Real Metaphysics

Essays in honour of D. H. Mellor

**Edited by Hallvard Lillehammer and
Gonzalo Rodriguez-Pereyra**

Real Metaphysics

Does time flow? Do the past and future exist? What are facts? What is causation? Do truths have truthmakers? If so, what are they? The chapters in this collection offer new answers to these fundamental questions, which have preoccupied Hugh Mellor, one of the outstanding metaphysicians of our time and author of titles including *The Matter of Chance* (1971), *Matters of Metaphysics* (1991) and *The Facts of Causation* (Routledge 1995).

Real Metaphysics brings together new articles by leading metaphysicians to honour Mellor's contribution to metaphysics. Some of the most outstanding minds of current times shed new light on all the main topics in metaphysics: truth, causation, dispositions, properties, explanation, and time. At the end of the book, Hugh Mellor responds to the issues raised in each of the fourteen contributions and gives us new insight into his own highly influential work on metaphysics.

Real Metaphysics stands as a highly original exploration and assessment of some of the most central issues in metaphysics, and will make fascinating reading for anyone interested in contemporary philosophy.

Contributors: David Armstrong, Alexander Bird, Tim Crane, Chris Daly, Frank Jackson, Arnold Koslow, Isaac Levi, Hallvard Lillehammer, David Lewis, Hugh Mellor, Peter Menzies, Paul Noordhof, L. Nathan Oaklander, Gonzalo Rodriguez-Pereyra, Gideon Rosen, Peter Smith.

Routledge studies in twentieth-century philosophy

Real Metaphysics

Essays in honour of D. H. Mellor

Edited by Hallvard Lillehammer and Gonzalo Rodriguez-Pereyra

Contents

Contributors

David Armstrong is Emeritus Professor of Philosophy at Sydney University. He is the author of, among other books, *A Materialist Theory of the Mind* (Routledge & Kegan Paul 1968) and *A World of States of Affairs* (Cambridge University Press 1997). He currently works in metaphysics and truthmaker theory.

Alexander Bird is Reader in Philosophy at the University of Edinburgh. He is the author of *Philosophy of Science* (UCL 1998) and *Thomas Kuhn* (Acumen 2000).

Tim Crane is Professor in Philosophy at University College London and Director of the Philosophy Programme of the School of Advanced Study, University of London. He is the author of *The Mechanical Mind* (Penguin 1995; 2nd edn, Routledge 2003) and *Elements of Mind* (Oxford University Press 2001).

Chris Daly is Lecturer in Philosophy at Manchester University. He works mainly in metaphysics.

Frank Jackson is Professor of Philosophy, Research School of Social Sciences, The Australian National University. He has held positions at the University of Adelaide, La Trobe University and Monash University and a number of visiting appointments outside Australia. He is the author of, among other books, *Perception* (Cambridge University Press 1977) and *From Metaphysics to Ethics* (Oxford University Press 1998).

Arnold Koslow is Professor of Philosophy at the Graduate Center, City University of New York. His recent work has focused mainly on logic, the philosophy of mathematics and the philosophy of science. He is the author of *A Structuralist Theory of Logic* (Cambridge University Press 1992).

Isaac Levi is John Dewey Professor of Philosophy at Columbia University. He has written extensively on topics concerning decision-making and changes in

beliefs and values from a pragmatist standpoint. He is the author of, among other books, *The Fixation of Belief and its Undoing* (Cambridge University Press 1991) and *The Covenant of Reason* (Cambridge University Press 1997).

David Lewis (deceased) was Class of 1943 University Professor of Philosophy at Princeton University. He wrote extensively on metaphysics, philosophy of language, philosophy of mind and epistemology. He was the author of, among other books, *Counterfactuals* (Basil Blackwell 1973) and *On the Plurality of Worlds* (Basil Blackwell 1986).

Hallvard Lillehammer is Fellow of King's College and University Assistant Lecturer in the Faculty of Philosophy at Cambridge University. He works mainly in ethics.

D. H. Mellor is Emeritus Professor of Philosophy at Cambridge University. He is the author of, among other books, *The Facts of Causation* (Routledge 1995) and *Real Time II* (Routledge 1998).

Peter Menzies is Associate Professor and Head of the Philosophy Department, Macquarie University, Sydney. Much of his published research centres on the subject of causation.

Paul Noordhof is Lecturer in Philosophy at the University of Nottingham. His book, *A Variety of Causes*, is forthcoming from Oxford University Press.

L. Nathan Oaklander is Professor and Chair of the Philosophy Department at the University of Michigan-Flint. He is the author of *Temporal Relations and Temporal Becoming* (University Press of America 1984) and (with Quentin Smith) *Time, Change and Freedom* (Routledge 1995).

Gonzalo Rodriguez-Pereyra is Gilbert Ryle Fellow of Hertford College and Lecturer in the Faculty of Philosophy at Oxford University. He is the author of *Resemblance Nominalism* (Oxford University Press 2002).

Gideon Rosen is Associate Professor of Philosophy at Princeton University. He is the author (with John Burgess) of *A Subject with No Object* (Oxford University Press 1997).

Peter Smith is a Fellow of Jesus College and Lecturer in the Faculty of Philosophy at Cambridge University. He was editor of *Analysis* for 12 years, and is the author of *Explaining Chaos* (Cambridge University Press 1998) and (with O. R. Jones) *The Philosophy of Mind* (Cambridge University Press 1986).

Introduction

Hallvard Lillehammer and Gonzalo Rodriguez-Pereyra

The following chapters, all previously unpublished, have been written in honour of Hugh Mellor, who has recently retired as Professor of Philosophy at the University of Cambridge. The chapters are all concerned with metaphysical topics about which Mellor has written. They are followed by Mellor's replies.

Hugh Mellor was born in London on 10 July 1938. He read Natural Sciences and Chemical Engineering at Pembroke College, Cambridge, from 1956 to 1960. His first formal study of philosophy was in the United States, where, on a Harkness Fellowship (1960–2), he studied for an MSc degree in Chemical Engineering at the University of Minnesota. While there, he completed a Minor in Philosophy of Science under Herbert Feigl. Later, after a year working for ICI as a chemical engineer, he returned to Pembroke in 1963 as a PhD student, supervised by Mary Hesse, and submitted his thesis, 'The Matter of Chance', in 1968. He became a Research Fellow of Pembroke in 1964, and University Assistant Lecturer in the Faculty of Philosophy in 1965. In 1986 he was elected Professor of Philosophy, a chair from which he retired in 1999. After his retirement he was University Pro-Vice-Chancellor for Research for two years, 2000–1. He has been a Fellow of Darwin College, Cambridge, since 1970, and became a Fellow of the British Academy in 1983.

Hugh Mellor's contribution to philosophy is rich, varied and original. In the early 1980s, with the publication of *Real Time* (Mellor 1981), he revived the debate on McTaggart's paradox about the A-series and became one of the main defenders of the B-theory of time. He revised and refined his position in *Real Time II* (Mellor 1998). He has also developed an original theory of causation (Mellor 1995), according to which causes raise the chances of their effects and facts can be causes and effects. In addition, he has produced original and influential work on dispositions, laws, properties and mind. Probability has been another continuous interest. *The Matter of Chance* (Mellor 1971) was his first published book, and he is currently co-writing a textbook on probability with Arnold Koslow.

In all his writings Mellor has remained faithful to the Cambridge tradition of straight thinking, clear writing and sharp argument. Mellor is a combative philosopher, and this feature is present in the replies to the chapters in the present volume. Yet Mellor does not pursue philosophical combat for its

own sake, but as a means in the pursuit of truth (his main contributions to philosophy are *positive*: a theory of time, a theory of causation, etc.).

