledge. For paradigmatically normative notions like rationality and justification, however, the prospect of full naturalization seems dim (as I have argued elsewhere).[10]

Second, not just any non-instrumental desire is plausibly thought to provide a ground for a practical reason. Such a desire can be, for instance, self-destructive. It can also be directed toward something that it is plainly not rational to want for its own sake, such as being burned. Such desires can be implanted, even if they are difficult to take as real possibilities otherwise. How can a naturalist rule them out as unsuitable bases for action? It is true that they are not 'natural', but what is the relevant notion of the natural? A statistical notion will not do. Even an evolutionary one will not adequately serve. For one thing, survival of the species is possible at various levels; mere survival need neither be good nor imply that whatever conduces to it is good.

My third point is a special case of the second, but cannot be seen to be so if one makes the common mistake of taking the distinction between intrinsic and instrumental desires to exhaust the territory of conation. Intrinsic desire is commonly contrasted with instrumental desire, and the latter is often taken to be the only kind of extrinsic (non-intrinsic) desire. It is not the only kind. Intrinsic desire (the kind Hume had in mind in speaking of passions), is desire for something *for its own sake*, where this implies wanting it on account of something taken to be intrinsic to it, such as the refreshing qualities of water wanted on a hot day. But a desire can be merely non-instrumental: *neither* intrinsic nor instrumental, as where a desire that is utterly isolated from one's interests is induced by posthypnotic suggestion and one has no idea why one has it. I could be caused by brain manipulation to have a merely non-instrumental desire to visit the North Pole. This desire would not give me a reason for action.

If we keep in mind that there are two kinds of non-instrumental desire, we can see a surprising consequence for instrumentalism. At least for pure instrumentalism, it makes no difference whether a desire is intrinsic or *merely non-instrumental*. Suppose one forgets why one wanted something, such as to go to the hardware store, but enters the store. The action is a response to the desire and may occur even when one has forgotten, but does not realize one has forgotten, why one set out to go there, yet also has no (intrinsic) desire to go there for its own sake. Since one has forgotten this and has no new motivating desire, the action, although plainly in some *historical* way desire-based, is not a candidate to be rational as a means to any further end. Still, *that* one desires something other than as a means (non-instrumentally) is all that matters. Such a desire – call it a *residual desire* – can have a clearly attainable object, and it defines a goal for action as clearly as does an intrinsic desire. We can act to realize a residual or other merely non-instrumental desire, get what we want in so acting, and do so efficiently or inefficiently in the same sense that applies to instrumental action in general.

This apparent commitment of instrumentalism to the reason-giving power of merely non-instrumental desires is entirely consistent with the related *psychological* point that *believing* one does not know why one wants to enter the hardware store can *eliminate* the desire to do so, especially if one realizes that one originally wanted to do it *only* as a means. Normatively, the desire is, for instrumentalism, still not irrational: it is indeed capable of providing a perfectly good *reason* for action, say to open the door.

A consequence of conceding reason-giving power to a merely non-instrumental desire is this. Even though this desire is not intrinsic and (typically) would be easily given up upon realizing that one has no idea why one wants its object, wanting to do something purely as a means to realizing this object would still provide a reason for this action. The reason would be claimed to derive from the 'grounding' of this instrumental desire in the non-instrumental desire to which it is subordinate.[11] Is there, however, any (normative) reason to perform the second action, say, to ask someone to delay a meeting so that one can go to the hardware store? Doing this might be reasonable on the basis of wanting to recall what one wanted, but *not* in order to fulfill the merely non-instrumental desire to go to that store.[12] I conclude that the problem of residual desires is a serious challenge to a naturalization of practical reason along instrumentalist lines. Mere non-instrumental desire is not a ground for action.

Moral realism

If what I have said so far is sound, then ethical naturalism is far from established and indeed faces very serious difficulties. Do we have a plausible alternative theory? I believe we do. This section will outline it (a full-scale statement is provided in Audi 2004).

It might seem that if cognitivism in ethics is sound, as I suggest it is, we should also be moral realists. The possibility of cognitivism without realism, however, has been defended (or in any case presupposed) by Kantian constructivists. We would do well to consider this line. If it succeeds, there may be more room for non-naturalistic moral theories.

As I understand it, Kantian constructivism is in outline the view that moral principles are, in the basic cases, constructed rather than discovered. The categorical imperative, then, lays down a method for determining what our procedures should be; contrary to an impression not unnatural in reading the *Groundwork,* it does not provide a master axiom whose truth is knowable *a priori* and from which we may deduce principles to guide everyday life. One way to put this is to say that it guides us in a way that leads to a 'reflective endorsement' of principles constitutive of our practical identity.[13]

This view deserves much more discussion that I can provide here. But any appraisal must surely explore what *constraints* operate in any reflections that may be plausibly claimed to yield true moral principles. In a way, the

question is similar to one confronting Brandt's qualified instrumentalism: what besides logic, conceptual clarity, and relevant facts, is presupposed in determining a correct application of the constructive procedure? Kant himself apparently thought that we could eliminate both mistaken maxims governing particular actions and the false principles they instantiate by appeal to a contradiction test. Taken by itself, this test has failed to convince even loyal Kantians. But Kant also used his intrinsic end version of the categorical imperative, which requires that we treat all persons never merely as means but always as ends in themselves. Moral philosophers and other readers have found this plausible. But this, I take it, has both normative and descriptive content. Above all, treating people merely as means is a kind of *preponderantly* instrumental treatment – of which paradigms are exploitive deeds – and treating them as ends is a case of doing something toward them that is at least in part motivated by a concern with their good for its own sake. If this is correct, then there apparently are normative properties that play a role in constraining the basic use of the categorical imperative. It does not *create* moral truths without depending on some basic moral, or at least normative, assumptions.

If, however, there are normative constraints on the application of the categorical imperative – or on any adequate procedure for constructing moral principles – the question arises whether its proper use does not after all imply a commitment to moral realism. The property of being right, for instance, might be that property an act has in virtue of its accord with a principle that emerges from a correct application of the categorical imperative, constrained by such things as a prohibition on treating persons merely as means. This is a *filtered moral realism* as opposed to the more *direct moral realism* implicit in treating some lower-order act-types such as promise-keeping and lying as *prima facie* right or wrong. But it still appears to be a version of realism.

None of these points implies that there is not an epistemic role for Kantian construction. The kind of reflection one may do under the inspiration of the categorical imperative may both refine the principles one already holds and lead to discovery of new ones. There is much room here for a quest for reflective equilibrium among one's intuitions about cases, one's ordinary working principles – such as those governing veracity, fidelity, and justice – and the categorical imperative itself. But without constraints of a normatively substantive kind, the procedure provides neither a plausible account of moral justification nor an adequate way to avoid the unacceptable relativism that looms up when there are insufficient objective constraints on the grounding of moral judgment.

I believe, then, that constructivism, if it is indeed an anti-realist view, does not provide a satisfactory alternative to a realist ethical theory. I shall assume that the best ethical theories available will be both cognitivist on the semantic side and realist on the ontological side. What about the epistemological side?

It might seem that for a philosophical naturalist, an empiricist moral epistemology is the only satisfactory kind. That is true for a naturalist who takes all truths to be truths of nature and considers the *a priori* non-natural. It should also hold for a scientific naturalist who want to be an ethical cognitivist and thinks that all truths are in principle confirmable using scientific method. But one could be a naturalist about moral properties and still take some of the truths in which they figure to be *a priori*. I shall, however, leave this possible position – which I doubt is occupied by any major theorist in ethics – aside. My own preference is a moderately rationalist moral epistemology, and I now turn to that.

Major rationalist moral epistemologists include Aquinas, probably Thomas Reid, Kant, and of course Moore and Ross. I do not take the view to depend on non-naturalism in moral ontology; but non-naturalism is apparently the position of at least Moore and Ross, and I think it is preferable for a plausible rationalist epistemology. My own view in this territory is an integration of intuitionism with Kantian ethics and with a connected account of intrinsic value.[14] I hold that the sorts of principles of *prima facie* duty that Ross formulated are not empirical. That there is a *prima facie* duty, hence some moral reason, not to lie, break promises, or injure seems to me knowable *a priori*. This does not entail that it cannot be known empirically, that it is obvious, that our justification for it is indefeasible, or that it cannot be supported by more general principles, such as the categorical imperative. Moral rationalism does not have to embrace stereotypes of the self-evident that have bedeviled the understanding of ethical intuitionism at least since Moore.[15]

Moreover, the view I have outlined does *not* imply that singular moral judgments, such as that I should do a particular charitable act, are *a priori*. These are not only existential and empirical; they are also knowable only in the light of at least a presupposition that no other duty overrides the one in question. This is not the place to present my view in detail. I simply want to bring it forward as a moderately rationalist, realist, pluralistic moral theory that does not require or embrace naturalism.

Ethical objectivity

If we do not take basic moral principles to be empirical, and if we grant that moral judgments are knowable only in the light of a plurality of potentially conflicting moral considerations, have we undermined the objectivity of moral judgment? The question is particularly important in the context of appraising naturalism in ethical theory. For suppose moral truths are natural. Then they may be plausibly taken to be, as utilitarians generally suppose, confirmable by common-sense observation and scientific procedures. They would thus be objective in basically the same way as scientific judgments. The point is particularly important in connection with ethical naturalism. That position is in good part motivated by a desire to

achieve objectivity in deciding moral questions, much as naturalism in general is in good part motivated by a desire to achieve objectivity in dealing with intellectual questions.

The question here is not *ontic* objectivity, the kind that goes with the existence of a fact of the matter as to whether an act does or does not have a moral property. This kind of objectivity (at least bypassing certain problems about vagueness) apparently comes with any plausible metaphysical realism. The important question here is *epistemic objectivity*. In broad terms, this occurs in ethics provided: (1) there is an intersubjectively usable method for answering moral questions; (2) using it tends to lead to agreement among adequately rational persons who conscientiously use it in the light of the same empirical data; and (3) beliefs grounded in this agreement tend to be justified and, if true, to constitute knowledge.

In meeting criteria of epistemic objectivity, utilitarian empiricism in ethics may seem to have an advantage over rationalist moral theories. We can surely tell, using ordinary observation and scientific method, what kinds of acts tend to produce pleasure and what kinds tend to produce pain, and on the same basis we can make quantitative comparisons. Whether an act is obligatory, then, is knowable, at least in principle. Still, as critics of utilitarianism have noted, it is not easy in practice to determine such information. Mill specifically addressed this question in answering the objection that in practice moral agents lack time to calculate the effects of their options:

> mankind have been learning by experience the tendencies of actions . . . [and] Nobody argues that the art of navigation is not founded on astronomy, because sailors cannot wait to calculate the Nautical Almanac . . . only in cases of conflict between secondary principles is it requisite that first principles should be appealed to.
>
> (Mill 1979, Chapter 2)

By contrast, intuitionists argue that the same kinds of principles – such 'secondary principles' as that lying is wrong and promise-keeping is obligatory – that Mill must show to be derivable from the principle of utility are instead *a priori*.[16] This intuitionist effort has met with many objections. One is that it opens the way to dogmatism in ethics. The suggestion is that anyone who does not see the truth of an *a priori* principle is either deficient in understanding or in any case clearly wrong-headed. But a moderate rationalism takes even *a priori* justification to be in general defeasible; it also distinguishes the self-evident from the obvious, and it grants that moral principles – particularly given the complexity of the *prima facie* qualification – need not be obvious.

Even if there are *a priori*, intuitively knowable moral principles, then, no licence for dogmatism is implied. Moreover, any plausible moral theory will grant that singular moral judgments, say that given a medical emergency, I must break a particular promise, are not self-evident and may require the same kinds

of factual evidences that plausible theories of any other kind would call for. One kind of factual evidence concerns what would happen to an accident victim if I did not render aid rather than press on to keep a prior promise.

To take a different example, it may be self-evident that killing people is *prima facie* wrong, but it is not self-evident that one ought not to kill someone who is attacking one with a lethal weapon and can be deterred only by taking deadly defensive action. Such examples raise a problem. Perhaps this moral proposition could not be known non-inferentially, even though the proposition might be 'intuitive' as judged in the light of the facts of the case. In responding to the dogmatism objection in this way, then – allowing that an intuitive judgment may depend on factual information and may be false – have we undermined the authority of intuition in ethics?

We have not: fallibility does not imply unreliability; and the epistemic dependence of singular moral judgments on factual information does not entail a like dependence of our justification for moral *principles*. In any case, the facts crucial for justification of singular moral judgments are often evident at a glance. This is one reason why, on the kind of intuitionist view I am sketching, non-inferential, intuitive knowledge of singular moral propositions is provided for in a huge range of everyday cases. Normally, I should simply keep my promises, tell the truth, avoid injuring people, and so forth. The related singular moral judgments (those favoring such conduct) can even be anchored – as an epistemological naturalist would wish – in a reliable connection between the non-moral facts on which moral properties are consequential and the belief that the act has that property, as where, on the basis of seeing someone cheat on an exam, we believe that an injustice has been done. If the cheating entails injustice, and our belief that injustice has been done is reliably grounded in the perception of the cheating, then that belief is ultimately based on the facts in virtue of which it is true. So it is with knowledge of natural facts as well. Reliability and indeed objectivity here come with grounding of the belief in the facts; neither reliability in the judgmental process nor the objectivity of the beliefs it generates requires that the cognizer's concepts all be natural.

Non-inferential, intuitive justification, then, is both commonly attainable and defeasible in the ways I have sketched. It can also be defended with premises: the fact that it arises in a quasi-perceptual, non-inferential way does not preclude defending it – or correcting it – by argument. Dogmatism is not a natural result of embracing the kind of intuitionist view I have sketched. This is not to deny that, for any ethics, dogmatism is, for at least some people in some cases, a liability. Conviction can easily lead to stubbornness; the desire to do the right and uproot the wrong can produce a zeal that oversteps its proper bounds. A good moral theory reduces these tendencies and, where they occur, provides for their identification and guides their elimination.

To put the main epistemological point here a bit differently, the naturalistic anchoring of *singular* moral judgments that goes with taking moral

properties to be consequential on non-moral ones must not be taken to undermine the view that there is genuinely moral knowledge. There is a sense in which the kind of ethics I have been outlining is autonomous: it enables us to acquire *general* moral knowledge on the basis of natural reason. For a rationalist moral epistemology of the kind in question, the epistemic autonomy of ethics is quite strong.

An important implication of the kind of ethical theory I have outlined should be emphasized here. Since this kind of ethical theory is not naturalistic, it is compatible not only with the existence of non-natural properties but also with theism. Indeed, it can be integrated with theism in a way that has important implications.

To see this, consider first the point that if basic moral truths are necessary, as they surely are if they are *a priori* in the way they seem to be, then they are known by God. If they are divinely known and, in at least that sense, divinely accepted, then God knows the truth of such principles as that lying and infidelity and killing are *prima facie* wrong. If we further assume, that God could not know this without wishing us to act accordingly,[17] then there is a kind of *authority* of moral principles that cannot be accounted for on naturalistic lines. Accounting for the authority of moral principles is an important task of moral theory. There are various routes to doing this. The authority of reason is enough for many philosophers. But for many others and for many non-philosophers, it is not enough. The moderate rationalist view I have sketched leaves open both theological and secular routes to accounting for the authority of moral principles.