Mellor's contribution to philosophy does not merely consist in his published work. During his years at Cambridge he contributed to the moulding of dozens of philosophers, many of whom have gone on to distinguished academic careers. Those who have been supervised by him know how supportive and dedicated a supervisor he is.

The following chapters honour Mellor's work in metaphysics. Here we offer a brief summary of their contents and Mellor's replies.

Truth, truthmaking and success

Truth has been a permanent topic of interest among philosophers. What is truth? Is it a property? Is there anything more to truth than the T-biconditionals? Does it consist in the correspondence of a proposition – or another truthbearer – with a fact? Do truths require truthmakers? These are some of the questions about truth that concern philosophers.

Since the late 1980s there has been a considerable interest in the notion of truthmakers – that in virtue of which a truth is true, that which makes a truth true – and the question of whether truths have truthmakers. In Mellor's work truthmakers occupy a central position. One of the main differences between *Real Time II* and its predecessor, *Real Time,* is precisely that *Real Time II* explains Mellor's tenseless theory of time in terms of truthmakers.

There are some areas of consensus in truthmaker research, for instance that e is the truthmaker of $<e$ exists$>$, and that the truthmakers of a conjunction are the truthmakers of its conjuncts. It also seems to be part of the consensus that truth-functions have truthmakers. What the truthmakers are of (a) simple predications, (b) negative truths, (c) universal generalizations and (d) modal truths is more controversial. In his chapter, David Armstrong advances two principles of truthmaker theory:

(1) Truthmakers necessitate the truths they truthmake (Truthmaker Necessitarianism).
(2) Every truth has a truthmaker (Truthmaker Maximalism).

He then goes on to propose truthmakers for modal truths that are consistent with the actualist character of his ontology, namely the idea that only what is actual exists and so there are no mere possible worlds and things. With the help of what he calls the Entailment Principle, namely that a truthmaker for a contingent truth is also a truthmaker for any truth entailed by the former truth, Armstrong argues that the truthmaker for the proposition that $<$not-p is possible$>$, where p is contingent, is the truthmaker for p. He also gives truthmakers for necessary truths, truths about alien properties and other modal truths. An important characteristic of Armstrong's truthmakers is that they are actual entities, thereby ensuring that modal truths by themselves do

not impose any sort of ontological inflation. (Armstrong's ontology is not the most economical. It contemplates entities such as states of affairs in general and, in particular, totality states of affairs, e.g. the state of affairs that, say, *a*, *b* and *c* are all the red objects in this room.)

Mellor's reply is a testimony to the controversy that surrounds truthmaker theory. First, he denies that necessary truths have truthmakers. This idea has been maintained by others and is often aired in discussion. But Mellor's attack on Truthmaker Maximalism goes beyond this. He also denies that some contingent truths have truthmakers. In particular, Mellor maintains that only *atomic* propositions have truthmakers. It follows that negative truths have no truthmakers. This allows Mellor to argue against Truthmaker Necessitarianism. Consider a universal generalization such as 'everything is **F**' and suppose *a* and *b* are the only two things there are. If truthmakers necessitate their truths then more than the truthmakers of '**F***a*' and '**F***b*' are needed to make 'everything is **F**' true. For since the truthmakers of '**F***a*' and '**F***b*' do not exclude the presence of something that is neither *a* nor *b*, they do not necessitate the truth of 'everything is **F**'. That there is nothing that is neither *a* nor *b* is a negative truth which does not need and does not have a truthmaker. So nothing necessitates the truth of 'everything is **F**', although whatever makes true '**F***a*' and '**F***b*' also makes true 'everything is **F**'. Mellor's rejection of Truthmaker Necessitarianism and Truthmaker Maximalism constitutes an important claim that allows him to bypass the notorious problems of finding truthmakers for negative truths and universal generalizations.

The controversy about the truthmakers of simple predications centres around the ontology of those truthmakers. Some think that tropes are truthmakers, whereas others think that facts play this role. Among those who postulate facts as truthmakers, some (like Mellor, who calls his truthmakers *facta*) take them to be constituted by particulars and universals, whereas others (resemblance nominalists) take them to be constituted by only resembling particulars. David Lewis, who used to deny that truths have truthmakers, now thinks that they do and, moreover, thinks that truthmakers are neither facts nor tropes, but are the ordinary particulars true propositions or sentences are about.

Consider a black cat called 'Long'. Then 'Long is black' is true. Normally, philosophers think that particulars, like Long, cannot be the truthmakers of such predications. The usual reason for this is that a cat like Long may fail to necessitate the truth of 'Long is black' (Armstrong 1997: 115). A further reason is that this answer makes Long a truthmaker not only of 'Long is black', but also of 'Long is hairy', 'Long is small', and so on. But surely Long is black in virtue of something different from that in virtue of which it is small or hairy. Thus 'Long is black', 'Long is hairy', 'Long is small', and so on, should have different truthmakers (Rodriguez-Pereyra 2000: 268–9).

Nevertheless, Lewis thinks that Long is the truthmaker of truths such as 'Long is black'. Lewis invokes his counterpart theory, according to which a thing *a* is possibly **F** just in case there is something else, *b*, in a different possible

world, that sufficiently resembles *a*, and *b* is **F**. But there are many different counterpart relations, so *b* might be a counterpart of *a* under one counterpart relation but not under another. Take a counterpart relation under which all of Long's counterparts are black. If we think of Long under this counterpart relation, we think of Long as essentially black. *Long qua black* is a name for Long that evokes this counterpart relation for Long. Similarly, *Long qua small* is a name for Long that evokes a counterpart relation under which Long is essentially small, i.e. a counterpart relation under which all of its counterparts are small. Thus Long qua black is the truthmaker of 'Long is black' and Long qua small is the truthmaker of 'Long is small'. No doubt Long qua black necessitates the truth of 'Long is black' and Long qua small necessitates the truth of 'Long is small'. But Long = Long qua black = Long qua small. This is how Long can be the truthmaker of predications like 'Long is black' and 'Long is small'. In a postscript to Lewis's chapter, Lewis and Rosen extend this idea to account for the truthmakers of negative existentials such as 'there are no unicorns'. The thought is to make the world qua lacking unicorns the truthmaker of 'There are no unicorns'. Lewis also criticizes Mellor's refusal to admit temporal parts, and his consequent ontology of indiscernible facta that accounts for the truthmakers of temporal predications. In his reply, Mellor explains why he does not accept temporal parts.

That truths have truthmakers does not tell us what truth is or whether it is a property. Ramsey, Mellor's philosophical hero, held a sort of deflationary theory of truth, according to which for it to be true that *p* amounts to no more than that *p*. This view of truth, according to which there is no property of truth, does not seem to require facts – at least if they are understood as anything more than mere true propositions. But Mellor, as we have seen, does postulate facts – his *facta* – to play, among others, the role of truthmakers. How, if at all, can we resolve the apparent tension between these views? This is the main topic of Peter Smith's contribution. The resolution of the tension is obtained by means of a Ramseyan thought that Mellor accepts: that the content of a belief is that *p*, just if, for any appropriate desire, actions caused by that belief combined with a desire will be successful in fulfilling the desire just in case that *p*. When the success condition obtains, the belief is true. But the success condition will normally be a causal condition, and causal conditions involve facta (objects instantiating universals, on Mellor's view). Thus, if a belief is true, certain facta must obtain. One may decide to call this a Truthmaker Principle. As Smith says, this is an apparently happy reconciliation, but one that leaves certain problems unresolved. For instance, what is the relationship between Mellor's facta required by the success condition of 'there is ice cream in the freezer'? The answer to this must depend on what facta there are and thereby on what particulars and properties there are. Smith explores questions like this and shows the relevance of Mellor's anti-physicalist ontology to these answers.

The Ramseyan thought about the content of a belief is easily converted into a thought about truth, known as *success semantics*. Roughly put, a true belief is

one that causes successful actions when combined with the appropriate desires. Indeed, for Mellor truth is 'the property of full beliefs that guarantees the success of actions based on them' (Mellor 1991: 275).

The notion of cause is thus used to explain truth, and causation is a topic on which Mellor has much to say. On the basis of some of Mellor's ideas about causation, for instance that there can be no simultaneous causation, that causal connections are contingent, and that some causation is probabilistic, Chris Daly objects to the idea that a true belief causes a successful action. For Daly, when an action is caused by a desire and a true belief, the action does not ensure the desire's fulfilment, it just makes this fulfilment probable. Daly also argues that Mellor's so-called Ramsey Test for properties commits him to the existence of a property of truth. Thus it is not merely the concept of truth that is elucidated by success semantics, but a property, the property of truth, that is characterized by it. In his reply, Mellor explains why he is not committed to there being a property of truth. His talk of a 'property' of truth is a mere *façon de parler*. He also provides responses and comments to the other objections by Daly, about what Mellor has to say about truth and what he has to say about communication.

Mind and causation

Two distinctive theses in Mellor's philosophy of mind are his objectivism and his non-physicalism. Mellor's objectivism consists in the denial of the claim that there are irreducibly subjective facts corresponding to subjective representations of the world. His non-physicalism consists in the denial of the claim that all facts are physical. Although all physical facts are objective, not all objective facts are physical. According to Mellor, different areas of inquiry have their own truthmakers, the truthmakers of physics being one set of objective truthmakers among others. Although mutually consistent, objectivism and non-physicalism might be thought to be odd bedfellows. Among the most prominent arguments against physicalism is Jackson's so-called 'knowledge argument', which claims that physical facts fail to exhaust reality because there are irreducibly subjective facts associated with subjective mental representation, such as facts about what it is like to see a red tomato (Jackson 1986). On this view, subjectivism is the natural bedfellow for non-physicalism, not objectivism. Mellor has rejected both this view and the argument on which it depends.

Crane questions this dialectic and argues that, suitably understood, the existence of subjective facts is admissible by physicalist and non-physicalist alike. Crane argues that the knowledge argument does, *pace* Mellor, establish the existence of subjective facts, understood as facts about the subjective character of experience. Suitably understood, subjective facts are objects of propositional knowledge the learning of which requires that one occupies a certain position in the world. So understood, it is no more controversial that Mary learns a new fact when she comes to experience red for the first time than that Vladimir learns a new fact when he exclaims 'I am here!', having

located himself on the map. In each case, an item of knowledge is acquired which can only be acquired by someone occupying a certain position in the world. In each case the item of knowledge is a subjective fact. Mellor, however, distinguishes between *facts*, which are mere 'shadows of truths', and *facta*, which are the metaphysically substantial truthmakers of truths. In light of this distinction, Mellor's objectivism could be interpreted as the denial of subjective facta, facts being metaphysically insubstantial. Crane questions this interpretation. A subjective factum is what would have to exist in order for a subjective fact to be learned. But in the case of Mary this is just a visual experience of red. If the subjective factum is the experience of red, then no-one should deny the existence of subjective facta, since all parties to the dispute agree about the existence of experiences. It is therefore unclear what Mellor could be denying in denying the existence of subjective facts. In his replies, Mellor objects to Crane's postulation of subjective facts on the grounds that they entail the claim that some facts are inexpressible. According to Mellor, no ontologically serious facts (in the sense of 'truthmaking facta') are inexpressible. Mellor therefore retains his original view of what Mary learns on her escape from the black-and-white room, namely an ability to imagine and recognize red things.

   Jackson himself no longer subscribes to the knowledge argument, having abandoned the subjectivist ship for the more traditional Australian stance of objectivist physicalism. His chapter addresses the question of what physicalism entails, in particular what descriptions framed in physical terms entail by way of descriptions framed in other terms, such as mental terms describing the character of Mary's experience of a red tomato. Jackson defends the principle of a priori passage, namely that some suitably rich account of the way things are physically a priori entails the way that things are in other respects, including experientially. Thus, taking Jackson's main argument, we can start with the physical premise '60 per cent of the earth is covered by H_2O', add the a priori truth 'Water is the stuff that plays the water role' and the physical truth 'H_2O is the stuff that plays the water role', and thereby infer a priori '60 per cent of the world is covered by water'. Central to Jackson's argument is the so-called 'stop clause', according to which enough physical information only entails all the truths in our world on the condition that it is accompanied by the further piece of information that the physical information is complete. According to Jackson, far from uncovering a substantial metaphysical issue, this requirement merely registers a common fact about any old deduction, such as the deduction of average height from a list of individual heights. Mellor does not share Jackson's faith in the explanatory powers of the sciences of the non-sentient, be it with respect to mental phenomena or non-mental phenomena like heat and water. Taking Jackson's own example, Mellor argues in his replies that it takes distinctively macroscopic kinds to account for the phenomenon of water. Even if water's relation to H_2O is like that of a heap of sand to its grains, it is just as misleading to say that water is H_2O as to say that people are human cells.

In one of its guises, the knowledge argument is an argument in favour of non-physicalist epiphenomenalism. On this view, some physical causes have irreducibly non-physical effects, namely mental states or events such as Mary's experience of red, where these effects themselves are causally inefficacious. The main attraction of epiphenomenalism is its consistency with the causal closure of the physical, or the view that all physical effects have sufficient physical causes. The construal of mental states or events as causally inefficacious has been thought to offer the only acceptable non-physicalist alternative to an interactionist dualism on which the effects of mental states and events would be overdetermined. Mellor has rejected this motivation for epiphenomenalism, arguing that some overdetermination exists between the mental and the physical, even though, as he explains in his replies, it is not systematic.

Noordhof's chapter lends independent support to the view that the non-physicalist is stuck with systematic overdetermination. According to Noordhof, the epiphenomenalist has no independently satisfactory account of the asymmetric causal character of mental facts or events. On the one hand, the epiphenomenalist is happy to appeal to facts of experience to defend the view that the physical gives rise to 'macro-surprises', such as Mary's irreducible experience of red. On the other hand, the epiphenomenalist rejects the appeal to facts of experience to defend the view that some of these 'macro-surprises' are causes of behaviour, such as Mary's exclamation 'So that's what red is like!'. Yet either fact of experience seems equally fundamental. Noordhof argues that any defence of epiphenomenalism requires appeal to the idea that causation involves asymmetric necessitation. Such appeal requires the epiphenomenalist to explain how causation is related to the fact that causes usually precede their effects. According to Noordhof, the only plausible account available to the epiphenomenalist is a causal theory of temporal precedence along Mellorian lines that is either faced with problems regarding the temporal location of mental facts or events or undermines the motivation for epiphenomenalism by facts of experience. Mellor agrees with Noordhof that the epiphenomenalist has no good reason to deny the causal efficacy of the mental. Yet he takes issue with a number of Noordhof's claims about Mellor's theory of causation, as well as one aspect of Noordhof's argument against epiphenomenalism. Where Noordhof questions the consistency of epiphenomenalism with the view that mental facts which lack effects have temporal locations, Mellor sees no difficulty with this combination of views.