Naturalism in ethics remains a powerful thrust on the contemporary scene. But I cannot see that it succeeds. The way of non-cognitivism leaves many unsolved problems and forces a logically unnatural construction of moral discourse. The path through reduction seems to be blocked by insurmountable obstacles. The rationalist view provides an anchor in the natural world for moral discourse, but it allows us to transcend the merely empirical in framing our moral principles. It preserves both epistemic objectivity and metaphysical realism; and it provides a plurality of plausible moral principles to guide everyday life.

Acknowledgements

For a detailed and helpful commentary on this paper, I am grateful to Antonella Corradini. The paper has also benefited from discussions with other participants in the Milan Conference on Naturalism held in June, 2003.

Notes

1 The Frege–Geach problem has been widely discussed. For a good short treatment, see van Roojen (1996).

2 See, e.g., Sturgeon (1988) and Boyd (1988). See also Harman (1986).

3 This is argued in Audi (1993).
4 On this matter see especially Boyd (1988), and Bloomfield (2001).
5 A detailed defense of this naturalistic substitution – or reconstructionist – view is provided by Brandt (1979).
6 My detailed critical response to Brandt's naturalistic reconstructionism is given in Audi (1983).
7 I argue for this in Audi (2004).
8 See, e.g., Williams (1981) and, for a recent defense of an instrumentalist view, Hubin (2001).
9 This view of instrumental beliefs and an extended critique of instrumentalism in the theory of practical reason are provided in Audi (2002a).
10 My most recent defense of internalism is Audi (2002b).
11 The reason is, however, *entirely* derivative. Otherwise, even where one wants *A* purely as a means to *B*, one could be credited with two reasons to *A* – that one wants to *A* and that one wants to *B* – which could outweigh an intrinsic desire to *C* stronger than the desire to *A*. At that rate one could prefer swimming to boating, but since the strength of one's desire for the latter, together with the strength of one's desire to rent a boat, as a means to it, could be greater than the strength of one's desire to swim, one could have better reason to go boating. And this could be so even if one has forgotten why one wanted to go boating (say, to meet a friend across the lake) and would hardly know what to do with the boat having rented it. Here conative agitation unsettles me after all, even though I do not know its cause or how to achieve quiescence.
12 If this case of a merely non-instrumental desire is one in which an object of desire is not desired under any aspect of goodness, the case also suggests why the view that, normally, what is intrinsically wanted is wanted under the aspect of goodness is plausible.
13 For a well-developed version of the reflective endorsement view, see Korsgaard (1996, 1997). Among the instructive recent critical discussions is Shafer-Landau (2003), especially Chapter 2. In Korsgaard (1996), e.g., she says of Kant, apparently approvingly, 'Kant, like Hume and Williams, thinks that morality is grounded in human nature and that moral properties are *projections* of human dispositions' (ibid.: 91, my italics). Later she clarifies this by maintaining, 'The form of realism I am endorsing here is procedural rather than substantive realism: values are constructed by a procedure, the procedure of making laws for ourselves' (ibid.: 113).
14 I defend this view in detail in Audi (2004).
15 This is argued in detail in Chapter 2 of Audi (2004).
16 For a plausible consequentialist account of how we might account for secondary principles, see Hooker (2000).
17 I presuppose here a distinction (found in Aquinas) between God's antecedent and consequent will. I have discussed this in relation to ethical theory in Audi (forthcoming). See the view that

> Principles of moral obligation constituted by divine commands are not timeless truths, because the commands are given by signs that occur in time. People who are not in a region of space-time in which a sign can be known are not subject to it.
>
> Adams (2000: 270)

It appears that the constitution relation intended, as opposed to the one posited in what I call divine commandability theory, by Adams is incompatible with *a priori* status for the relevant moral principles. For a short recent treatment of divine command ethics, see the exchange between Janine Marie Idziak and (representing the natural law side) Craig A. Boyd and Raymond J. Van Arragon in Idziak *et al*. (2003).

Non-reductive naturalism versus non-naturalism in ethics: How wide is the gap?

Comment on Audi's paper

Antonella Corradini

Audi's paper examines various issues related to moral naturalism, ethical objectivity and moral realism. In what follows, I intend to comment on some of the aspects related to moral ontology and moral epistemology.

Moral ontology

In the field of moral ontology Audi has explored several perspectives ranging from non-cognitivism as a form of naturalism to non-naturalism. In order to clarify these positions, he has resorted to a parallelism between philosophy of mind and ethics which, in the former domain, involves the relationship between mental and neurophysiological properties while in the latter it regards the connections between moral and natural properties (Audi, p. 205).

Table 9.1 illustrates the relationship between the philosophy of mind and ethics. It approximately adopts Audi's classification at least in so far as non-cognitivism, reductive and non-reductive naturalism are concerned. However, what appears in both fields of non-naturalism is, to a large extent, my own responsibility. The aim of the comments set out here is to concentrate on the following aspects: what sort of features distinguish non-reductive naturalism from non-naturalism? And to what extent does the parallelism between philosophy of mind and ethics hold?

Table 9.1 Relationship between philosophy of mind and ethics

	Eliminativism	*Reductive naturalism*	*Non-reductive naturalism*	*Non-naturalism*
Philosophy of mind	Eliminative materialism	Identity theories	Nomic supervenience	Emergentism
Ethics	Non-cognitivism	Concept or property-reductive views	Nomic supervenience	Conceptual supervenience

Non-reductive naturalism in philosophy of mind and ethics

The notion of supervenience is crucial for non-reductive naturalism. Let us assume that A denotes a set of neurophysiological properties and B a set of mental properties. B will supervene on A if there are no two objects with exactly the same A-properties and that differ in their B-properties.

Supervenience implies dependence of the supervenient properties on the base-properties while simultaneously allowing for the former to have a certain autonomy. B-properties do not have to correspond to a fixed set of subvenient properties; non-reductive naturalism implies the thesis of the multiple realizability of mental properties and is therefore compatible with functionalism.

A similar situation arises if we consider the moral domain. Moral properties supervene on natural ones. They are therefore dependent on natural properties but are also multiply instantiated by them. In order to highlight the peculiarity of this form of moral naturalism, I would add that: (1) Moral properties do not necessarily supervene only on physical ones; they may also supervene on properties referred to by social or human sciences and such properties can (but do not have to) supervene on physical properties; and (2) moral properties may only weakly supervene on natural properties, in which case the supervenience relation is not necessary but contingent.

A further common feature of non-reductive naturalism in philosophy of mind and ethics is that supervenience is conceived in both domains as a nomological relation, that is, as an *empirical* relation. The empirical status of the supervenience relation renders both philosophy of mind and ethics amenable to scientific methods. Such an aspect could have, for example, the important consequence that a unitary form of realism is made possible together with scientific realism.

The entities referred to by science are presumed to explain why certain events occur. Does this also apply to moral properties? Is it possible to attribute explanatory force, or even causal efficacy, to moral properties? Non-reductive naturalists say it is. They maintain that the moral explanatory level cannot be reduced to the natural explanatory level without loss of important aspects of the explanation (Brink 1989: 195 ff.). Saying that an uprising is caused by police brutality or seizures of property is not the same as saying that it is caused by injustice. In this regard, the claimed explanatory autonomy of the moral level becomes an element in favour of a naturalistic moral realism because, in a scientific framework, moral properties may perform an explanatory role only if they are constituted by natural properties endowed with causal efficacy.

More recent supporters of non-reductive naturalism, however, observe that supervenience by itself is not sufficient to grant causal autonomy to the moral level. Supervenience is the source of bottom-up causality, but it cannot account for top-down causality and therefore it gives rise to moral epiphenomenalism. In order to grant causal efficacy to moral properties, we

must assume a further postulate: moral properties must be emergent properties (Rottschaefer 1998: 7.4).

In my opinion, this position raises two difficulties. First, the emergence of higher-level properties from lower-level properties is not really compatible with a naturalistic frame. If, for example, we say that the process of perceiving an object's red colour is emergent, we imply two things. First, that the perception of red is not caused by neural processes alone; second, that it can exert causal efficacy on the neural level. However, this entails that physics is explanatorily incomplete: there are both neurophysiological processes that cannot explain higher emergent levels and neurophysiological processes that cannot be explained by other neurophysiological processes. Emergence implies breach of the causal closure principle of the physical world. It therefore entails the abandonment of physicalism and its replacement with a form of non-naturalism.[1]

Second, as we have just seen, emergentism takes non-reductive naturalism too far making it a form of non-naturalism. Yet emergentism is not required by the moral case. Non-reductive naturalists resort to emergentism in order to give causal efficacy to moral properties; but they need it precisely because they share the naturalistic assumption that moral properties are constituted by natural properties. If we admit, as non-naturalists do, that moral properties are *sui generis* properties, we can give causal efficacy to moral properties without resorting to emergence. Within a non-naturalistic frame moral properties are causally efficacious as *intentional objects* of our knowledge.[2] Let us suppose that an agent acts in order to avert an injustice, such as the suffering one person causes to another without reason. The subject is moved to action by, among other things, her belief that the state of affairs is morally wrong. However, the cause of her belief is *not* the natural state of affairs itself but the morally negative *value* of that state of affairs as conceived by the subject (Corradini 2003a, 2003b). Therefore, moral properties are not required to have causal efficacy like protons in a cloud chamber. On the contrary, values are able to explain a belief as intentional objects only if they are non-natural properties. As Audi puts it: 'They do not have direct explanatory power ... but they might be said to have contributory explanatory power, in the sense that they can make an indispensable contribution to the direct explanatory power of propositional attitudes in which they figure' (1997: 122).

If we now return to my initial analogy between philosophy of mind and ethics we find that it is no longer valid within a non-naturalistic perspective. As already anticipated in the initial scheme, a non-naturalistic philosophy of mind presupposes the validity of emergentism, which non-naturalistic ethics does not require. Moral properties simply belong to another domain, where there is no place for emergence or for supervenience conceived as a nomological relation. As Audi writes, 'they do not belong to the causal order' despite the fact that 'they are anchored in it' (ibid.: 122).

At this point we may ask: what are the features of plausible non-naturalistic ethics?

Non-naturalism in ethics

First, we must set aside what I would call a prejudice about non-naturalism, namely that it implies the existence of two unrelated sets of properties, the moral and natural ones. This is often argued as a consequence of Moore's assertion that moral properties are *sui generis*. However, Moore did not intend this assertion to be taken as implying the existence of mysterious and evanescent properties, with no relation to the 'real' world. Moore warns against this understanding of his *Principia Ethica* by observing that:

> I should never have thought of suggesting that goodness was 'non-natural', unless I had supposed that it was 'derivative' in the sense that, whenever a thing is good ... its goodness ... 'depends on the presence of certain non ethical characteristics' possessed by the thing in question.
> (Moore 1942: 588)

It is thus interesting to find that the concept of supervenience was first conceived by the 'arch non-naturalist' G.E. Moore (1922), as Audi calls him, and that the naturalistic interpretation of supervenience is only derivative. We must now determine the form taken by a non-naturalistic supervenience relation between moral and natural properties.

In the ethical domain (strong) supervenience expresses a necessity different from the nomic one, that is, *conceptual* necessity. Attention has recently been redirected to this sort of necessity by David Chalmers (1996: Chapter 2), according to whom consciousness supervenes naturally, but not analytically, on the neurophysiological properties. We can conceive creatures like *zombies* who are neurophysiologically identical to conscious creatures but do not actually have any conscious experience. By resorting to conceptual necessity, Chalmers seeks to refute psycho-physical identity: if this holds, he argues, it should hold analytically and not merely empirically. It would be beyond the scope of this paper to go deeper into Chalmers' proposal or to assess its merits. However, I submit that an argumentative strategy based on the notion of emergence is more useful for anti-physicalistic purposes in the philosophy of mind than are appeals to conceptual necessity.

I return now to the moral case. Conceptual necessity differs from both purely logical and physical necessity. If we use the possible-world jargon,[3] we can state the following: an assertion is physically necessary if it holds in all possible worlds compatible with natural laws, that is, in all physically possible worlds. For example:

x is a watery stuff iff x is H_20

This is an empirical law, which holds in all naturally possible worlds. We can therefore argue that it does not hold in worlds which are logically but not physically possible. In these worlds, watery stuff, for example, is identical

with XYZ. The physically possible worlds are a subset of the logically possible worlds.

Let us take another example:

If x is a pleasure, then x is good.

This is a normative principle which holds in all conceptually possible worlds. These latter belong to a subset of the logically possible worlds, that is, the subset in which meaning-postulates hold, in particular the meaning-postulate that pleasure is good. Note that pleasure is not tautologically identified with the good; the postulate refers to two semantically autonomous meanings that can be identified independently of their relationship. Something similar happens in the case of the assertion: 'Human beings are rational animals'. This does not simply state the substitution conditions between the terms involved in the definition; rather, by going beyond the purely linguistic level, it gives us information about the relationship holding between their *denotata*. Conceptual necessity, as it is expressed by a relationship of this kind, is thus to be understood as a necessity of content and not as a merely formal necessity. It applies to relationships between states of affairs that neither exclusively depend on logical principles nor are tied to mere empirical aspects of objects. If the property of being good supervenes conceptually on the property of being pleasant, this principle holds also in the worlds that are not physically possible. In a world where watery stuff is XYZ, the principle that pleasure is good still holds.

Moral epistemology

The meaning-postulate expressing the conceptual relationship between natural and moral properties is *normative* in character. To grasp this relationship we need *moral knowledge*. This is not a mysterious 'moral intuition' nor is it the result of a special 'moral sense': we can grasp the moral properties of a state of affairs in the same manner as we know other sorts of conceptual relationships: for example, in the aesthetic or mathematical domains. Normative conceptual necessity is only one case among others in the wider field of conceptual necessity.

Moral knowledge is *a priori*, for even if the subvenient properties are empirical, the relation between natural and moral properties is *a priori*. Accordingly, a non-naturalist cannot but be a rationalistic moral epistemologist, to use Audi's terminology. However, Audi also maintains that although

> the sorts of principles of *prima facie* duty that Ross formulated are not empirical ... the view I have outlined does *not* imply that singular moral judgments ... are *a priori*. These are not only existential and empirical; they are also knowable only in the light of at least a presupposition that no other duty overrides the one in question.
>
> (Audi, p. 213)

I disagree with Audi on this point. The level of generality in no way influences the rationalistic character of moral knowledge. The claim that moral knowledge is *a priori* is true at the most abstract level of the first principles but it is also true at the most concrete level of the ultimate practical judgement. By proceeding from the abstract to the concrete, the moral decision-maker gathers more and more empirical information and formulates more and more accurate normative bridge-principles. To be sure, the particular situation gives decisive clues as to the *hic et nunc* overriding duty, but the choice of the latter is not just the result of mere empirical observation. It still involves an *a priori* moral judgement, whose epistemological status does not undergo changes because it is embedded in a concrete situation.[4]

My final comment concerns the alleged indefeasibility of *a priori* knowledge. Audi argues against philosophical common-places on this matter by denying that the *a priori* character of moral knowledge necessarily implies that 'our justification for it is indefeasible ... Moral rationalism does not have to embrace stereotypes of the self-evident that have bedeviled the understanding of ethical intuitionism at least since Moore' (Audi, p. 213). For this and other reasons, Audi's point of view can be labelled a 'moderately rationalist' moral epistemology (Audi, p. 213).