Causation is also the topic of Menzies' chapter, in which he questions Mellor's arguments that causation cannot consist in a genuine relation. Menzies argues that the commonsense concept of causation is, indeed, one of an intrinsic relation. According to Menzies, causal relations can be understood as relativized to the contextual parameter of a lawful kind of system, where a lawful kind of system is one in which the intrinsic properties and relations evolve over time in conformity with a common set of laws. On this view, two property instances, such as someone being a smoker and that someone getting cancer, in a lawful system like the human body are causally related if and only

if there is a kind of intrinsic process that typically holds in human bodies when instances of cancer are counterfactually dependent on instances of being a smoker, and a process of this kind holds in the particular human body that includes the someone who smokes and gets cancer. An intrinsic process in a given system is a temporally ordered sequence of states that instantiate the intrinsic properties and relations which constitute that kind of system. Menzies claims that his relational theory of causation has three distinct advantages. First, it deals with cases of causal pre-emption and overdetermination. Second, it deals with cases where causes and effects are absences. Third, the analysis deals with cases of double causal prevention. In his replies, Mellor defends the view that causes raise the chances of their effects against Menzies' problem cases of late pre-emption and coincident causes. In the process of so doing, Mellor accepts Menzies' claim both that causation is embodied in intrinsic properties of law-based systems and that what we think causes something depends on what we hold fixed in assessing relevant counterfactuals. Yet Mellor denies that it follows that causation is a relation.

Dispositions and laws

The topics of dispositions and laws are one of Mellor's main philosophical interests. Dispositional predicates, Mellor believes, are those whose extension is given by a conditional such as 'would be **G** if it were **F**'. But, in general, not all such predicates correspond to properties, since for Mellor properties are those entities over which the quantifiers of the Ramsey sentence of all laws range.

Isaac Levi questions the connection between dispositions and conditionals. For Levi, statements that attribute a disposition (disposition statements) are not equivalent to subjunctive conditionals. There is a straightforward reason for this: Levi thinks that disposition statements have truth values whereas subjunctive conditionals do not. So, although 'the glass is fragile' may be true or false, 'if the glass were dropped, it would break' has no truth value. This means, of course, that disposition statements do not entail such conditionals, but this does not prevent Levi from claiming that belief in a disposition statement *supports* a certain subjunctive conditional. Levi thinks that this gives him an advantage over the view that belief in disposition statements entails subjunctive conditionals.

For Levi dispositional predicates are just placeholders in stopgap covering laws. But this does not mean that dispositional predicates fail to meaningfully apply to objects. All it means is that the laws in which dispositional placeholders appear are not completely adequate for the purposes of explanation. Further inquiry is required to integrate the placeholders into explanatorily adequate theories. Nevertheless, the dispositional placeholders help to provide sketches of explanation.

Since dispositional predicates apply to objects, dispositions are real. But the sense in which they are real, Levi claims, is not a sense that draws an

ontological distinction at the level of predicates, namely a distinction between predicates that have some sort of ontological correlate and predicates that do not. For Levi there are only methodological distinctions to draw between predicates. One such distinction would be that between problem-raising and problem-solving predicates. But this distinction, Levi claims, is relative to research programmes and the state of knowledge at a given time. And so it would seem that such a distinction cannot play the role of the distinction between predicates with ontological correlates and predicates with no ontological correlates that Mellor advocates. Mellor, in his reply, argues that Levi's distinction can fit his own needs.

In his chapter, Alexander Bird considers the issue of whether all properties are essentially dispositional. If this is the case then the instantiation of any property entails some subjunctive conditional. Bird revives a debate between Mellor and Elizabeth Prior about whether the instantiation of a structural property, such as *being triangular*, entails a subjunctive conditional (see Mellor 1974; 1982; Prior 1982). Mellor holds that *being a triangle* does entail such a conditional, and Prior denies this. According to Bird, Mellor is on the side of those who take all properties to be essentially dispositional. But, although there is a plausible story according to which a property such as *being triangular* does entail a subjunctive conditional, there is also a plausible story according to which it does not. And although this does not mean that all properties are dispositional, it does mean that what looked like a reason to reject the idea that all properties are dispositional is not a compelling reason at all. In his reply, Mellor explains why for him the war between those who take all properties to be dispositional and those who take them to be categorical is a phoney war: all properties, *being triangular* included, are both dispositional and categorical. Mellor also says why he does not take properties to be essentially dispositional.

The idea that laws and explanations reduce possibilities is an attractive one. Arnold Koslow's chapter makes a case for this idea. The first thing that Koslow notes is that for laws and explanations to reduce possibilities, a new concept of possibilities and their reduction is needed. Koslow starts by describing a new set of possibilities (he calls them *natural possibilities*), which includes things as varied as the truth values of sentences, the members of sample spaces and the outcomes of tossing a die. Natural possibilities can be abstract (numbers, numerical equations, truth values), concrete (a particular act, such as eating a banana), object-like, property-like, structured, structureless, and so on. However varied these and other possibilities are, Koslow shows why they are all genuinely modal. He does this by introducing a mini-theory of natural possibilities. The modal character of a set of natural possibilities N is explained by means of the notion of a certain implication relation defined on the power set of N.

Koslow explains how laws and explanations reduce possibilities. After showing how laws reduce possibilities, Koslow notes that Mellor's facticity condition on explanation, namely that A's explaining B entails both A and

B, entails that explanations that either are laws or involve laws as parts will reduce possibilities. But not all explanations are like that: some do not involve laws at all. Koslow notes that many models of explanation do not guarantee that explanations reduce possibilities but says that Mellor accepts certain constraints on explanations which yield the result that explanations in general reduce possibilities. The constraint in question is that A's explaining B entails that the chance of B given A is greater than the chance of B given the absence of A.

Mellor replies that he does not assume this constraint on explanations. For him only causes are required to raise the chances of their effects, and many explanations are not causal. Nevertheless, Mellor argues, even these explanations reduce possibilities. To show this only the facticity of explanation is required.

Change and time

All philosophers possess incompatible properties at different times, for example when they change their minds. Mellor has changed his mind about change, or the explanation of how philosophers can possess incompatible properties at different times. His previous view was an instance of what Rodriguez-Pereyra calls the relational theory of change, according to which changeable properties are relations between things and times. Subsequently, Mellor has come to reject this view, arguing that changeable properties are intrinsic to their objects. In his chapter, Rodriguez-Pereyra questions both the relational theory of change and Mellor's reasons for rejecting it. Appealing to such apparent relational properties as 'being in contact with' and 'having been murdered', Rodriguez-Pereyra argues that Mellor fails to show that no changeable properties are relational, and therefore that all changeable properties are intrinsic. Yet this fact may not rescue the relational theory of change. According to Rodriguez-Pereyra, the relational theory fails to explain how change is possible because the incompatible relations it postulates, such as 'holding-the-relational theory-at' and 'denying-the-relational-theory-at', are borne to different entities, namely times. Rodriguez-Pereyra argues that, for relational change to occur, a thing would have to bear incompatible relations to the same entity at different times. The relational theory fails to provide such a single entity. In his replies, Mellor rejects Rodriguez-Pereyra's objection both to the relational theory and to Mellor's own rejection of that theory. He also rejects the amendment to Mellor's own theory of change proposed in Lewis's chapter by rejecting the principle of unrestricted mereological composition of temporal parts on which this amendment depends.

One issue on which Mellor's views remain unchanged is the metaphysics of time: he has repeatedly defended the so-called 'tenseless' theory, or the view that all times past, present and future are equally real. Mellor's tenseless theory, and his use of McTaggart's paradox to demonstrate the unreality of tensed facts, has generated a substantial literature in defence of presentism,

or the view that only the present exists, as a way to get around the inconsistent attribution of past, present and future tense to all times on which McTaggart's paradox depends. Oaklander's chapter further defends the tenseless theory against Craig's version of presentism, which attempts to give an ontological foundation for irreducibly past- and future-tensed statements without falling prey to McTaggart's paradox. According to Oaklander, Craig fails in his attempt to extend the conceptual irreducibility of tense to the level of ontology. Craig needs both to have presently existing truthmakers for past- and future-tensed statements and to avoid the countenance of past and future existences, on pains of contradiction. He therefore claims that past- and future-tensed facts exist at present but are not what ultimately makes these statements true. Oaklander argues that Craig's attempt to show that past- and future-tensed facts are not ultimate can succeed only by reintroducing a tenseless ontology, thereby undermining presentism and reintroducing McTaggart's paradox. Mellor accepts Oaklander's attack on Craig's presentism, and uses his reply to return to Prior's presentist analysis of time. According to Mellor, Prior's failure to complement the semantics of time with an ontology thereof makes presentism both vacuous and question begging.

References

Armstrong, D. M. (1997) *A World of States of Affairs*, Cambridge, UK: Cambridge University Press.

Jackson, F. (1986) 'What Mary didn't know', *Journal of Philosophy* 83: 291–5.

Mellor, D. H. (1971) *The Matter of Chance*, Cambridge, UK: Cambridge University Press.

—— (1974) 'In defense of dispositions', *Philosophical Review* 83: 157–81.

—— (1981) *Real Time*, Cambridge, UK: Cambridge University Press.

—— (1982) 'Counting corners correctly', *Analysis* 42: 96–7.

—— (1991) *Matters of Metaphysics*, Cambridge, UK: Cambridge University Press.

—— (1995) *The Facts of Causation*, London: Routledge.

—— (1998) *Real Time II*, London: Routledge.

Prior, E. (1982) 'The dispositional/categorical distinction', *Analysis* 42: 93–6.

Rodriguez-Pereyra, G. (2000) 'What is the problem of universals?', *Mind* 109: 255–73.

1   Truthmakers for modal truths

David Armstrong

I praise Hugh Mellor for his important contributions to scientific realism and to empirical metaphysics and for his espousal of the notion of truthmakers, that in the world in virtue of which truths are true. These three doctrines, or perhaps directions of thought, interlock in a natural and powerful way. To many of us they seem to give a charter for progress in philosophy, however slow and struggling that progress may be.

1   Introduction

Truth, I think, attaches fundamentally to propositions. We may then define *realism* about the truth of a particular true proposition as the contention that its truth is determined by something that lies *outside* that proposition. This is at least a plausible thesis for the vast majority of true propositions, and I take this plausibility to be the charter for truthmaking theory. In general, propositions that are true have this property of truth in virtue of some portion or portions of the world. (In some cases the word 'portion' must not be too narrowly construed.) It is these portions of the world that we truthmaker theorists call truthmakers.

I go on to make two rather strong claims. The first is Truthmaker *Necessitarianism*. The determining of a truth by a truthmaker is a necessitation, an absolute necessitation. Notice that we should not say that it is an entailment (as I have wrongly said in the past). Entailment can hold only between propositions, and generally at least the truthmaker for a truth will not be a proposition. The connection is cross-categorial. The simplest, if somewhat uninteresting, example of such a necessity holds between any truthmaker, T, and the truth <T exists>.[1]

Notice that necessitarianism seems to require that we take truths as propositions rather than as beliefs, statements, and such. Truthmakers, entities in the world, can hardly necessitate beliefs and statements about these entities, generally at least. What are propositions, then? I think that they are the *intentional objects* of actual or possible beliefs, statements and so on. I hope to give a naturalist, empiricist and, to a degree, deflationary account of intentional objects. All this, however, must be left aside here.

The second principle I uphold is Truthmaker *Maximalism*. Every truth has a truthmaker. (I do not, of course, assert that each truth has its own *unique* truthmaker. Truthmaker theorists, to a man and a woman I think, reject the metaphysics that such a postulation of unique truthmakers demands.[2]) The two principles of necessitarianism and maximalism may still turn out to be too strong. But let us set out in hope. Maximalism, in particular, is central to my argument, as will emerge. One pressing question for a truthmaker maximalist is to suggest plausible truthmakers for modal truths. That is my present enterprise.

2   Truthmakers for mere possibilities

Let us begin by considering truths of possibility: <it is possible that pigs should fly>, and so on.[3] And let us, in particular, concentrate on truths of 'mere possibility': truths having the form 'is possible that p', but where p itself is false. One of the major curiosities of analytic metaphysics in recent decades is that a number of very important and highly regarded philosophers have held that we can do ontological justice to these truths only by huge postulations. David Lewis has postulated his pluriverse, within which all possible worlds exist, 'the worlds in all their glory' in his phrase. Alvin Plantinga and others have not accepted the pluriverse, but have reified 'ways that the world might have been but is not' as what they call 'abstract entities' that exist distinct from the way the world actually is. These philosophers, it seems to me, have not so much brought in men to do a boy's work but have rather brought in giants to do a child's work. My thesis will be that a perfectly good truthmaker for <it is possible that pigs should fly> is the truthmaker for the contingent *truth* <it is not the case that pigs fly>.

Notice here that, by Truthmaker Maximalism, the truth that <it is not the case that pigs fly> has a truthmaker – something that some truthmaker theorists who are not maximalists might deny. Thus maximalism is a very important premise in my proposal for cutting truthmakers for the 'mere possibilities' down to size. This I think of as a consideration in favour of maximalism, although others may wish to argue in the reverse direction.

My argument now requires another premise, a premise that will require some discussion. Here is a first pass at that premise. Suppose that p has truthmaker T, and suppose that p entails q. Then, it seems, in general (perhaps only in general), T must be a truthmaker for q. Call this the Entailment Principle. It may be symbolized informally thus:

$$1 \quad T \rightarrow p$$
$$2 \quad p \text{ entails } q$$
$$\therefore 3 \quad T \rightarrow q.$$

Remember that the arrow is a cross-categorial necessity holding between a portion of the world and a proposition, and so is not a sentential connective.

I have deliberately not substituted any symbol for the word 'entails' in order to allow for the possibility of plugging in different conceptions of entailment here.

First, suppose that entailment here is taken to be classical entailment. Then there is trouble for any attempt to provide truthmakers of any serious interest for necessary truths. Suppose p to be any contingent truth. The maximalist, at least, will hold that it has a truthmaker. By classical entailment, the contingent truth will entail any necessary truth. As a result, accepting this reading of the Entailment Principle will have the consequence that the truthmaker for *any* contingent truth will equally be a truthmaker for *any* necessary truth. (And if any truth that something is possible is itself a necessary truth, as it is in the appealing S5 system of logic, then the truthmaker for any contingent truth will be truthmaker for any modal truth.)

These consequences do not, strictly, refute truthmaker theory, but they do trivialize such a theory for the case of necessary truths, perhaps for all modal truths. And, indeed, many truthmaker theorists seem to be prepared for such a retreat. But I am unwilling to see the theory restricted in this way. It is true that the sorts of truthmaker that I am suggesting for the truths of mere possibility are themselves rather deflationary. But I think that my suggested truthmakers do preserve some natural connection between modal truths and their truthmakers, a connection that is lost if the trivializing account is accepted.

What is to be done, then, by way of getting to a satisfactory version of the Entailment Principle? One thing worth trying is substituting some more restrictive conception of entailment for classical entailment, some conception that does not allow the unfortunate trivializing explosion of truthmakers for modal truths. The suggestion is not that we should abandon classical entailment altogether. The idea would only be not to use it in the formulation of the present Entailment Principle. Horses for courses. This line is taken by Restall (1996), and certain non-classical entailments are endorsed by Stephen Read (2000) for the purposes of truthmaking theory (systems R and E).

Suppose, however, that such proposals are not satisfactory for some reason. One might still retain classical entailment in the formula, but restrict the scope of the application of the principle in some way or ways that would evade the explosion.