Now, if moral knowledge meets fallibilistic criteria of justification, one may ask whether coherentism is not a better epistemic alternative to intuitionism. If we deny the equivalence between *a priori* and indefeasible knowledge, we are obliged to admit that no self-justifying belief exists. A belief is always justified by other beliefs, pertaining to the background knowledge available to us. In order to avoid a vicious regress, we must claim that justification of second-order beliefs results from the mutual support of the totality of beliefs, and this corresponds to coherentistic criteria. The well-known objection that fairy tales are also coherent and consistent can be easily rebutted if we are ready to combine epistemic coherentism with ontological realism. If we are realists, we can maintain that second-order beliefs are realistic beliefs about the reliability of our beliefs about the world. As Brink puts it: 'Our realist second-order beliefs include beliefs about our psychological make-up, our cognitive and perceptual equipment, and their hookup to the world' (1989: 127).

In conclusion, a moderately rationalistic moral epistemology should perhaps dare to end its forced marriage with intuitionism and seek to draw new energies from a coherentistic perspective combined with moral realism.

Notes

1 I am aware that there are many views on emergence and that not all of them are incompatible with physicalism. This is not the place to explore this important and now much-discussed topic. My own allegiance is to an interpretation of emergence like that put forward by Crane (2001) where he explains why emergence is at odds with physicalism.

It goes without saying that other positions in the philosophy of mind are at odds with physicalism, such as Cartesian substance dualism, for example. But, once again, I agree with Crane that one merit of emergentism is that it represents a midway point between reduction and substance dualism.

2 As such, they perform the same role as descriptive, non-moral properties inasmuch as they are conceived as *abstracta*. See Lowe's and McCann's papers in this volume.

3 For more technical details, see Corradini (1995).

4 On the relevance of the distinction between the abstract and the concrete level of the normative domain, see Corradini (2003c: Section IV). Another point of minor disagreement between Audi and myself is whether from a rationalistic point of view basic moral principles must be synthetic. According to Audi, rationalists in moral epistemology are, *as such*, committed only to the existence of a priori basic moral knowledge; they have often argued for its syntheticity, but that is not entailed by their general rationalist commitments alone (1997: 96). I claim, on the contrary, that rationalists are, *as such*, committed to the synthetic *a priori* character of moral principles, or, as I would prefer to say, to their *material-analytical character* (see previous section).

10 Resisting naturalism

The case of free will

Hugh J. McCann

It is no easy matter to define philosophical naturalism, but I am going to take it in a very simple and straightforward way. I shall understand naturalism as the view that the problems of philosophy are to be addressed, and their solutions framed, in terms acceptable to natural science, particularly physical science. Naturalism has both a negative and a positive thrust. The negative part is that to the extent the concepts philosophers employ have no home in the language of physics – if they are irreducibly mentalistic, say, or value laden, or peculiarly religious – those concepts have no bearing on the world of ordinary experience, which (it is implied) is the only world of which we have knowledge. Right-thinking philosophers will therefore eschew them. The positive thesis is that the concepts of the physical sciences will be found to suffice: that is, that they can and ultimately will provide a thorough and sufficient understanding of all that we experience. I think it is fair to say that naturalism represents a kind of orthodoxy within analytic philosophy. But its origins are much older, and like most orthodoxies it has seldom come close to achieving universal acceptance. One thing that impedes its success is the free will problem, which has always resisted solution in terms congenial to naturalism, and continues to do so. The reason is essentially twofold: friends of naturalism have never succeeded in providing a deterministic solution to the problem, and an indeterministic or libertarian solution violates the tenets of naturalism. In this paper, I shall chart the case for these claims, and then say just a bit about what they imply about the prospects for analytical philosophy without naturalism.

Some history

Naturalism as we know it today has its roots in the scientific revolution and the philosophical movements that emerged in that era. These tended to portray the physical universe as a grand mechanical contraption, in which all that takes place does so in accordance with a few deterministic laws. This is, of course, a long way from our present conception of the physical world, and it leaves open some questions on which contemporary naturalists have strong views. Notoriously, it is silent on the relation of the mental to the

physical: about whether psychological phenomena are merely aspects or by-products of the material world, or instead make up a separate realm, and possess their own peculiar efficacy. The image of the universe as machine is mute about other issues as well, for example, about the origin of the universe, the development of life, and, at least to some extent, the status of values.

But there were some subjects on which the scientific revolution spoke with unmistakable clarity, and what it had to say left little or no room for free will. The chief of these subjects was what counts as an appropriate explanation. For the new sciences, the explanation of physical phenomena was to be mechanistic, not teleological. A good explanation would proceed not in terms of supposed purposes for which things existed, of ends they naturally sought, or of goods their behavior achieved, but in terms of diachronic relations of cause and effect. Things moved as they did because they were impelled to do so by prior causes, whose operation was presumed to be deterministic, precisely measurable, and completely describable in mathematically formulated laws. There was, of course, the question of how far those laws would be found to reach, of how much of our experience would prove describable as just a matter of bodies in motion. But, then as now, the success of science inevitably prompted efforts to assimilate as much as possible to physical or quasi-physical models. Human action was a legitimate target for such efforts, because paradigm cases of human action are at least partly physical: I cannot move my arm without the arm moving; I cannot fire a gun without the gun firing. But any attempt to deal with human action in a mechanistic way was bound to produce a clash over free will.

The primary reason for this was, of course, the issue of determinism itself. Free will requires that the formation and execution of intention be, as we say, 'up to the agent'. That is, it requires that human decisions and actions involve the exercise of a certain sort of power – a power that appears, at least, to be contracausal. When the cue ball strikes the object ball in a game of billiards, it is not up to the object ball how it moves, precisely because the conditions in which it is placed, together with the features of the collision, determine completely what is to happen next. Only if this were false could anything be up to the object ball. The same goes for human decision and action. If I am deciding where to go after this conference, the condition of my mind does guarantee some aspects of the outcome. I will not decide to go somewhere I have never thought of, or to a place I have no reason to visit. And because these things are settled by my mental condition, they are not up to me. But my mental state seems not to guarantee everything. If more than one option presents itself to me as having something to offer, then (most of us would say) it *is* up to me which one I choose. This, we would say, is part of what it means to claim I am free, and it suggests that I am very much different from a billiard ball. It suggests, at least, that the condition of my mind does not determine what I decide – that I might have had the very same motives and beliefs, and yet have decided differently.

There is a second point on which the mechanistic view conflicts with free will. When I decide or act freely, what I do can only be made sense of by an explanation that is explicitly psychological. If, for example, I decide to visit France after this conference, I will do so because such a visit makes sense to me, and the sense it makes is what will explain my decision and subsequent action. The real import of this point is obscured in the Cartesian setting, where it gets tangled up with the question of how events conceived to occur in an immaterial soul can have material consequences. In my view, that is not the main issue. Rather, as I shall argue below, ordinary explanations of action do not proceed in terms of events, even mental events. They proceed in terms of the content of mental events, in terms of envisioned outcomes, portrayed as desirable. They explain decision and action as undertaken by the agent in pursuit of an end deemed worth having. This runs counter to the mechanistic approach, first, because mental content is very difficult to treat plausibly in naturalistic terms. Second, this kind of explanation is explicitly teleological – the very thing the proponents of the new sciences found most objectionable in traditional approaches.

Philosophers of the seventeenth and eighteenth centuries had few tools for attempting to analyze the relation of the mental to the physical, and so made little in the way of a concerted effort to explain away the mental dimension of behavior, or to reduce it to processes that could be viewed as acceptably physical. There was, however, great respect for the idea of determinism, at least as applied to the physical realm – enough respect that among major thinkers of the era, few were prepared to defend free will in a form that would threaten it. A notable exception is Thomas Reid, who, rejecting the mechanistic picture entirely, boldly asserts that only minds have genuine causal power (Reid 1969: I, Chapter V), that in exercising that power they are undetermined (ibid.: IV, Chapter VI–VIII), and that the 'motives' that explain behavior are not events at all but, as I just suggested, *abstracta* (ibid.: IV, Chapter IV). If these claims are true, and if our actions in the physical world accord with our wills, then the physical realm cannot be expected always to behave deterministically, and the explanation for what occurs in it will not always be causal. The approach of Immanuel Kant, though more radical in some ways, is far less combative. For Kant, freedom pertains only to the noumenal realm: it is a necessary presupposition of deliberation and action, but strictly a practical postulate, not an item of knowledge in the theoretic sense (Kant 1998: A 538–541, B 566–569). On the contrary: to observation action will always appear deterministic, since in the phenomenal realm the law of causation is universal. The effect of this strategy is to affirm free will, but to sequester it from the world of mechanistic causation. We may in practical contexts take ourselves to be free; but empirical observation, which alone provides legitimate knowledge, must always claim the opposite.

But, of course, Reid and Kant were swimming against the tide. The most common response to the mechanistic picture of decision and action –

manifested prominently by Locke and Hume in Britain, and by Leibniz on the Continent – was to accept it, and to postulate a weakened view of freedom that would comport with determinism. Thus Locke speaks of liberty as a power to do or forebear a particular action according to the determination of the mind (Locke 1961: Book II, Chapter XXI, Section 8), Hume gives substantially the same characterization, calling it a 'hypothetical liberty' given which if we choose to remain at rest, we may, and if we choose to move, we also may (Hume 1955: Section VII, Part I). For both authors, 'liberty' is no more than freedom of action – that is, freedom to act as we choose, or in accordance with what we will. As for the will itself, it is determined by the agent's strongest motive, and neither philosopher accords any explicit freedom to the will as such; indeed, Locke is at pains to reject this whole idea, which he says leads to an infinite regress of willings to will (Locke 1961: Section 23). In fact, however, one can adapt the Lockean analysis of freedom to the will also, holding that decision and volition are free even though ruled by the agent's strongest motive, provided only that a dominant motive in the opposite direction would have caused an opposite result. This is in effect the strategy of Leibniz, who, though troubled by Locke's views on the will itself (Leibniz 1949: Book II, Chapter XXI, Section 8), finally takes an essentially similar stance, holding that the only necessity that would destroy moral responsibility must be an 'absolute' necessity, such that our actions would occur no matter whether we wish them or not. The conditional or 'hypothetical necessity' through which our motives determine our behavior is not of this kind, for in this case a different motive would cause a different outcome (Leibniz 1985: 381). Thus, Leibniz was able to assert, somewhat misleadingly, that motives incline the will without necessitating it (ibid.: 382). What this means, really, is that determining motives do not render our conduct *logically* necessary. And since they do not, Leibniz is prepared to consider us free. He would not deny, however, that our motives necessitate our wills hypothetically; indeed, the very thing Leibniz calls 'hypothetical necessity' is what Hume calls 'hypothetical liberty'.

Some more recent history

It was against this background that analytic philosophy concerning the will and action took shape. Early analytic philosophers were little influenced by European philosophy after Kant. But they had great respect for the empiricist tradition, and shared much of its motivation, including great respect for science; indeed, at least in the beginning, the prevailing orthodoxy was that empirical observation was our sole source of all but conceptual knowledge, and that since philosophers do not engage in systematic observation, they ought to confine their endeavors to the conceptual realm. And I think it is fair to say that for about half a century, it was largely so – first with the logical positivists' project of formulating an ideal empiricist

language for science, and later with the ordinary language movement, which sought to demonstrate that philosophical problems rested on confusion that could be alleviated through conceptual analysis. It was not to be expected, however, that this austere self-discipline would last very long – especially in light of the fact that the enterprises of both positivism and ordinary language philosophy met with little success. Inevitably, the penchant of philosophers for building their own view of the world would reassert itself, so that today little remains of analytic philosophy's once-vaunted disdain for metaphysics, or its ostensible ban on constructive theorizing.

But a lot of things do remain, many of which were there from the beginning – among them, ironically enough, a prevailing metaphysics: that of naturalism. For even when it tried its hardest to eschew metaphysics, analytic philosophy was guided, as any approach to philosophy must be, by a picture of the way things are, and that picture was of a world congenial to physical science. Here, however, there was a second irony, for while science had undergone dramatic change, the picture had not changed at all. The analytic movement was in part precipitated by a new scientific revolution, one dimension of which was the discovery that the behavior of subatomic particles seemed not to be deterministic after all, but was instead only probabilistic. One might have thought that this development would be welcomed by philosophers of action, as setting the stage for theory that would at once preserve free will and harmonize with the latest science. Indeed, precisely this suggestion came from some members of the scientific community.[1] But the idea did not gain great currency among philosophers, for reasons about which I shall speculate below. Rather, it was the older, mechanistic conception that dominated, and the strategies by which the notions of freedom and responsibility were to be brought into conformity with it were mostly those of the earlier era.

Things were not entirely the same. The Kantian strategy of sharply distinguishing the theoretical and practical realms and then confining notions of freedom to the latter was not much favored, especially in the early days, and it was never endorsed in the form Kant gave it. A version of this strategy did, however, gain favor among certain philosophers of the ordinary language movement – a typical example being A.I. Melden (1961). The essence of this approach is to apply the Wittgensteinian notion of a language game to discourse about human behavior. Such discourse may occur in any of several games, or patterns of discourse. Which game is going on depends on the interests of the speakers in question, and the concerns of speakers engaged in different games need not intersect. Thus, to a scientist interested in human behavior, what is observable when I raise my arm is a mere physical happening: the arm's going up. He will be interested in the inner, physiological causes of this event, and will seek to construct a deterministic explanation of it in terms of them – all of this, by Melden's lights, a perfectly legitimate endeavor. The concerns of everyday discourse, however, are altogether different. To the ordinary speaker, it was alleged,

the very same event which is my arm going up counts as a human action: my raising my arm. As such, it belongs to an equally legitimate framework of explanation, but one to which the language of deterministic causation is entirely alien. We explain actions in terms of the agent's reasons – that is, by citing his motives and intentions, which for Melden are *logically* related to action, and therefore cannot possibly cause it. And depending on how the agent was motivated – for example, on whether he was operating under duress – we may or may not decide that he could have behaved differently. But this decision belongs to the language game of action, not that of mechanistic causation. Indeed, even to raise the question whether an *action* is causally determined is to be guilty of a kind of logical howler, to mingle together incommensurate forms of discourse, thereby creating a false appearance of conflict. If we follow the proper course we will confine each language game to its peculiar sphere, and the conflict will be banished.