In his article Restall discusses one such restriction, which he attributes to Frank Jackson. Although not directly relevant to the question of truthmakers for mere possibilities, this limitation is of sufficient interest to merit a brief digression. Jackson's idea was to restrict the propositions p and q in the formula to contingent truths. To this Restall objected that if for q we substitute the conjunction of contingent p with any necessary truth N, then that conjunction is still a contingent truth because one conjunct is contingent. But p classically entails p & N, so a truthmaker for p is a truthmaker for p & N. And then, by the highly plausible principle that a truthmaker for a conjunctive truth is a

truthmaker for each conjunct, the truthmaker for contingent *p* still becomes truthmaker for N.

But could not Jackson reply by retreating a little further? Define a 'purely contingent' truth as one that contains, at any point in the given structure of the truth, contingent truths alone. The idea (which I may have not caught quite satisfactorily, though it seems catchable) is that it should be a contingent truth through and through. This revised Jackson principle then seems very plausible for such purely contingent truths, and is, I think, a valuable addition to the general principles of truthmaking.[4]

We can now return to the mere possibilities. Given that *p* is true, and given that *p* is contingent, then it can surely be concluded that <not-*p* is (merely) possible>. In earlier days we would probably have said that the entailment is analytic. Even if we do not say this, however, can we not say that the entailment holds in virtue of what contingency *is*? It therefore seems likely that the Entailment Principle holds in such a context, even with classical entailment. That is to say, the truthmaker for *p* should also be the truthmaker for <not-*p* is possible>. The proposition <not-*p* is possible> may well be a necessary truth (it is in S5 at any rate). We are therefore moving beyond the revised Jackson principle. Nevertheless, the extension looks to give us a valid argument.

Here is the suggestion presented as a formula:

$$1 \quad T \rightarrow p$$
$$2 \quad T \rightarrow \langle p \text{ is contingent} \rangle$$
$$\therefore \quad 3 \quad T \rightarrow \langle p \& p \text{ is contingent} \rangle$$
$$4 \quad \langle p \& p \text{ is contingent} \rangle \text{ entails } \langle \text{it is possible that not-}p \rangle$$
$$\therefore \quad 5 \quad T \rightarrow \langle \text{it is possible that not-}p \rangle \text{ (by the Entailment Principle)}.$$

A discussion of Premise 2 is in order. T is something in the world, some state of affairs or other entity depending on just what truthmakers are postulated, a matter that depends on one's whole metaphysics. Whatever T is, in the cases we are considering it is a contingent being. Could the contingency of T lie outside T? It does not seem possible. It cannot be a relation that T has to something beyond itself. So T is the truthmaker for the proposition <*p* is contingent>. Whether we need a *property* of contingency *in re*, a special categorial property of the truthmaker, is a difficult question of metaphysics that I trust need not be entered into here. For myself, I hope to avoid having to postulate such a property. But, however one resolves that matter, it is difficult to quarrel with the idea that any truthmaker for *p* is also the truthmaker for <*p* is contingent>.

The step from the conjunction of 1 and 2 to 3 seems even less controversial. The mereological sum of the truthmakers for two propositions should be a truthmaker for the conjunction of the two propositions. In this case the sum is T + T, which in mereology sums to T. Hence, using the Entailment Principle, the idea that the truthmaker for a contingent truth is also a truthmaker for the *possibility* that it is not true is upheld.

With that, the *need* for more far-fetched truthmakers, for instance really existing possible worlds, seems to be removed. One might still wish to postulate as truthmakers a realm of possibilities in ontological addition to actualities. Nothing in my argument rules this out. But the pressure to make this ontological addition seems very much reduced. Occamist considerations become very weighty. Notice, also, that one is not committed to the idea that the truthmaker for a contingent truth, even if it is a minimal truthmaker for that truth, is necessarily a *minimal* truthmaker for the associated mere possibility that it is false. The inquiry into minimal truthmakers for truths of possibility is an important topic, but one that will not be pursued further in this chapter. But that any truthmaker for a contingent truth is also truthmaker for the possibility of the contradictory proposition seems a most important result for a one-world and naturalistic metaphysics. A big step is made towards providing deflationary, yet relevant, truthmakers for all modal truths.

This argument just presented for a this-worldly account of truthmakers for truths of mere possibility is somewhat elaborate. There is a much simpler argument that may have weight. Consider the totality of contingent beings. If any of these beings were not to exist and/or contingent beings that do not exist were to exist, then the mere possibilities would have automatically to co-vary with these differences. That is to say, the mere possibilities supervene upon the contingent beings with absolute necessity. This consideration, of course, does not show us in any detail what the truthmakers for the truths of mere possibility are. But it again casts some cold water on the need for the wildly ambitious truthmakers that have been proposed by many contemporary metaphysicians.

The question does arise of what metaphysical interpretation we should place on this supervenience. I should like to interpret it as showing that the mere possibilities are no addition of being to the contingent beings. Compare the necessary supervenience of the mental on the physical – a controversial doctrine, of course, but one that is normally taken to mean that the mental is no addition of being to the physical. 'No addition of being' does not mean that the mental does not exist. It is not a charter for eliminativism about the mental. In the same way, the supervenience of the mere possibilities does not mean that there are no such possibilities. But it does mean, I contend, that these possibilities are not something ontologically *additional* to the contingent existences.

3 Aliens

One very interesting sort of mere possibilities – well, interesting to metaphysicians anyway – are the aliens. We owe the use of the term 'alien' in metaphysics to David Lewis. An alien, for him, is something that neither exists in our world nor is combinatorially constructible from things that do exist in this world. (Consider the way that centaurs and such like are combinatorially constructible. They are therefore not aliens.) Take alien properties first. It may well be,

Lewis thinks (epistemic 'may'), that there are properties, in particular, which are aliens to our world but which are instantiated in other worlds. Why should not there be many other worlds that are much more property rich than our world? Given the Lewisian pluriverse, this sounds a very plausible argument, but obviously not an apodeictic one, for the presence of property aliens in many other worlds. As for particulars in other worlds, for Lewis they are, strictly, all of them alien to our world. This is because he analyses mere possibilities in our world as mere counterparts of the particulars of this world. For him, as is well known, there is no (strict) trans-world identity of particulars.

For a one-world chauvinist like myself, however, the aliens, properties as well as particulars, can only be *mere* possibilities. The question is, though, what can be the truthmaker for the assertion that aliens are possible? In the past I have not handled this question well. One suggestion about properties, in particular, that I embraced for a while is that, although the concept of an alien property is thinkable, it is not really a genuine possibility. An alien property, on this view, is like squaring the circle, which is thinkable but in fact impossible. But if, like me, you think of properties as contingent entities, and uninstantiated properties as non-existent, such a priori limitation of the possible properties can hardly be maintained. So that idea had to be discarded. The situation is still worse with alien particulars. Surely there might be alien particulars – a duck in this room now, say – which are not identical, partly or wholly, with any actual particular?

It seems, though, that the treatment already given of truthmakers for truths of mere possibility will serve us in dealing with the (apparent) modal truth that there might have been aliens. Consider properties in particular. There is an actually existing entity (I am taking existence omnitemporally) that is the totality of properties (all of them instantiated, according to me). You can imagine this totality as recorded in a list, perhaps an infinite one, yielding a true proposition.

What is the truthmaker for the truth that a certain class of properties is the class of *all* the properties? Provided you concede that this truth has a truthmaker, which I at least am committed to by Truthmaker Maximalism, then there seem to be two options. First, you might think that just the sum of the members of the class constitute a satisfactory truthmaker. I cannot agree with this, because I am also committed to truthmakers *necessitating* their truths, and it seems clear that there might have been more properties 'on the list' and so in the class. I think the ontology of *allness*, if you will allow me the word, demands a special sort of fact or state of affairs (here I am close to Russell, and differ from the Wittgenstein of the *Tractatus*).