This maneuver is less radical, but the analogy between it and the one Kant attempted is obvious: if we cannot sequester the operations of voluntariness themselves from the arena of causality, we can at least sequester talk of what is voluntary from talk of what is caused – and so hope to reach a peaceful settlement between proponents of determinism and defenders of free will. Popular though it was for a time, however, this was not the settlement at which most analytic philosophers aimed. At least through the mid-twentieth century, most opted for what Kant had pronounced a wretched subterfuge (Kant 1956: 99) – that is, for a compatibilism which, though often more sophisticated in detail, was at bottom little changed from the heyday of the Enlightenment. A free action was an action that would have gone otherwise had the agent chosen, or perhaps tried, to do otherwise; a free decision was one that would have been different had the agent possessed adequate or sufficient motivation to decide differently. To say an agent could have chosen or done otherwise, it was urged, is simply to invoke these conditional relationships. And to hold people responsible for their deeds is to take advantage of the supposed deterministic sequence from motive to overt behavior, hoping to alter the effect by altering the cause. In one version or another, this type of compatibilism was defended early on by G.E. Moore (1912: Chapter VI), later by authors such as R.E. Hobart (1934), Moritz Schlick (1939) and A.J. Ayer (1954), and still more recently by Alvin Goldman (1970: Chapter 7) and Donald Davidson (1973).

It is fair to say, I think, that neither of these strategies met with much success. The most obvious failure was the sequestering maneuver, which ran afoul of the fact that no matter how diverse the language games of action and of physiological psychology may be, they speak of the same phenomena, and in terms that contradict. It may or may not be correct to say that when I perform the action of raising my arm, my action is identical with the event of my arm rising.[2] But it is true that when I raise my arm I normally

take myself to be bringing about my arm's rising, and to be doing so freely, so that its occurrence is up to me. Were it not so, I would not consider myself responsible for the event, or its consequences. By libertarian lights, however, my arm's rising will not be up to me if, as the physiologist says, that very event is caused by occurrences in my brain, which are themselves no more than a barely noticeable part of the vast causal network that constitutes the physical world, and whose present state Laplace's demon could have predicted with certainty a thousand years ago. So while players of the physiological language game and players of the human action language game may certainly agree to ignore one another, that doesn't change the fact that both groups cannot be right: in conversing among themselves, each group contradicts what the other has to say about human behavior.

The efforts of analytic philosophers to defend classical compatibilism also encountered serious obstacles. In his famous paper 'Ifs and Cans', J.L. Austin pointed out that not all *if*-clauses in English are conditional – so that even if, as G.E. Moore had maintained, to say 'I can' is as much as to say 'I can, if I choose', it would be wrong to jump to the conclusion that statements about what we can do have reference to a conditional relation between choice and action (Austin 1961). Even if such a relation is involved in freedom of *action*, moreover, it seems insufficient to capture what we have in mind by freedom of the *will*, or the ability to do otherwise. Thus, Roderick Chisholm, though not fully persuaded by Austin's rejection of compatibilism (Chisholm 1964), still thought a decisive argument was available. That an agent could have done otherwise cannot, he claimed, mean no more than that he would have done otherwise had he so chosen. For it may yet be the case – and if determinism holds, it certainly *would* be the case – that the agent could not have chosen otherwise. And if that is so, then he still could not have done otherwise (Chisholm 1997: 146–147). The essential insight here is that freedom *of the will* pertains not to the question what obstacles an agent may face in achieving the ends he wills, but rather to the etiology of willing itself. It may well be that if I choose to visit France after this conference, nothing will prevent me. But it seems empty to claim on this ground that it is up to me whether I will go there if my decision to go is settled in advance, a mere element of a uniformly causal framework in which all that takes place has been determined from the world's beginning. If my choice is the inevitable consequence of events over which I have no control, then it too is beyond my control. Peter van Inwagen has shown that this kind of argument can be made perfectly general (van Inwagen 1975). If determinism is true, then all that we do is settled by the distant past and the laws of nature, neither of which are in our power to alter. But if we cannot alter these, then we cannot alter what we do. It follows that there is no freedom; we can never do otherwise – and so, it seems, are never responsible.

There is, however, a third strategy to consider, unlike the first two in its claims about freedom, but still compatibilistic in spirit. Perhaps partly in

response to criticisms of classical compatibilism, a number of philosophers have followed Harry Frankfurt in maintaining that if by freedom is meant the legitimate possibility of doing otherwise, then responsible agency does not require free will after all (Frankfurt 1969). This, they argue, is because the conditions that rule out alternative possibilities need not enter into the actual etiology of choice or action – so that despite their presence, agency is still exercised normally. Suppose, for example, that I decide in the normal way that following this conference, I will visit France. Since no obvious compulsion is at work in my decision, we would doubtless take it as an exercise of free will. Suppose, however, that my old friend Pierre, who is bent on having me visit France, has implanted a device in my brain that would enable him to force me to decide to do so. As it turns out, Pierre doesn't have to use the device in this case, since I decide that way anyhow. But, the argument runs, Pierre also has means of knowing what I am about to decide; and had I been about to decide differently, he would simply have pushed a button on his hand-held computer, thereby forcing me to decide to visit France. This is a case, then, in which I could not have decided differently. Yet, given what actually occurred, it appears I decided responsibly.

The appeal of such examples is that if they are legitimate, we may approach issues of moral responsibility in an essentially compatibilist spirit, while avoiding implausible claims about what constitutes freedom. For even though, in our example, Pierre is poised to control my decision if need be, the decision actually occurs without his intervention. Accordingly, the argument runs, my faculties of deliberation and decision can be assumed to have operated normally. And as long as the etiology of my decision is normal, I can be held responsible for it. We need not worry about whether my choice was determined from the remote past, or about whether I could have chosen differently, if that is what constitutes free will. For, it is alleged, the example shows that this kind of freedom is not necessary for moral responsibility.

But are such examples legitimate? Several philosophers, most prominently David Widerker,[3] have argued plausibly that they are not, at least from a libertarian perspective. We may assume that it takes some time for a decision to take place – though doubtless it requires very little – and let t be the moment in our example when I commenced to decide to visit France. The question is, might I have done otherwise at t? Might I, for example, have commenced to make some other decision? The answer depends on how Pierre would have known that I was going to do so. If he would have had to wait until t – that is, until the moment I commenced to decide not to visit France – then the example fails. For in fact I could have done otherwise: I could have undertaken to decide not to visit France. I might not, of course, have completed the decision if Pierre was able to interfere quickly enough. But that does not matter. Freedom is freedom, no matter how brief its exercise. In order for the example to succeed, therefore, Pierre must have some reliable sign *prior to* t as to what I will decide – that is, there must be

some signal whose occurrence prior to *t* is *guaranteed* if I am about to make any other decision. But if the signal is guaranteed, then its occurrence is in essence a causally necessary condition for my deciding not to visit France. Consequently, although it is true that I could not, at *t*, have decided differently, the reason has little or nothing to do with Pierre. It is because a causally necessary condition of my doing otherwise – namely, the prior occurrence of the signal – was lacking. What this indicates is that Frankfurt-style examples can be constructed only by begging the question against libertarian notions of freedom. If the claim that a person could have done otherwise means what libertarians take it to mean, such examples would appear to be impossible,[4] and if that is so then the third strategy by which analytic philosophers have sought to develop a theory of responsible action in line with deterministic principles also fails.

Indeterminism and naturalism

One cannot, of course, consider the matter closed; there is always the possibility that some future effort to bring peace between determinists and free will advocates will be found to succeed. For the present, however, the hostilities continue. Yet one may wonder why allegiance to determinism is so popular among those who would favor a naturalistic account of human behavior. It was mentioned above that part of the inspiration for the analytic movement in philosophy came from developments in science, among them the rise of quantum mechanics, according to which subatomic phenomena do not obey deterministic principles. We should note that this claim at first encountered considerable resistance among scientists. Today, however, quantum theory is solidly entrenched, and if it is true there is reason for thinking that, at least as far as indeterminacy goes, libertarian free will can in principle be brought into line with naturalism. The key requirement is that quantum-level indeterminacies, which ordinarily cancel one another out at the macro-level, be able to have impact at the level of neurological firings – that is, at the level of the physiological processes we take to correspond to mental occurrences. If this were so, then the neurological events that correspond to acts of decision and volition might after all turn out to be indeterministic, so that exercises of free will would be secured as to their physiological counterparts. Robert Kane has argued that this kind of impact is not at all implausible to imagine. His suggestion is that, especially in situations of moral temptation, the tension of the deliberative process corresponds to a chaotic situation in the brain that is able to amplify quantum-level perturbations, so that the neurophysiological sequence that corresponds to the agent's decision is rendered indeterminate (Kane 1996: 129–130). How likely it is that this particular suggestion will turn out correct is, of course, anybody's guess. But the sources Kane cites indicate that if chaos theory can be wedded to quantum theory, macro-level sensitivity to quantum-level events could turn out to be widespread in the

nervous system; and if that is so, then some account along the lines Kane suggests may well be found to succeed.

Here, then, is a possibility that deserves to be explored. Yet, at least until recently, it has received rather little in the way of philosophical attention, and is still not much discussed. Part of the reason, certainly, has been the state of scientific knowledge. Until the development of chaos theory, no one had any clear idea how quantum perturbations might be magnified so as to be reflected in the physiology we think goes with decision making. And even if we view this problem as now on the way to being surmounted, we remain woefully ignorant of the precise details of the physiology in question. We know a little about the brain functions involved in deliberation and decision, but we are nowhere close to being able to say what specific brain events might correlate with an act on my part of deciding to visit France, and how they might differ from those which would correlate with my deciding to remain in Italy. Any effort along the lines Kane proposes would therefore have to be considered speculative and uncertain. But I do not think this is the whole answer, for in my experience philosophers who find the idea of a quantum mechanical treatment of free will attractive tend not to have strong naturalist sentiments anyway. Rather, they are first and foremost defenders of free will; and they find the quantum mechanical option attractive not as a step toward a naturalist theory of human decision and action, but as a partial counter to the objection that libertarian freedom is an irrational notion, one which cannot be squared with a properly scientific view of the world. By contrast with this perspective, I suspect that anyone truly dedicated to developing a naturalist theory of decision and action would find the quantum mechanical gambit rather unattractive. Why should this be?

The answer, I think, is that quantum mechanics is in itself offensive to the sentiments that underlie philosophical naturalism. Not that it doesn't fit with the definition of naturalism adopted at the beginning of this paper. Quantum mechanics is certainly physics, and Kane's views certainly represent a step in the direction of treating free will in terms of science as we know it. The problem is that in this case science as we know it runs counter to what I described earlier as the positive thesis associated with naturalism: that the concepts of physical science will ultimately provide a thorough and complete understanding of all that we experience. Even for strictly physical phenomena, quantum mechanics does not do that. Just the opposite: it is fundamentally statistical, and on the usual interpretation finds the world to be irreducibly probabilistic. If it is true, then nothing recognizable by contemporary standards as physics is going to be able to tell us why a particular atom of radium decays when it does, or why a particular photon in the classic double-slit experiment strikes the reflecting screen where it does. Is this an embarrassment for physics? Not really, in my opinion. The world is what it is, and if quantum theory describes the world correctly, then as science it must in the end be considered not just a success but a triumph.

For naturalism, however, quantum mechanics represents a defeat; if it is true, then a central thesis of naturalism is false.

This, I suggest, is what explains the hesitancy of naturalistically minded philosophers to embrace the idea of a quantum-theoretic treatment of the underlying physiology of decision and volition. Even if such a treatment were found to succeed it would be a disappointment; it would leave us without a full, scientific accounting for exercises of human agency. Worse yet, from the naturalist's perspective, it would invite philosophers of other persuasions to try to fill the explanatory gap with some sort of teleological explanation in terms of the agent's motives or intentions – something of which contemporary naturalists are no more fond than were their empiricist forebears. I shall have more to say below about such explanations, which I think are legitimate and valuable. In fact, however, I don't think the naturalist need fear that they will close the gap in our understanding of decision and action that would be left by a failure of determinism. If we are free, then no consideration we can cite will provide the kind of explanation for decision and volition that a determinist would want – that is, an explanation that closes off all alternative possibilities. All the same, a good, orthodox naturalist should find the association of free will with quantum indeterminacy an unwelcome prospect. Should it succeed, it would cement quantum theory in place more firmly than ever. And it would make the uncertainty naturalists are wont to complain about when it comes to free will part and parcel of an uncertainty that goes straight to the bottom of reality, physically understood.

Naturalism and psychological explanation

But perhaps the naturalist will be prepared to make a concession at this point. After all, he may reason, the success of quantum mechanics makes it unlikely in any case that a completely satisfying physical account of the world will be found. The discomfiture this brings to naturalism may be increased if quantum uncertainty turns out to be manifested in human decision and action, but the basic reality would be unchanged. Moreover, the really important tenet of the naturalist's creed – that the world is to be understood in terms of physics – could end up surviving such a discovery fairly nicely, just as it survived the discovery of quantum mechanics. After all, to the extent we understand quantum phenomena at all, we do so in terms of concepts that are recognizably and strictly physical. And, the argument would run, the same will eventually prove true of human behavior. Determined or not, human decision and volition will finally be understood as physical phenomena. That is, they will ultimately be reduced to physiological processes in the human brain – phenomena whose explanation, however complete or incomplete, will proceed in physical terms, the only terms appropriate to legitimate explanation.

It is, I think, on this point that we that we begin to approach a final and unbridgeable impasse between naturalism and libertarian free will. The

reason for this has to do with consciousness. From a libertarian perspective, the real value of free will lies not in the mere fact that it separates human willing from the natural causal order, but in the fact that this separation makes it possible for individual agents to exert some control over their destinies, to guide their lives according to purposes that *they* set. Paradigmatically, this is accomplished through conscious deliberation – that is, by the agent envisioning possible futures for himself, and deciding among them. And it is vital that this process be conscious. That is one reason why we never hear of anyone reaching a decision in dreamless sleep, or forming an intention while in a coma. It makes no sense to think of an entity that is not conscious envisioning a future, or finding in it any value worth pursuing. The naturalist position, by contrast, is that consciousness does not matter, that a true comprehension of the relation between an act of will and the reasons for which it occurs are to be had in terms that are physiological, not psychological. Even if the relation between reasons and willing turns out not to be fully deterministic, therefore, the naturalist can still insist that that relation will one day be replaced in our theories by relations between the physiological processes to which reason and willing are reduced. Only when this occurs, he will say, can we have a scientifically respectable account of human decision and action.

I think, however, that the difficulty here runs far deeper than the naturalist, or at least the naturalist I have depicted, takes it to do. This can be seen by considering what it is that counts as a 'reason' for which an action might be performed. Naturalists – and often their opponents too – are disposed to think of a reason as a mental state or event with the potential to give rise to action. Thus, the event of my experiencing a desire to visit my friend Pierre might be viewed as a reason for which I might decide to visit France. Should I proceed to make the decision, the event of my feeling the desire to see Pierre would be expected to figure prominently in the explanation of my decision – a deterministic explanation if determinism turns out to be true after all, or some kind of statistical account if it does not. In fact, however, what is treated in common life as a reason for deciding or acting cannot figure in either of these sorts of explanation. A reason for deciding has to be something I can *think* – and I cannot think an event, even the mental event of my experiencing a desire to visit Pierre. Rather, I must think the *content* of the event – something we might express in the optative mood as: *Would that I visit good old Pierre*. This, unlike the event of my thinking it, is an abstraction or *ens rationis* – a thought, rather than the thinking of it. We have already taken note of Reid's claim that this kind of entity cannot figure in a causal explanation of an act of will. He was right about that: a cause has to be an event or state in the real world, not an abstract entity, and the same goes for anything we might point to as 'statistically causing' or explaining something.

When we speak of desires as explaining decision or action, then, what is at work is neither a causal nor a statistical form of explanation. What kind

of explanation is it? Obviously, a teleological one, an effort to account for what the agent does in terms of the envisioned objective at which he aims. It is appropriate to think of such explanations as encapsulating the agent's own practical reasoning. Thus, if we explain my decision to travel to France by citing my desire to visit Pierre, we treat the objective whose ostensible fitness for achievement I apprehend through my desire as the ground for my decision to pursue it. The reasoning by which I reach my decision can be summarized in a kind of Aristotelian practical syllogism:

> Would that I visit Pierre.
> If I travel to France, I can visit Pierre.
> Therefore, I shall travel to France.

Here my decision is presented as justified by the objective my desire sets before me, in that by deciding to visit France I form an intention which, if carried out, will count as a step toward visiting my friend, a goal which will eventually be achieved by timely formation and execution of further intentions directed toward it.

There is much to be said about this kind of reasoning, but I will content myself with two points. First, this reasoning is practical not just as to content, but functionally as well. It does not end as 'practical reasoning' is sometimes held to end – that is, in a judgment as to what is best, which is then left to blind chance (or should I say blind causation?) as to whether it will lead to intention formation. Rather, this reasoning terminates in decision itself, which *is* intention formation. It is too much to go into here, but I think similar reasoning, terminating in the activity of volition, is what occurs in the execution of intention – so that both these types of active willing count also as ratiocination, or drawing a reasoned conclusion. Second, there is a clear and very satisfying logic to this kind of reasoning. This has eluded many philosophers, because they tend to assess all reasoning along epistemic lines – that is, as though it were an effort to learn more about the world, by drawing the right conclusions from independently confirmed evidence. On this understanding, there is no way the practical syllogism cited above can be judged valid or cogent. How does one deduce an intention – that is, the content of my decision – from the optation, 'Would that I visit Pierre', and a belief about how to move toward satisfying that optation? Clearly, one cannot; there are no logical principles for deducing intentions from desires and beliefs. This approach to evaluating our practical syllogism is, however, misguided. The function of intention formation is not epistemic; it is not to bring the mind into conformity with the world. Rather, it is to bring the world into conformity with the mind: to form and carry out intentions directed at changing the world to suit our preferences. Viewed in this light, our practical syllogism makes perfect logical sense. For if its conclusion – that is, my decision to travel to France– is successfully carried out, and if I am correct in believing that this will put

me in a position to visit good old Pierre, then by good old *modus ponens*, I will be that much closer to accomplishing my objective. Or, to state the point a bit more formally, if the minor premise and conclusion of the syllogism are satisfied, the major premise will progress toward being satisfied as well – exactly the relationship we should be looking for if the 'direction of fit' is world to mind, rather than vice versa (Kenny 1975: Chapter 5, especially 80–82).

Brief as they are, these points begin to illustrate the advantages of realizing that when, in common life, we speak of the reasons that explain a decision or action, we refer to what are in fact abstractions, not states or events. To view the operations of voluntary willing in this way is to see them as ratiocinative processes, wherein responsible agents guide their conduct toward consciously chosen ends. It is far from clear, however, that any of this can be done justice on a naturalist approach. From that perspective, my desire to visit Pierre would likely be viewed not as content but as the thinking of the content – that is, as a psychological event, to be understood in terms congenial to science. In recent years, that has meant submitting the event to functional analysis. That is, my desiring to visit Pierre would be treated as though what is essential to it *qua* mental is not the content of which I am conscious when it occurs, but rather the way it interacts with the rest of what goes on in my head when I deliberate over where to go after this conference. Having completed the functional analysis, the naturalist would next point out that what goes on in my head must after all be physiological, so that the desire-event may be identified with whatever brain process is found empirically to correlate to it, no mention of conscious content being thereafter necessary. Or, it might be held that any conscious goings on associated with it are at best supervenient, and so count only as epiphenomena. And, of course, the same treatment can be given for my belief as to how I may go about visiting Pierre, and my decision to travel to France to see him. Finally, it will be pointed out that since all of these events are at bottom physiological, the explanatory relations that join them, whatever they are, must be physiological as well. Thus, the naturalist program is fulfilled.

I think it is obvious, however, that this approach to understanding decision making will not satisfy defenders of libertarian freedom. For even if the explanatory relations of which it speaks turn out not to be deterministic, they leave conscious content completely out of account. Yet, as we have seen, there is no point to libertarian freedom, except as a precondition for agents to play a meaningful, autonomous and responsible role in carving out their personal destinies – something that is possible only on the supposition that they can entertain and select objectives. The libertarian will argue that it is vacuous to think someone could prosecute such an enterprise except by working with conscious content – vacuous to hold, in effect, that a zombie could as truthfully be said to make decisions as you or I. Moreover, he will say, it is a mistake to think the content of our desire and belief states is

merely epiphenomenal. The simple truth is that citing the desires and beliefs out of which a decision is made does provide understanding. A complete account of the operations of the will has to explain why it does so, not ignore or pass over the fact.

What all of this comes to is that, in the context of science as we know it today, the really important disagreement between the naturalist and the libertarian is not about indeterminism, but about the reality of consciousness, and the role it plays in behavior. And I do not see how this disagreement can be overcome. To do so would require two things. First, it would take a functional account not just of mental states and events – of the thinking of content – but of content itself. And of course there have been efforts along these lines. I cannot consider them here, but they face significant obstacles. Mental items such as desires, beliefs and intentions are not concrete entities but *abstracta*. They are, that is, proposition-like, and not to be identified with any kind of sign or physical representation, even in the brain itself, of which they might be held to constitute the meaning or significance. Furthermore, entities like this share logical relations, not functional ones, and it is in terms of these relations that they are explanatory. Thus, although the mental events of which they are the content must certainly play a functional role in the etiology of decision and action, the explanatory role of the content cannot reduce to the functional role of the events. Indeed, one could go further: there is no obvious reason why the functional role of the events could not turn out to be grounded in the explanatory relations among their contents, rather than the other way around.

But even if these points can successfully be addressed, there remains a final and I think insuperable barrier separating the libertarian from the naturalist: the fact of consciousness itself. The simple truth is that there is, in the phrase now popular, something it is like to be an agent, and a sensible libertarianism must maintain that were this not so, free will would be an empty shell. Whatever else might be said about mental content, it is only *as presented to the agent* that it makes possible the significant exercise of moral autonomy. Yet this final thing, consciousness itself, is the one thing naturalism cannot accommodate. To include it in our theories – to include, that is, what I actually experience when I desire to visit Pierre – is to admit an entirely new sort of datum, a kind that cannot be represented in the vocabulary of physics. And to admit that kind of datum is to give up being a naturalist.

A final note

If the foregoing arguments are correct, a libertarian account of free will cannot be a naturalistic one. Is it, then, the case that libertarian views of the will are not to be considered analytic philosophy, in the truest sense of the term? It is certainly possible to make such a claim. One need only characterize analytic philosophy narrowly enough that it turns out to

exclude any view of the will that is not strongly naturalistic. And there may have been a time – in the days of the Vienna Circle, perhaps – when such a characterization would have had a certain persuasiveness. But those days are long gone. The radical claims of logical positivism and the ordinary language movement have today very few defenders, and the movement that began with figures like Frege, Russell and Wittgenstein has long since broadened its scope to recover much of the historical heritage it once pretended to leave behind. There remain, to be sure, prevailing orthodoxies, naturalism among them. But the tradition of discourse that today goes by the name 'analytic philosophy' has by now come to encompass many perspectives, and is easily tolerant enough that one can comfortably consider oneself both a defender of libertarian freedom and an analytic philosopher. After all, I am one myself.

Notes

1 See, for example, Compton (1935).
2 I have argued elsewhere that it is not correct. See McCann (1998, Chapter 3).
3 Widerker (1995). See also Kane (1996: 142–143).
4 For further discussion of this issue see Widerker and McKenna (2003).

Practical reasoning and action

Comment on McCann's paper

Francesca Castellani

In his paper, McCann examines the origin and the development of the gap between naturalism and free will, the efforts to bridge it, and how such efforts have been received in analytic philosophy. His analysis has the merit of reminding us that analogous versions of the problem now confronting us in updated manner have already emerged in other cultural settings. This updating largely concerns the developments in science, which, as some claim, is now able to incorporate uncertainty with regard to free will as well as it is able to handle quantum uncertainty.

McCann leads us step by step to the point where the gap between naturalism and free will seems insuperable. Given the manifold facets of naturalism, we should bear in mind that McCann focuses on its main thesis that meaningful philosophical problems 'are to be addressed, and their solution framed, in terms acceptable to natural science, particularly physical science' (McCann, p. 225). As to free will, here too there are different views: McCann focuses on libertarian free will. He emphasizes that the crucial point regarding free will is not whether its operations, and ultimately intentional actions, are either deterministic or indeterministic; rather, it is whether or not they are subject to nomic regularities. In fact, whether such regularities can be related to universal or probabilistic laws does not matter, since in both cases the explanatory model is one and the same, namely the covering-law model according to which human action is a particular case of a general regularity. Besides *a priori* considerations on how action is constituted as such, namely as distinct from mere observable behaviour, we have good *a posteriori* evidence that a nomological explanation of intentional action is unsatisfactory, in that it completely neglects the subjective dimension constituted by the reasons out of which agents perform their actions. The crucial point is the one emphasized by McCann: according to the nomological model, the process by which an agent forms her intentions, chooses the ends towards she directs her conduct, is an obscure, mysterious and even irrational phenomenon releases each of us is well aware of the reasons on which our intentions and selected ends are grounded.

For these and other reasons cited by McCann, it is advisable to dispense with the nomological model and switch to a teleological explanation of

action based on the agent's reasons for undertaking it. As numerous philosophers have argued, at least since von Wright (von Wright 1971), and as McCann argues as well, such teleological explanation encapsulates the agent's practical reasoning. Let us suppose that after the conference McCann went to France, and we want to explain why he did so. The explanation will consider the practical reasoning of agent McCann and will divide into two steps. The first concerns intention formation; the second concerns the realization of that intention through performance of the corresponding action.

The practical syllogism presented by McCann covers the first step and shows that such reasoning involves the conceptual contents of our mental states or processes:

1 Would that I visit Pierre (= content of desire)
2 If I travel to France, I can visit Pierre (= content of belief)
3 (Therefore) I shall travel to France (= content of intention)

The logical structure of this reasoning may be rendered as follows:

$1'$ Desires x, that p
$2'$ Believes x, that $q \rightarrow p$
$3'$ (Therefore) Intends x, that q

where Desires, Believes, and Intends are all intensional operators: the clause 'that' is followed by a statement, but what matters is the proposition it expresses, its intension. Such a clause introduces not actual states of affairs but possible, desired, intentioned states of affairs, or ones believed to be true.

McCann claims that there is a clear logic to this kind of reasoning, one that meets criteria of practical validity, rather than ones of theoretical validity, in the following sense: if the conclusion and the minor premise are satisfied, a step toward satisfying the desire expressed in the major premise has been accomplished. Therefore, according to McCann, practical reasoning is not to be assessed along epistemic lines. Now I agree that practical reasoning 'on active service', as Anscombe puts it (Anscombe 1957: 60), is intended to change the world so that it suits our desires and intentions. But explanation, even in practical contexts, has a theoretical-epistemic function: it is not about forming a decision or an intention; rather, it is about knowing the reasons why the agents decided to do something, for example, to travel to France. McCann maintains that, from a theoretical point of view, one cannot 'deduce an intention – that is, the content of my decision – from the optation, "Would that I visit Pierre", and a belief about how to move toward satisfying that optation' (McCann, p. 237). Here McCann seemingly introduces a cleavage between practical and theoretical rationality which I am unable to reconcile with the fact that he introduces practical reasoning in an explanatory context. The fact that practical reasoning justifies intention by working backwards, from the conclusion and minor premise to

the major premise, likens it to any kind of explanation: we start from something that has already occurred or been accomplished – which we want to explain, and we trace back to what it is that explains it. Is McCann suggesting that the logic of action is different from the logic of explanation? Does the practical reasoning by which an agent decides what to do in order to achieve some goal follow a logic different from the one that presides over the logical reconstruction of the same reasoning pursued for the purpose of explanation? I do not think this is what McCann means, since he introduces practical reasoning in the context of teleological explanation. There is no doubt that the context in which practical reasoning is directed towards deciding what to do differs from the context in which it is directed towards explaining what has been done, but this is only a pragmatic difference. In the first context of use, the argumentative steps are, so to speak, processed in imprecise manner, at a glance. Because the world rapidly changes, and so does our knowledge about it, we have little time to make up our minds, and our practical reasoning is often elliptic or in any case flawed. By contrast, when we explain why we acted as we did, we reason, so to speak, when the game is over: the decision has already been taken, the situation is not going to change, the conditions are fulfilled, so that we are able to reconstruct our practical reasoning more accurately. I wonder why this pragmatic difference should give rise to a logical difference.

I now turn to some matters of detail connected with the foregoing discussion. The first concerns the minor premise of practical reasoning. This states that the content of agent McCann's belief is that if he travels to France, he can visit Pierre; q is a sufficient condition for p. The agent believes there are alternative means to satisfy the desire expressed by the major premise. This seems to be a normal condition for the free exercise of will in the majority of practical contexts. Moreover, it matches the fact that we make most of our decisions without any careful consideration: suffice it that the option chosen is adequate to achieve our goals. Yet, it may also happen that the agent believes that the means to satisfy the desire or the intention expressed by the first premise are necessary; in other words, it may well be that the agent believes that q is a necessary condition for p, that he can visit Pierre only if he travels to France.

The second comment concerns how we realize our intention by performing the corresponding action – that is, how action follows from an intention: this is the second step of practical reasoning. McCann only mentions this issue in passing. Yet I deem it important and I would like to raise it. For intention to be followed by action, the agent must feel committed to achieving the intended ends: that is, she must assume those ends as obligatory in the given situation. The question is: why, if we intend to do something, do we feel committed to doing it? This commitment to act is something like a felt obligation: we feel obliged to act so as to achieve the intended ends. Whence does this feeling of obligation derive? All obligations refer to an axiological ordering. Consider a concrete and subjective axiological

ordering, like the one that orders the preferences of a subject in the given concrete situation where she happens to be, and includes everything that she believes to be positive for herself *hic et nunc*, all things considered. Such concrete subjective preferences do not depend solely on the intention to pursue abstract values: they also depend on the subject's inclinations, desires, concrete expectations in a given situation.