But perhaps this dispute can be bracketed, provided it is accepted that there is a truthmaker *of some sort* for the truth that a certain class of properties is the class of all the properties that there are. What, then, of the modal truth that it is possible (merely possible) that what is in fact the class of all the properties is a mere sub-class of the class of all the properties? This is the truth that we need a truthmaker for. But will it not fall within the scope of the Entailment

Principle, indeed within the scope of the amended Jackson principle, if properties are, as I contend, contingent beings? If so, the truthmaker for the truth that *these* are all the properties will also be the truthmaker for the truth that it is possible that these should *not* have been all the properties.

This does not quite get us to the aliens, because the class of merely possible properties includes non-alien properties: those that can be reached combinatorially from the real properties. Consider, however, the modally mixed class that comprises the union of the real properties and all the unreal but combinatorially accessible properties. It would seem that the sub-class of the non-existent but combinatorially accessible properties is necessitated by the actual, real, properties. For it is because of the intrinsic nature of the actual properties that they are combinable or uncombinable in the forming of merely possible properties of particulars. The combinatorially accessible properties supervene.[5] So the truthmaker for the modally mixed class of actual properties plus the merely possible properties, but excluding the alien possible properties, is the conjunction of the actual properties, plus whatever makes it true that they *are* all the actual properties.

And, if the Entailment Principle applies here (and why should it not?), that truthmaker is also the truthmaker for the possibility of extra properties, the ones whose possibility does not depend upon the combinability of actual properties, in short the aliens. Once again, a notable ontological economy is achieved with respect to truthmakers.

4   Is it possible for there to be nothing at all?

Perhaps, only perhaps as we shall see, the Entailment Principle will throw light on the traditional philosopher's question 'Is it possible that there be nothing at all?' Notice first that if there are necessary beings, then a negative answer must be given to this question. So let us very temporarily bracket the question of necessary beings, and ask only if it is possible that there be no contingent existences. Then consider the proposition <there is at least one contingent being>. This would appear to be true, and indeed to have innumerable truthmakers, each of them sufficient by itself for the truth of this proposition. If it is false, then this can only be because the world as a whole is a necessary being, 'the hideous hypothesis of the atheist Spinoza' as Hume so delightfully put it. That is an epistemic possibility that we once again bracket.

Now consider the conjunction of propositions <there is at least one contingent being and <this proposition is a contingent truth>>. Suppose this is true, and suppose that the Entailment Principle holds for antecedents having the form p and p *is contingent*, something which I have argued is very plausible. Then, it would seem, it can be inferred that <it is possible that it is not the case that there is at least one contingent being>. A universe empty of contingent beings appears to be a possibility.

This result, however, runs into a difficulty. Suppose it to be true that there

is no necessary being, a proposition I greatly incline to accept. Given this, the putative possible world that is empty of contingent beings is empty, period. But now, in this supposed world, what is there to be the truthmaker for the truth that there is nothing? This is a nasty objection for me, holding as I do that every truth has a truthmaker. I suppose that a devoted Meinongian might say that the truthmaker is 'the state of nothing at all'. I draw the line at this.

What is to be said, then? For entailment to go through here we need the truth of <there is at least one contingent being>, but also the proposition that <this is a contingent truth>. This suggests a possible way out. Perhaps the second conjunct is not contingent after all, but necessary. Perhaps, supposing necessary beings to be impossible as I incline to think, there has to be at least one contingent being. I am attracted by this idea, but it has its own difficulties. So I will leave the matter in suspense.[6]

5   Truthmakers for necessary truths

Not all truths of possibility are truths of mere possibility. What is actual is also possible, but truthmakers for contingent truths will automatically be truthmakers for their possibility (though usually they will not be minimal truthmakers). What is necessary is also possible, but we can subsume that enquiry under the question of truthmakers for necessary truths, which is our present business.

What I now want to argue is that necessary truths, in so far as they are necessary, give us no information about the *existence* of anything at all. They are concerned with *possibilities* alone. A corollary is that the rational sciences of mathematics and logic, where the bulk of interesting necessary truths are to be found, are concerned not with existence but only with possible existence. The sphere of necessity, and so the sphere of the rational sciences, is wider than the sphere of the actual. It is the sphere of the possible.

How is this position to be supported? From necessary truths alone, no contingent conclusions can be derived. (An interesting asymmetry here is that, using classical entailment, from contingent truths any necessary truth can be derived.) So if any considerations in the rational sciences lead us to postulate actual existents, then these will have to be *necessary* beings. The question then is whether we have any reason to postulate necessary beings, in particular whether the rational sciences give us any reason to postulate such beings.

Consider the case of the numbers. Have not mathematicians shown that all sorts of numbers exist? A particularly striking illustration is Cantor's demonstration that there are an indefinite, indeed an infinite, number of infinite numbers. Cantor's proof, using the beautiful diagonal argument, proceeds, like all mathematical arguments, purely a priori. Proceeding from necessary truths by necessary steps, the existence of these numbers is shown to obtain. Therefore they cannot be contingent existences. Will they not be necessary beings? And once we have admitted them as necessary beings, it will be hard

to deny that more ordinary mathematical entities, such as the humble natural number 7, are also necessary beings.

This, of course, creates a problem for empiricists. It is a descendent of Kant's famous, and well-justified, question about synthetic a priori knowledge. How is it possible that mathematics, and (presumably) logic, should be able to yield such extraordinary knowledge within the realm of necessity? Mathematicians investigate all sorts of very different, but very general, topic-neutral, *structures*. Their method is definition and proof, and their conclusions, if not absolutely certain, are more nearly certain than anything else that we are able to attain to. How is such knowledge possible? A great difficulty, I suggest. One way of dealing with it, of course, is to give up empiricism and embrace the idea of a non-natural realm to which a special faculty of mind ('reason') gives us access. Notoriously, this was what Gödel did.[7]

I am not willing to give up empiricism. I therefore suggest that we embrace what may be called Possibilism[8] in mathematics and logic. When the mathematician or logician demonstrates the existence of some entity we should understand it as demonstrating the *possibility* of existence of some structure in the empirical world which instantiates the entity in question. Thus, we can think of 7 + 5 as a very abstract, but empirical, structure. 'Abstract' here is, of course, the commonsense use of the word, and does not connote going beyond the empirical realm. (It is linked to the topic neutrality of mathematics and logic.) This structure is instantiated by innumerable cases where there are seven things and five further things (7 and 5 being themselves simpler empirical structures.) These structures will also, of course, be instantiated *within* mathematics and logic. But they begin life, as it were, as empirical structures.

Some abstract structures dealt with by mathematics and logic, however, may not be instantiated. That enables us to deal with the infinite numbers. If the world nowhere contains any infinity, a proposition that may be true for all we know (but equally it may be false), then the infinite numbers are structures that are not instantiated. Supposing this to be so, it is still very plausible that they are *possible* empirical structures. For it is hard to believe that it is *impossible* that there is infinity somewhere in the structure of the empirical world. One would want a proof of this impossibility and, in the unlikely event of a satisfactory proof being found, then the infinite numbers would join the round squares in ontological oblivion. But in fact the question of whether there is infinity in the empirical world appears to be an empirical issue, even if one that could never be conclusively verified or falsified. Science may eventually cast some light on it, all the same.