Now I believe it plausible (1) that preferential beliefs, in the absence of obstacles, are conducive to action; and (2) that there is a clear correspondence between preferential beliefs and will, so that willing something is constitutively equivalent to being convinced that it is optimal as far as our concrete preferences are concerned (Galvan 1992: 189–211). In other words, the object of our willing, the content of our intention, is what we believe to be optimal *vis-à-vis* our preferences in the given situation. There is no volition if the object of one's willing is not deemed optimal *vis-à-vis* one's own preferences in the given situation. Vice versa, it would be senseless to judge something optimal *vis-à-vis* one's own concrete preferences and not to will it. Note that the equivalence proposed between willing and preferential belief does not entail that the agent intends to do something because it is optimal or because the agent believes it is optimal with reference to an abstract axiological ordering; rather, it states that the agent intends to do it because she judges it optimal in the light of her present expectations and preferences. Therefore, what an agent believes to be optimal may vary depending on different concrete situations and may not match what would be abstractly optimal for her. On the basis of this correspondence between the epistemic level of preferential beliefs and the intentional level of the will, we obtain analytically the principle according to which will, or intention, in the absence of impediment, is conducive to action. This principle is not put forward as an empirical law stating an empirically testable regularity according to which all those who intend to do something, and are not prevented from doing so, act correspondingly – without exception or in a given percentage; rather it is introduced as an analytical principle, as a meaning postulate for the operator of willing and intending, stating that volition is constitutively related to the willed action. In other words, it states that the content of willing is such that, in all worlds where the consciously ordered concrete preferences of the subject are realized, the object of willing is achieved as well.

Now, this analysis of a volitional or intentional context as a particular deontic context occurring within the context of the agent's beliefs may account for the commitment to action which characterizes intention. Though in itself motivationally inert, the deontic context of optimality becomes efficacious when it is mediated by the agent's beliefs. The presence of the deontic optimality operator within the relevant epistemic context, gives formal account of the constraint exerted by the intention to pursue consciously selected ends; that is to say, if an agent believes *hic et nunc* that realizing those ends is optimal for her in the given situation, she will feel committed to achieving them.

Finally, I wish to emphasize that I fully agree with McCann concerning his crucial claim that the autonomy and responsibility of agents in modelling their own personal destinies involves their ability to select their objectives – an ability which requires the elaboration of conscious contents. It is precisely on this point, as McCann appropriately emphasizes, that the attempt to naturalize will by explaining intentional action in functional terms breaks down. I believe that this failure is illustrated in exemplary fashion by artificial intelligence architectures to construe autonomous 'agents' based on data structures representing beliefs, desires and intentions (BDI), and on functions which preside over the practical reasoning by which such agents decide what action to undertake (Wooldridge 1999: 27–77). For these artificial agents, acting intentionally means following a certain procedure with a view to achieving an objective: the intention is represented in this case as an inner state with characteristic functions. A BDI agent is said to be autonomous in that it is able to generate its own goals. Yet it should be emphasized that this ability only concerns means-to-end reasoning: that is to say, it only concerns the generation of intermediate or secondary goals to be accomplished with a view to achieving the primary objective for which the agent was programmed. But BDI models leave the commitment to action unexplained: they have it spring spontaneously from intention as armed Minerva from Jupiter's head. Moreover, primary objectives fall outside the scope of a BDI agent in that they are imposed from outside. In other words, although a BDI agent is autonomous, it is programmed to pursue certain purposes rather than others. Which is a remark that appears obvious if by 'intentionality' is meant the presence within the agent of conscious mental contents: BDI agents are able to perform the functions associated with intention, but they are devoid of intentionality. Nevertheless I think that this is a remark which, like analogous ones, should be constantly reiterated and stressed. Otherwise we run the serious risk that, following the success of artificial intelligence and of the neurosciences, and who knows of what else in the future, the original philosophical meaning of numerous terms – like intention, belief, desire, memory, and even consciousness – will be lost, being replaced by surrogates whose advantage is that they do not clash with the vocabulary of the natural sciences.

Bibliography

Adams, R.M. (2000) *Finite and Infinite Goods*, Oxford: Oxford University Press.

Aitchison, I. and Hey, A. (1989) *Gauge Theories in Particle Physics: A Practical Introduction*, 2nd edn, Bristol: Adam Hilger Publishing Company.

Anscombe, G.E.M. (1957) *Intention*, Oxford: Basil Blackwell.

Armstrong, D. (1997) *A World of States of Affairs*, Cambridge: Cambridge University Press.

Atkins, P.W. (1992) *Creation Revisited*, New York: W.H. Freeman.

Audi, R. (1983) 'An Epistemic Conception of Rationality', *Social Theory and Practice* 9, 2–3: 311–334.

——(1989) *Practical Reasoning*, London: Routledge.

——(1993) 'Ethical Naturalism and the Explanatory Power of Moral Concepts', in S. Wagner and R. Warner (eds) *Naturalism: A Critical Appraisal*, Notre Dame, IN: University of Notre Dame Press, pp. 95–115.

——(1996) 'Naturalism', in *The Encyclopedia of Philosophy Supplement*, ed. by D.M. Borchert, New York: Macmillan, pp. 372–374.

——(1997) *Moral Knowledge and Ethical Character*, Oxford: Oxford University Press.

——(2000) 'Philosophical Naturalism at the Turn of the Century', *Journal of Philosophical Research*, XXV: 27–45.

——(2002a) 'Prospects for a Naturalization of Practical Reason: Humean Instrumentalism and the Normative Authority of Desire', *International Journal of Philosophical Studies,* 10, 3: 235–263.

——(2002b) 'An Internalist Theory of Normative Grounds', *Philosophical Topics,* 29 (1,2): 19–46.

——(2004) *The Good in the Right: A Theory of Intuition and Intrinsic Value*, Princeton, NJ: Princeton University Press.

——(forthcoming) 'Divine Command Morality and the Autonomy of Ethics'.

Austin, J.L. (1961) 'Ifs and Cans', in J.L. Austin, *Philosophical Papers*, ed. J.O. Urmson and G.J. Warnock, Oxford: Clarendon Press, pp. 153–160.

Ayer, A.J. (1936) *Language, Truth, and Logic*, London: Penguin.

——(1954), 'Freedom and Determinism', in *Philosophical Essays*, London: Macmillan, pp. 271–284.

Barnes, J. (1995) 'Metaphysics', in *The Cambridge Companion to Aristotle*, ed. J. Barnes, Cambridge: Cambridge University Press, pp. 66–108.

Barrow, J.D. (1991) *Theories of Everything*, Oxford: Clarendon Press.

——(1994) *The Origin of the Universe*, New York: Basic Books.

Barrow, J. and Tipler, F. (1986) *The Anthropic Cosmological Principle*, Oxford: Clarendon Press.

Bartholomew, D.J. (1984) *God of Chance*, London: SCM Press.

Beilby, J. (ed.) (2002) *Naturalism Defeated? Essays on Plantinga's Evolutionary Argument Against Naturalism*, Ithaca, NY: Cornell University Press.

Bergmann, M. (1997) 'Internalism, Externalism and the No-Defeater Condition', *Synthese,* 110: 399–417.

——(2000) 'Deontology and Defeat', *Philosophy and Phenomenological Research*, 60: 87–102.

Black, M. (1962) 'Review of *The Natural Philosophy of Time*', *Scientific American,* 206: 181–183.

Block, N. (1995) 'On a Confusion about a Function of Consciousness', *Behavioral and Brain Sciences*, 38: 227–247.

Bloomfield, P. (2001) *Moral Reality*, Oxford: Oxford University Press.

Bludman, S.A. (1984) 'Thermodynamics and the End of a Closed Universe', *Nature,* 308: 319–322.

Bolzano, B. (1827) *Athanasia*, Sulzbach: Seidel.

——(1994) *Wissenschaftslehre*, Stuttgart-Bad Cannstatt: Frommann.

Bondi, H. and Gold, T. (1948) 'The Steady State Theory of the Expanding Universe', *Monthly Notices of the Royal Astronomical Society,* 108: 252–270.

Boolos, G. (1971) 'The Iterative Concept of Set', *The Journal of Philosophy*, 68: 215–232.

Borde, A. and Vilenkin, A. (1994) 'Eternal Inflation and the Initial Singularity', *Physical Review Letters,* 72: 3305–3308.

Boyd, R.N. (1988) 'How to Be a Moral Realist', in G. Sayre-McCord (ed.) *Essays in Moral Realism*, Ithaca, NY: Cornell University Press, pp. 181–228.

Brandt, R.B. (1979) *A Theory of the Good and the Right*, Oxford: Oxford University Press.

Brentano, F. (1924) *Psychologie vom empirischen Standpunkt*, Leipzig: Meiner.

Brink, D.O. (1989) *Moral Realism and the Foundations of Ethics,* Cambridge: Cambridge University Press.

Broad, C.D. (1938) *An Examination of McTaggart's Philosophy,* 2 vols, Cambridge: Cambridge University Press.

Buber, M. (1983) *Ich und Du*, Darmstadt: Wissenschaftliche Buchgesellschaft.

Bunge, M. (1985) *Treatise on Basic Philosophy*, vol. 7: *Epistemology and Methodology III: Philosophy of Science and Technology,* Part I*: Formal and Physical Sciences*, Dordrecht: D. Reidel.

Carr, B.J. and Rees, M.J. (1979) 'The Anthropic Cosmological Principle and the Structure of the Physical World', *Nature*, 278, 12 April: 605–612.

Chalmers, D.J. (1995) 'Facing up to the Problem of Consciousness', *Journal of Consciousness Studies,* 2: 200–219.

——(1996) *The Conscious Mind*, Oxford: Oxford University Press.

Chisholm, R.M. (1964) 'Austin's *Philosophical Papers*', *Mind* 73: 1–26.

——(1997) 'Human Freedom and the Self', *The Lindley Lecture*, University of Kansas. Reprinted in *Free Will*, ed. D. Pereboom, Indianapolis, IN: Hackett, pp. 143–155.

Collins, R. (1998) 'A Defense of the Probabilistic Principle of Indifference', lecture to History and Philosophy of Science Colloquium, University of Notre Dame, Oct. 8. Presently unpublished.

——(2002a) 'The Argument from Design and the Many-Worlds Hypothesis', in W.L. Craig (ed.) *Philosophy of Religion: A Reader and Guide*, Trenton, NJ: Rutgers University Press, pp. 130–148.

——(2002b) 'God, Design, and Fine-Tuning', in R. Martin and C. Bernard (eds) *God Matters: Readings in the Philosophy of Religion*, New York: Longman Press, pp. 119–135.

——(2003) 'The Evidence for Fine-Tuning', in N.A. Manson (ed.) *God and Design: The Teleological Argument and Modern Science*, London and New York: Routledge, pp. 178–199.

Coliva, A. (1999) 'Wittgenstein: certezza versus scetticismo e naturalismo', in D. Marconi (ed.) *Naturalismo e naturalizzazione*, Vercelli: Edizioni Mercurio, pp. 169–187.

Compton, A.H. (1935) *The Freedom of Man*, New Haven, CT: Yale University Press.

Conee, E. (1996) 'Plantinga's Naturalism', in J.L. Kvanvig (ed.) *Warrant in Contemporary Epistemology: Essays in Honor of Plantinga's Theory of Knowledge*, Lanham, MD: Rowman & Littlefield, pp. 183–196.

Corradini, A. (1995) 'Supervenienz und Universalisierbarkeit. Eine Auseinandersetzung mit R.M. Hare', in C. Fehige and G. Meggle (eds) *Zum Moralischen Denken*, Frankfurt am Main: Suhrkamp, pp. 194–228.

——(2003a) 'Der naturalistische Fehlschluss in der naturalistischen Ethik: Herbert Spencer und die gegenwärtige Debatte', *History of Philosophy and Logical Analysis*, 6, Paderborn: Mentis, pp. 219–235.

——(2003b) 'Naturalism and Moral Agency', in J. Quitterer and E. Runggaldier (eds) *Persons: An Interdisciplinary Approach,* Vienna: Hölder-Pichler-Tempsky, pp. 199–207.

——(2003c) 'On the Normativity of Human Nature: Some Epistemological Remarks', *Journal of Medicine and Philosophy*, 28, 2: 239–254.

Craig, W.L. (1990) *Divine Foreknowledge and Human Freedom: The Coherence of Theism I: Omniscience*, Leiden: E.J. Brill.

——(1992) 'The Origin and Creation of the Universe: a Reply to Adolf Grünbaum', *British Journal for the Philosophy of Science*, 43, 2: 233–240.

——(1994a) 'Prof. Grünbaum on Creation', *Erkenntnis*, 40, 3: 325–341.

——(1994b) 'A Response to Grünbaum on Creation and Big Bang Cosmology', *Philosophia naturalis*, 31: 237–249.

——(1996a) 'Tense and the New B-Theory of Language', *Philosophy,* 71: 5–26.

——(1996b) 'The New B-Theory's *Tu Quoque* Argument', *Synthese,* 107: 249–269.

——(1996c) 'Timelessness and Creation', *Australasian Journal of Philosophy,* 74, 4: 646–656.

——(1997) 'Hartle–Hawking Cosmology and Atheism', *Analysis,* 57, 4: 291–295.

——(1998a) 'McTaggart's Paradox and the Problem of Temporary Intrinsics', *Analysis,* 58: 122–127.

——(1998b) 'Divine Timelessness and Personhood', *International Journal for Philosophy of Religion*, 43, 2: 109–124.

——(1999a) 'Tensed Time and our Differential Experience of the Past and Future', *Southern Journal of Philosophy,* 37: 515–537.

——(1999b) 'The Presentness of Experience', in M. Wegener (ed.) *Time, Creation, and World Order*, Aarhus: University of Aarhus Press, pp. 107–120.

——(2000a) *The Tensed Theory of Time: A Critical Examination*, Synthese Library, Dordrecht: Kluwer Academic Publishers.

——(2000b) *The Tenseless Theory of Time: A Critical Examination*, Synthese Library, Dordrecht: Kluwer Academic Publishers.

——(2000c) 'On Truth Conditions of Tensed Sentence Types', *Synthese,* 120: 265–270.

——(2000d) 'The Extent of the Present', *International Studies in the Philosophy of Science*, 14: 165–186.

——(2001) 'On Wishing It Were Now Some Other Time', *Philosophy and Phenomenological Research,* 62: 159–166.

——(2002) 'Divine Eternity and the Special Theory of Relativity', in G. Ganssle and D. Woodruff (eds) *God and Time*, New York: Oxford University Press, pp. 129–152.

——(2003) 'In Defense of Presentism', in A. Jokic and Q. Smith (eds) *Time, Tense, and Reference*, Cambridge, MA: MIT Press, pp. 390–408.

——(forthcoming) 'Tooley on Tense and Temporal Priority', *Analysis*.

Crane, T. (2001) 'The Emergence of Consciousness', in Fondazione Carlo Erba (ed.) *The Emergence of the Mind*, Milan: Fondazione Carlo Erba, pp. 183–191.

Damasio, A.R. (1994) *Descartes' Error: Emotion, Reason, and the Human Brain*, New York: Grosset/Putnam.

——(1999) *The Feeling of What Happens: Body and Emotions in the Making of Consciousness*, New York: Harcourt.

Dancy, J. (2000) *Practical Reality*, Oxford: Oxford University Press.

Danto, A. (1967) 'Naturalism', in P. Edwards (ed.) *The Encyclopedia of Philosophy* Vol. 5, London: Routledge, pp. 448–450.

Davidson, D. (1973) 'Freedom to Act', in T. Honderich (ed.) *Essays on Freedom of Action*, London: Routledge & Kegan Paul, pp. 137–156.

——(1980) *Essays on Actions and Events*, Oxford: Clarendon Press.

——(1980a) 'Actions, Reasons, and Causes', in D. Davidson, *Essays on Actions and Events*, Oxford: Clarendon Press, pp. 3–19.

Davies, P.C.W. (1978) 'Spacetime Singularities in Cosmology and Black Hole Evaporation', in J.T. Fraser, N. Lawrence and D. Park (eds) *The Study of Time III*, Berlin: Springer-Verlag, pp. 74–91.

——(1981) *The Edge of Infinity*, New York: Simon & Schuster.

——(1995) 'The Birth of the Cosmos', in J. Gready (ed.) *God, Cosmos, Nature and Creativity*, The Templeton Lectures, University of Sydney, 1991–1994, Edinburgh: Scottish Academic Press, pp. 1–27.

Dennett, D.C. (1991) *Consciousness Explained*, Boston: Little, Brown and Company.

DeRose, K. (1999) 'Solving the Skeptical Problem', in K. DeRose and T. Warfield (eds) *Skepticism: A Contemporary Reader*, New York: Oxford University Press, pp. 183–219.

DeWitt, B.S. (1983) 'Quantum Gravity', *Scientific American* 249, 6, December: 104–115.

Devlin, K. (1997) *Goodbye, Descartes: The End of Logic, and the Search of a New Cosmology of the Mind*, New York: John Wiley and Sons, Ltd.

Dummett, M. (1960) 'A Defense of McTaggart's Proof of the Unreality of Time', *Philosophical Review,* 69: 497–504.

Eddington, A. (1933) *The Expanding Universe*, New York: Macmillan.

Einstein, A. (1952) 'Cosmological Considerations on the General Theory of Relativity', in A. Einstein *et al*., *The Principle of Relativity*, New York: Dover Publications, pp. 177–188.

Evans, G. (1982) *The Varieties of Reference*, Oxford: Clarendon Press.

Frankfurt, H. (1969) 'Alternate Possibilities and Moral Responsibility', *The Journal of Philosophy* 66: 829–839. Reprinted in *The Importance of What We Care About: Philosophical Essays*, Cambridge: Cambridge University Press, 1998, pp. 1–10.

Frede, M. and Patzig, G. (1988) *Aristoteles: Metaphysik Z*, Munich: Beck.

Friedman, A. (1922) 'Über die Krümmung des Raumes', *Zeitschrift für Physik,* 10: 377–386.

Gale, R.M. (1968a) 'The Static versus the Dynamic Temporal: An Introduction', in R.M. Gale (ed.) *The Philosophy of Time: A Collection of Essays*, New Jersey: Humanities Press, pp. 65–85.

——(1968b) *The Language of Time*, London: Routledge and Kegan Paul.

Galvan, S. (1992) 'Inferenza pratica e teoria del volere', in S. Galvan (ed.) *Forme di razionalità pratica*, Milan: Angeli, pp. 189–211.

Gasperini, M. (1999) 'Looking Back in Time Beyond the Big Bang', *Modern Physics,* Letters A 14, 16: 1059–1066.

——(2000) 'Inflation and Initial Conditions in the Pre-Big Bang Scenario', *Physics Review,* D 61: 87301–87305.

Geach, P. (1972), 'Some Problems about Time', in *Logic Matters*, Berkeley, CA: University of California Press, pp. 302–318.

Gill, M.L. (1989) *Aristotle on Substance: The Paradox of Unity*, Princeton, NJ: Princeton University Press.

Goldman, A.I. (1970) *A Theory of Human Action*, Englewood Cliffs, NJ: Prentice-Hall.

——(1994) 'Naturalistic Epistemology and Reliabilism', *Midwest Studies in Philosophy*, 19: 301–320.

Gott, J.R. III (1982) 'Creation of Open Universes from de Sitter space', *Nature,* 295: 304–307.

Gott, J.R. III, Gunn, J.E., Schramm, D.N. and Tinsley, B.M. (1976) 'Will the Universe Expand Forever?' *Scientific American,* 234 (March): 62–79.

Gott, J.R. III and Li-Xin Li (1998) 'Can the Universe Create Itself?' *Physical Review,* D 58 (2): 23501–23544.

Greig, J.Y.T. (ed.) (1932) *The Letters of David Hume*, 2 vols, Oxford: Clarendon Press.

Gribbin, J. (1976) 'Oscillating Universe Bounces Back', *Nature,* 259: 15–16.

——(1986) *In Search of the Big Bang*, New York: Bantam Books.

Grünbaum, A. (1967) 'The Anisotropy of Time', in T. Gold (ed.) *The Nature of Time*, Ithaca, NY: Cornell University Press, pp. 149–186.

——(1989) 'The Pseudo-Problem of Creation in Physical Cosmology', *Philosophy of Science,* 56: 373–394.

——(1991) 'Creation as a Pseudo-Explanation in Current Physical Cosmology', *Erkenntnis,* 35: 233–254.

——(1994) 'Some Comments on William Craig's "Creation and Big Bang Cosmology"', *Philosophia naturalis,* 31(2): 225–236.

——(2000) 'A New Critique of Theological Interpretations of Physical Cosmology', *British Journal for the Philosophy of Science*, 51, 1: 1–43.

Guth, A. (1981) 'Inflationary Universe: A Possible Solution to the Horizon and Flatness Problems', *Physical Review,* D 23: 247–256.

——(1997) *The Inflationary Universe: The Quest for a New Theory of Cosmic Origins*, Reading, MA: Addison-Wesley.

Guth, A. and Sher, M. (1983) 'The Impossibility of a Bouncing Universe', *Nature,* 302: 505–506.

Harman, G. (1986) 'Moral Explanations of Natural Facts – Can Moral Claims Be Tested Against Moral Reality?', *Southern Journal of Philosophy*, supplementary vol. 26: 57–68.

Harré, R. and Madden, E.H. (1975) *Causal Powers: A Theory of Natural Necessity*, Oxford: Blackwell.

Hartle, J. and Hawking, S. (1983) 'Wave Function of the Universe', *Physical Review*, D 28: 2960–2975.

Hasker, W. (1992) 'Evolution and Alvin Plantinga', *Perspectives on Science and Christian Faith,* 44: 150–162.

Hawking, S. (1976) 'Breakdown of Predictability in Gravitational Collapse', *Physical Review,* D 14: 2460–2473.

——(1988) *A Brief History of Time*, New York: Bantam Books.

Hawking, S.W. and Penrose, R. (1973) 'Space-Time Singularities', in S.W. Hacking and G.F. R. Ellis (eds) *The Large-Scale Structure of Space-Time*, Cambridge: Cambridge University Press, pp. 256–298.

——(1996) *The Nature of Space and Time*, The Isaac Newton Institute Series of Lectures, Princeton, NJ: Princeton University Press.

Hick, J. (1960) 'God as Necessary Being', *Journal of Philosophy,* 57, 22/23: 725–734.

Hinchliff, M. (1994) 'The Puzzle of Change', paper presented at the Pacific Division meeting of the American Philosophical Association, 2 April.

Hobart, R.E. (1934) 'Free Will as Involving Determinism and Inconceivable Without It', *Mind,* 43: 1–27.

Hochberg, D., Molina-Paris, C. and Visser, M. (1999) 'Tolman Wormholes Violate the Strong Energy Condition', *Physical Review,* D 59: 44011–44019.

Hogarth, W. ([1753] 1955) *The Analysis of Beauty*, Oxford: Clarendon Press.

Hooker, B. (2000) *Ideal Code, Real World*, Oxford: Oxford University Press.

Hoyle, F. (1948) 'A New Model for the Expanding Universe', *Monthly Notices of the Royal Astronomical Society,* 108: 372–382.

——(1975a) *Astronomy Today*, London: Heinemann.

——(1975b) *Astronomy and Cosmology: A Modern Course*, San Francisco: W.H. Freeman.

Hubble, E. (1929) 'A Relation between Distance and Radial Velocity among Extra-galactic Nebulae', *Proceedings of the National Academy of Science,* 15: 168–173.

Hubin, D.C. (2001) 'The Groundless Normativity of Instrumental Rationality', *Journal of Philosophy,* 98, 9: 445–468.

Humboldt, W. von (1960) *Werke in fünf Bänden*, ed. A. Flitner and K. Giel, Darmstadt 1960–81: Wissenschaftliche Buchgesellschaft.

Hume, D. (1947) *Dialogues Concerning Natural Religion*, ed. with an Introduction by N. Kemp Smith, Indianapolis: Bobbs-Merrill.

——(1951) *Treatise of Human Nature*, ed. L.A. Selby-Bigge, Oxford: Clarendon Press.

——(1955) *An Inquiry Concerning Human Understanding*, New York: The Liberal Arts Press.

Humphrey, N. (2000) 'The Privatization of Sensation', in L. Huber and C. Heyes (eds) *The Evolution of Cognition,* Cambridge, MA: MIT Press.

Husserl, E. (1970) *The Crisis of European Sciences and Transcendental Phenomenology: An Introduction to Phenomenological Philosophy*, trans. by D. Carr, Evanston, IL: Northwestern University Press.

Idziak, J.M., Boyd, C.A. and Van Arragon, R.J. (2003) 'Is Morality Based on God's Commands?', in M.L. Peterson and R.J. Van Arragon (eds) *Contemporary Debates in Philosophy of Religion*, Oxford: Blackwell, pp. 290–314.

Irwin, T. (1988) *Aristotle's First Principles*, Oxford: Clarendon Press.

Isham, C. (1988) 'Creation of the Universe as a Quantum Process', in R.J. Russell, W.R. Stoeger and G.V. Coyne (eds) *Physics, Philosophy and Theology: A Common Quest for Understandin*g, Vatican City: Vatican Observatory, pp. 375–408.

——(1990) 'Space, Time, and Quantum Cosmology', paper presented at the conference 'God, Time, and Modern Physics', March.

——(1993–1996) 'Quantum Theories of the Creation of the Universe', in R.J. Russell, N. Murphey and C. Isham (eds) *Quantum Cosmology and the Laws of Nature*, Vatican City: Vatican Observatory, pp. 51–89.

——(1994) 'Quantum Cosmology and the Origin of the Universe', lecture presented at the conference 'Cosmos and Creation', Cambridge University, 14 July.

Jaki, S.L. (1974) *Science and Creation,* Edinburgh: Scottish Academic Press.

Kane, R. (1996) *The Significance of Free Will*, New York: Oxford University Press.

Kanitscheider, B. (1990) 'Does Physical Cosmology Transcend the Limits of Naturalistic Reasoning?' in P. Weingartner and G.J.W. Doen (eds) *Studies on Mario Bunge's 'Treatise',* Amsterdam: Rodopi, pp. 337–350.

——(2004) 'Heuristische Ideen an den Grenzen kosmologischen Wissens', in W. Löffler and P. Weingartner (eds) *Knowledge and Belief. Proceedings of the 26th International Wittgenstein Symposium*, Kirchberg (Austria) 2003, Vienna: öbv-hpt, pp. 301–315.

Kant, I. (1956) *Critique of Practical Reason*, trans. L. White Beck, Indianapolis, IN: Bobbs-Merrill.

——(1998) *Critique of Pure Reason*, trans. P. Guyer and A.W. Wood, Cambridge: Cambridge University Press.

Kenny, A. (1975) *Will, Freedom and Power*, Oxford: Basil Blackwell.

Keynes, J.M. (1921) *A Treatise on Probability*, London: Macmillan.

Kline, M. (1972) *Mathematical Thought: From Ancient to Modern Times*, vol. 1, Oxford: Oxford University Press.

Koppelberg, D. (1999) 'What Is Naturalism Today – And What Should It Be Tomorrow?', in D. Marconi (ed.) *Naturalismo e naturalizzazione*, Vercelli: Edizioni Mercurio, pp. 23–45.

Korsgaard, C.M. (1996) *The Sources of Normativity*, Cambridge: Cambridge University Press.

——(1997) 'The Normativity of Instrumental Reason', in G. Cullity and B. Gaut (eds) *Ethics and Practical Reason*, Oxford: Oxford University Press, pp. 215–254.

Kvanvig, J.L. (ed.) (1996) *Warrant in Contemporary Epistemology: Essays in Honor of Plantinga's Theory of Knowledge*, Lanham, MD: Rowman & Littlefield.

Leftow, B. (1991) *Time and Eternity,* Cornell Studies in Philosophy of Religion, Ithaca, NY: Cornell University Press.

Leibniz, G.W. (1949) *New Essays on Human Understanding*, trans. A.G. Langley, LaSalle, IL: Open Court Publishing Company.

——(1951) *The Monadology and Other Philosophical Writings*, trans. R. Latta, London: Oxford University Press.

——(1985) *Theodicy*, trans. E.M. Huggard, ed. A. Farrer, La Salle, IL: Open Court.

Lehrer, K. (1996) 'Proper Function versus Systematic Coherence', in J.L. Kvanvig (ed.) *Warrant in Contemporary Epistemology: Essays in Honor of Plantinga's Theory of Knowledge*, Lanham, MD: Rowman & Littlefield, pp. 25–45.

Lemaître, G. (1927) 'Un univers homogène de masse constante et de rayon croissant, rendant compte de la vitesse radiale des nébuleuses extragalactiques', *Annales de la Société scientifique de Bruxelles,* 47: 49–59.

Lenzen, W. (1980) *Glauben, Wissen und Wahrscheinlichkeit*, Vienna: Springer.

Leslie, J. (1993) 'Is It All Quite Simple?', *Times Literary Supplement,* 29 January: 3–4.

Lewis, D. (1983) 'New Work for a Theory of Universals', *Australasian Journal of Philosophy*, 61: 343–377.

——(1986) 'Persons, Morality, and Tenselessness', *Philosophy and Phenomenological Research*, 47: 305–309.

Lifschitz, E.M. and Khalatnikov, I.M. (1963) 'Investigations in Relativist Cosmology', *Advances in Physics,* 12: 185–249.

Linde, A.D. (1983) 'Chaotic Inflation', *Physics Letters,* 129: 177–181.

——(1984) 'The Inflationary Universe', *Reports on Progress in Physics,* 47: 925–986.

——(2002) 'Cyclic Universe Runs into Criticism', *Physics World,* June, 8.

Linde, A., Linde, D. and Mezhlumian, A. (1994) 'From the Big Bang Theory to the Theory of a Stationary Universe', *Physical Review,* D 49: 1783–1826.

Locke, J. (1961) *An Essay Concerning Human Understanding*, New York: E.P. Dutton & Co.

Löffler, W. (1999) 'Naturalisierungsprogramme und ihre methodologischen Grenzen', in J. Quitterer and E. Runggaldier (eds) *Der neue Naturalismus. Eine Herausforderung an das christliche Menschenbild*, Stuttgart: Kohlhammer, pp. 30–76.

Loizou, A. (1986) *The Reality of Time*, Brookfield, Vermont: Gower.

Loux, M.J. (1991) *Primary Ousia*, Ithaca, NY: Cornell University Press.

Lowe, J. (2002) *A Survey of Metaphysics*, Oxford: Oxford University Press.

Mackie, J.L. (1977) *Ethics: Inventing Right and Wrong*, Harmondsworth: Penguin Books.

Malcolm, N. (1958) *Ludwig Wittgenstein: A Memoir*, London: Oxford University Press.
Mally, E. (1914) 'Über die Unabhängigkeit der Gegenstände vom Denken', *Zeitschrift für Philosophie und philosophische Kritik*, 155: 37–52.
Manson, N.A. (ed.) (2003) *God and Design: The Teleological Argument and Modern Science*, London and New York: Routledge.
McCann, H. (1998) *The Works of Agency*, Ithaca, NY: Cornell University Press.
McGilvray, J. (1979) 'A Defense of Physical Becoming', *Erkenntnis,* 14: 275–299.
McGinn, C. (1983) *The Subjective View*, Oxford: Clarendon Press.
Meinong, A. ([1904] 1971) *Über Gegenstandstheorie*, in *Gesamtausgabe*, vol. II, *Abhandlungen zur Erkenntnistheorie und Gegenstandstheorie*, Graz: Akademische Druck- und Verlagsanstalt.
——([1917] 1968) *Über emotionale Präsentation*, in *Gesamtausgabe*, vol. III, Graz: Akademische Druck- und Verlagsanstalt.
Meixner, U. (2003) 'Die Aktualität Husserls für die moderne Philosophie des Geistes', in U. Meixner and A. Newen (eds) *Seele, Denken, Bewusstsein*, Berlin: De Gruyter.
——(2004) *The Two Sides of Being: A Reassessment of Psycho-Physical Dualism*, Paderborn: Mentis.
Melden, A.I. (1961) *Free Action*, London: Routledge & Kegan Paul.
Mellor, D.H. (1981) *Real Time*, Cambridge: Cambridge University Press.
——(1991) *Matters of Metaphysics*, Cambridge: Cambridge University Press.
——(2003) 'Too Many Universes', in N.A. Manson (ed.) *God and Design: The Teleological Argument and Modern Science*, London and New York: Routledge, pp. 221–228.
Merricks, T. (1994) 'Endurance and Indiscernibility', *Journal of Philosophy,* 91: 165–184.
Mill, J.S. (1979) *Utilitarianism*, Indianapolis, IN: Hackett.
Misner, C.W., Thorne, K. and Wheeler, J. (1973) *Gravitation*, San Francisco: W.H. Freeman.
Moore, G.E. (1912) *Ethics,* Oxford: Oxford University Press.
——(1922) *Philosophical Studies*, Boston: Routledge & Kegan Paul.
——(1942) 'A Reply to My Critics', in: P. Schilpp (ed.) *The Philosophy of G.E. Moore*, La Salle, IL: Open Court, pp. 533–677.
Moreland, J.P. (1994) 'Theistic Science and Methodological Naturalism', in J.P. Moreland (ed.) *The Creation Hypothesis,* Downer's Grove, IL: Inter-Varsity Press, pp. 41–66.
——(1998) 'Libertarian Agency and the Craig/Grünbaum Debate about Theistic Explanation of the Initial Singularity', *American Catholic Philosophical Quarterly,* 81: 539–554.
Moreland, J.P., Meyer, S.C. and Bube, R.H. (1994) 'Conceptual Problems and the Scientific Status of Creation Science: A Discussion', *Perspectives on Science and Christian Faith,* 46: 2–25.
Naber, G.L. (1988) *Spacetime and Singularities: An Introduction*, Cambridge: Cambridge University Press.
Nelkin, N. (1996) *Consciousness and the Origins of Thought,* Cambridge: Cambridge University Press.
Novikov, I.D. and Zeldovich, Y.B. (1973) 'Physical Processes near Cosmological Singularities', *Annual Review of Astronomy and Astrophysics,* 11: 387–412.
O'Connor, T. (2000) *Persons and Causes: The Metaphysics of Free Will*, Oxford: Oxford University Press.
Papineau, D. (1993) *Philosophical Naturalism*, Oxford: Blackwell.
Parfit, D. (1998) 'Why Anything? Why This?' *London Review of Book,* 20, 2: 24–27.
Penrose, R. (1965) 'Gravitational Collapse and Space-Time Singularities', *Physical Review Letters,* 14: 57–59.
——(1981) 'Time-Asymmetry and Quantum Gravity', in C.G. Isham, R. Penrose and D.W. Sciama (eds) *Quantum Gravity 2*, Oxford: Clarendon Press: 245–272.

——(1982) 'Some Remarks on Gravity and Quantum Mechanics', in M.J. Duff and C.J. Isham (eds) *Quantum Structure of Space and Time*, Cambridge: Cambridge University Press, 3–10.

——(1997) *The Large, the Small, and the Human,* Cambridge: Cambridge University Press.

Plantinga, A. (1974) *The Nature of Necessity*, Oxford: Clarendon Press.

——(1992) 'On Rejecting The Theory of Common Ancestry: A Reply to Hasker', *Perspectives on Science and Christian Faith,* 44: 258–263.

——(1993a) *Warrant: The Current Debate*, Oxford: Oxford University Press.

——(1993b) *Warrant and Proper Function*, Oxford: Oxford University Press.

——(1996a) 'Methodological Naturalism?', in J. van der Meer (ed.) *Facets of Faith and Science*, Lanham, MD: University Press of America. Online available: http://id-www.ucsb.edu/fscf/library/plantinga/mn/home.html

——(1996b) 'Respondeo', in L. Kvanvig (ed.) *Warrant in Contemporary Epistemology: Essays in Honor of Plantinga's Theory of Knowledge*, Lanham, MD: Rowman & Littlefield, pp. 307–378.

——(2002) 'Reply to Beilby's Cohorts', in J. Beilby (ed.) *Naturalism Defeated? Essays on Plantinga's Evolutionary Argument Against Naturalism,* Ithaca, NY: Cornell University Press, pp. 205–225.

Plantinga, A., Van Till, H.J., Pun, P., and McMullin, E. (1991) 'Symposium: Evolution and the Bible', *Christian Scholar's Review,* 21: 8–109.

Post, J. (1991) *Metaphysics: A Contemporary Introduction*, New York: Paragon House.

Prior, A.N. (1968) 'Changes in Events and Changes in Things', in *Papers on Time and Tense*, Oxford: Clarendon Press, pp. 1–14.

Ramsey, J. (2002) 'Naturalism Defended', in J. Beilby (ed.) *Naturalism Defeated? Essays on Plantinga's Evolutionary Argument Against Naturalism*, Ithaca, NY: Cornell University Press, pp. 15–29.

Rea, M. (2000) 'Naturalism and Material Objects', in W.L. Craig and J.P. Moreland (eds) *Naturalism: A Critical Analysis*, London: Routledge, pp. 110–132.

Rees, M. (1997) *Before the Beginning,* with a Foreword by S. Hawking, Reading, MA: Addison-Wesley.

Reeves, H., Audouze, J., Fowler, W.A. and Schramm, D.N. (1973) 'On the Origin of Light Elements', *Astrophysical Journal*, 179: 909–930.

Reid, T. (1969) *Essays on the Active Powers of the Human Mind*, Cambridge, MA: The MIT Press.

——(1983) *Essays on the Intellectual Powers of Man* in *Thomas Reid's Inquiry and Essays*, ed. R. Beanblossom and K. Lehrer, Indianapolis, IN: Hackett Publishing Co.

Rottschaefer, W.A. (1998) *The Biology and Psychology of Moral Agency*, Cambridge: Cambridge University Press.

Runggaldier, E. (1998) 'Sortal Continuity of Material Things', *Erkenntnis,* 48: 359–369.

Russell, B. and Copleston, F.C. (1964) 'The Existence of God', in J. Hick (ed.) *The Existence of God*, New York: Macmillan, pp. 167–191.

Sartre, J.P. (1943) *L'Être et le néant: Essai d'ontologie phénoménologique*, Paris: Gallimard.

Schlesinger, G. (1980) *Aspects of Time*, Indianapolis: Hackett.

Schlick, M. (1939) 'When is a Man Responsible?', in *Problems of Ethics*, trans. by D. Rynin, Englewood Cliffs, NJ: Prentice-Hall, pp. 143–156.

Searle, J.R. (2001) *Rationality in Action*, Cambridge, MA: MIT Press.

Senor, T.D. (1993) 'Divine Temporality and Creation *ex nihilo*', *Faith and Philosophy,* 10: 86–92.

Shafer-Landau, R. (2003) *Moral Realism*, Oxford: Oxford University Press.

Shoemaker, S. (1969) 'Time without Change', *Journal of Philosophy,* 66: 363–381.
Silk, J. (1989) *The Big Bang*, 2nd edn, San Francisco: W.H. Freeman.
Smith, Q. (1986) *The Felt Meanings of the World,* Lafayette, IN: Purdue University Press.
——(1991) *Language and Time*, New York: Oxford University Press.
——(1993a) 'The Uncaused Beginning of the Universe', in W.L. Craig and Q. Smith (eds) *Theism, Atheism, and Big Bang Cosmology*, Oxford: Clarendon Press, pp. 108–140.
——(1993b) 'The Concept of a Cause of the Universe', *Canadian Journal of Philosophy,* 23: 1–24.
——(1996) 'Causation and the Logical Impossibility of a Divine Cause', *Philosophical Topics,* 24: 169–191.
——(1997) 'Simplicity and Why the Universe Exists', *Philosophy,* 71: 125–132.
Strawson, P.F. (1985) *Skepticism and Naturalism*, New York: Columbia University Press.
Sturgeon, N. (1988) 'Moral Explanations', in G. Sayre-McCord (ed.) *Essays in Moral Realism*, Ithaca, NY: Cornell University Press, pp. 229–255.
Suárez, F. (1861) *Disputationes metaphysicae*, ed. C. Berton, vols XXV and XXVI, Paris.
——(2002) *On Creation, Conservation, and Concurrence: 'Metaphysical Disputations 20, 21 and 22'*, trans. with an Introduction and Notes by A.J. Freddoso, South Bend, IN: St Augustine's Press.
Swinburne, R. (1991) *The Existence of God*, rev. edn, Oxford: Clarendon Press.
Tarski, A. (1949) 'The Semantic Conception of Truth and the Foundations of Semantics', in H. Feigl and W. Sellars (eds) *Readings in Philosophical Analysis*, New York: Appleton-Century-Crofts, pp. 52–84.
Tegmark, M. (1998) 'Is "the Theory of Everything" Merely the Ultimate Ensemble Theory?', *Annals of Physics*, 270: 1–51.
——(2003a) 'Parallel Universes', *Scientific American,* May 2003: 40–51.
——(2003b) 'Parallel Universes,' at www.arXiv:astro-ph/0302131 v1, accessed 7 February 2003. Also in J.D. Barrow, P.C.W. Davies and C.L. Harper (eds) *Science and Ultimate Reality: From Quantum to Cosmos*, Cambridge: Cambridge University Press, pp. 459–491.
Tryon, E. (1973) 'Is the Universe a Vacuum Fluctuation?' *Nature,* 246: 396–397.
van Fraassen, B. (1989) *Laws and Symmetry*, Oxford: Clarendon Press.
——(1996) 'Science, Materialism, and False Consciousness', in J.L. Kvanvig (ed.) *Warrant in Contemporary Epistemology: Essays in Honor of Plantinga's Theory of Knowledge*, Lanham, MD: Rowman & Littlefield, pp. 149–181.
van Inwagen, P. (1975) 'The Incompatibility of Free Will and Determinism', *Philosophical Studies,* 27: 185–199.
——(1990) 'Four Dimensional Objects', *Noûs*, 24: 245–255.
——(2002) 'Free Will Remains a Mystery', in R. Kane (ed.) *The Oxford Handbook of Free Will*, Oxford: Oxford University Press, pp. 158–177.
Van Roojen, M. (1996) 'Expressivism and Irrationality', *Philosophical Review*, 105: 311–333.
Vilenkin, A. (1982) 'Creation of the Universe from Nothing', *Physical Letters*, B 117: 25–28.
——(1983) 'Birth of Inflationary Universes', *Physical Review*, D 27: 2848–2855.
von Kutschera, F. (1994) 'Global Supervenience and Belief', *Journal of Philosophical Logic*, 23: 103–110.
——(1999) *Grundlagen der Ethik*, 2nd edn, Berlin: W. de Gruyter.
——(2001) 'Concepts of a set', in A. Newen, U. Nortmann and R. Stuhlmann-Laeisz (eds) *Building on Frege*, Stanford, CA: CSLI, pp. 319–327.
——(2003) *Jenseits des Materialismus*, Paderborn: Mentis.
von Wright, G.H. (1971) *Explanation and Understanding*, London: Routledge & Kegan Paul.
Vries, P. de (1986) 'Naturalism in the Natural Sciences: A Christian Perspective', *Christian Scholar's Review* 388–396.

Ward, K. (1996) *God, Chance, and Necessity*, Oxford: One World.

Weatherford, R. (1983) *Philosophical Foundations of Probability Theory*, London: Routledge and Kegan Paul.

Weinberg, S. (1992) *Dreams of a Final Theory,* New York: Vintage Books.

Weingard, R. (1982) 'Do Virtual Particles Exist?', in P. Asquith and T. Nichols (eds) *Proceedings of the Philosophy of Science Association,* 2 vols, East Lansing, MI: Philosophy of Science Association, vol. 1, pp. 235–242.

Wheeler, J.A. (1980) 'Beyond the Hole', in H. Woolf (ed.) *Some Strangeness in the Proportion: A Centennial Symposium to Celebrate the Achievements of Albert Einstein*, Reading, MA: Addison-Wesley, pp. 341–375.

Widerker, D. (1995) 'Libertarianism and Frankfurt's Attack on the Principle of Alternative Possibilities', *The Philosophical Review,* 104: 247–261.

Widerker, D. and McKenna, M. (eds) (2003) *Freedom, Responsibility and Agency: Essays on the Importance of Alternative Possibilities*, Aldershot: Ashgate Press.

Williams, B. (1981) 'Internal and External Reasons', in *Moral Luck*, Cambridge: Cambridge University Press, pp. 101–113.

Wittgenstein, L. (1974) *On Certainty,* ed. G.E.M. Anscombe and G.H. von Wright, Oxford: Blackwell.

Wooldridge, M. (1999) 'Intelligent Agents', in G. Weiss (ed.) *Multiagent Systems*, Cambridge, MA: MIT Press.

Yates, J.C. (1990) *The Timelessness of God*, Lanham, MD: University Press of America.

Index