If this is on the right track, then we can link it up with the claims made in Section 2 of this chapter about truthmakers for truths of possibility. Suppose, for the sake of argument, that there is no infinity in the structure of the world. Then claims about the infinite numbers amount to claims that among the

(mere) possibilities for the world are structures that instantiate those infinite numbers. For instance, there would seem to be the (mere) possibility that the class of all the electrons has the number of the natural numbers, the smallest infinite number, aleph$_0$. The truth, though, as we are supposing, will be that the number of this class is a large but finite number. The truthmaker for that truth will be the actual class, with its actual finite number, whatever that number is. In accordance with the Entailment Principle, this truthmaker can then be proposed for the modal proposition <it is possible that the number of the electrons is not finite>. Suppose, however, that it is instead true that there are an infinite number of electrons. The actual class of the electrons (and an infinite number of sub-classes of this class) will be the truthmaker for that proposition.[9]

The treatment just given may perhaps suffice for 'necessities of existence', in particular proofs of the existence of entities within mathematics and logic. But we have also to deal with necessary connections. Given 7 + 5, these numbers must add up to 12. What sorts of truthmaker can we offer for such necessary truths? It is not sufficient to point out that, strictly, what are here necessarily linked together by the relation of equality are possibilities rather than actualities. (One might point out, in defence of this, that in a very small world the number 12, say, might be a structure that was not instantiated.) Granted that, still these possibilities are necessarily connected. Instantiate the antecedent and the consequent must be instantiated, and vice versa. What is the truthmaker for this necessary connection?

I want to argue here that the *terms* of the relation are sufficient as truthmakers. The relation of equality holding between these terms is an internal one, and for all *internal* relations, I suggest, they are not an ontological addition. The two terms, 7 + 5 and 12, are all that is required. The relation of equality supervenes. Internal relations are not unreal, but they are not an addition of being to their terms. The contrast, of course, is with contingent truths such as <the number of the apostles is 12> or <some roses are red>. With the necessary truths the terms are internally connected, but they are not so related in the case of the contingent truths.

Suppose that the two terms of a dyadic and *internal* relation, R, are given. The two terms, *a* and *b*, taken together, will be the truthmakers for the truth <*a* R *b*>. So we have:

$$a + b \rightarrow\ <a\ \mathrm{R}\ b>.$$

The arrow will be the cross-categorial relation of necessitation, absolute or metaphysical necessitation. Notice that I have given no apodeictic argument for the absence of any further truthmaker, for instance *necessary* states of affairs involving internal relations. But, once again, there seems to be no need to postulate such additional truthmakers in the ontology.

6   Further matters

Impossibilities

Something may be said about truths of impossibility. It is a truth that it is impossible that there should be a round square, or an angle that has been trisected using ruler and compasses alone. The truthmaker for the first of these truths, sticking to this one for simplicity's sake, seems to be just the property *being round* together with the property *being square*. Here we may find some use for the notion of a *falsemaker*. If something is a truthmaker for something being round, then that very same truthmaker is a falsemaker for that same something being square. And, of course, a truthmaker for some proposition will always be a falsemaker for the contradictory (external negation) of that proposition.

It is interesting, though, to consider what a paraconsistentist, who holds that some contradictions are true, could make of truthmaking applied to such an alleged truth. (Since they want to be realists about certain contradictions, they ought, I think, to embrace truthmaking theory.) Suppose p & $\neg p$ to be such an alleged truth. I suppose the position would have to be that both conjuncts have the *identical* truthmaker, perhaps a contradictory state of affairs.

Analytic truths

These have been given a hard time in recent decades. The definition that I have always liked is that a truth is analytically true if and only if it is true solely in virtue of the meanings of the words in which it is expressed. We do not, however, want to say that the truth is a truth about words. That would seem to turn it into a contingent truth, which it is not. But how else is the phrase 'in virtue of' to be construed? Truthmakers can come to the rescue, I suggest. <A bachelor is an unmarried adult male> is *about* bachelors, and attributes certain properties to them. But the *truthmaker* for this proposition is solely the meanings of the words in which it is stated, in particular the meaning of the word 'bachelor' and its identity with the meaning of 'adult unmarried male'. We need not answer the difficult question of what meanings *are*. For myself I think that there are such things, whatever account ought to be given of them. It will be seen that the notions of *reference* and *truthmaker* come apart here. I do not see any particular objection to this, but it may be that confusion between the two notions is the factor that has made it hard to give an intelligible account of analytic truths.

Conceptual truths

If there is a distinction between them and analytic truths, as I incline to think that there is, conceptual truths can be treated in the same way. Suppose it to be – as I incline to think it is – a conceptual truth that veridical perception

conceptually necessitates that what is perceived is the cause of the perception. That is a truth about veridical perception and causation. But its truthmaker is our *concepts* of perception and causation (whatever concepts are).

How wide is the scope of the analytic/conceptual necessities? If the deflationary line that I have been arguing for about the ontology that lies behind necessary truths is correct, one might consider reviving the idea that necessary truths are *all of them* analytic and/or conceptual. That is a speculative suggestion, though.

Notes

1   I assume that existence is not a property, although 'existence' is a perfectly good predicate. So there is no fact or state of affairs of T's existing.
2   The metaphysics would come out as a realist parody of Paul Horwich's (1990) minimalist theory of truth. Given a forced choice between this inflated metaphysics and the minimalist theory, the minimalist theory seems the more attractive!
3   The example is to a degree controversial. If you hold that the laws of nature are metaphysical necessities, then pigs flying may be argued to be impossible. Truths of impossibility will be considered briefly in Section 6.
4   I thank Glenn Ross for very useful discussion of the Entailment Principle. The sorts of contingent truth that one finds oneself dealing with in truthmaking theory are, in general, fairly obviously 'purely contingent' if contingent at all.
5   My own combinatorialist account of possibility looks to a promiscuous recombination of existences that are – wholly – distinct existences. But that is not at issue here. The only premise that I need for the present argument is that combinability or non-combinability is determined solely by the nature of the terms involved.
6   Thanks to Greg Restall for discussion here.
7   See Gödel (1944).
8   For much more technical discussion, to which I am able to contribute little, see Putnam (1975) and Hellman (1989).
9   In this sort of case, we have truths that lack *minimal* truthmakers, a point spotted by Restall (unpublished work). I have said nothing about classes in the body of this chapter. But considering the iterative set-theoretical hierarchy, I think that we can distinguish between empirical classes, which are structures actually found in the world, and non-empirical classes, which are no more than the possibility of such structures. If for instance one takes the whole world – the totality of being – then the singleton class of which the world is the sole member is a non-empirical class. See Armstrong (1997: Ch. 12). The treatment is parallel to the treatment of the infinite numbers proposed in this chapter.

References

Armstrong, D. M. (1997) *A World of States of Affairs*, Cambridge, UK: Cambridge University Press.
Gödel, K. (1944) 'Russell's mathematical logic', in P. A. Schilpp (ed.) *The Philosophy of Bertrand Russell*, *The Library of Living Philosophers*, Vol. V, Evanston, IL: Northwestern University.

Hellman, G. (1989) *Mathematics without Numbers: Towards a Modal–Structural Interpretation*, Oxford: Clarendon Press.

Horwich, P. (1990) *Truth*, Oxford: Basil Blackwell.

Putnam, H. (1975) 'Mathematics without foundations', in *Philosophical Papers*, Vol. 1, Cambridge, UK: Cambridge University Press.

Read, Stephen (2000) 'Truthmakers and the disjunction thesis', *Mind,* 109: 67–79.

Restall, Greg (1996)'Truthmakers, Entailment and Necessity', *Australasian Journal of Philosophy*, 74: 331–40.

2 Things qua truthmakers

David Lewis

1 Truth and being

Any proposition has a subject matter, on which its truth value supervenes. Suppose that a certain proposition is entirely about styrofoam. Then its truth value supervenes upon the totality of the world's styrofoam. If two possible worlds were just alike with respect to their styrofoam – if they had styrofoam of just the same kind at just the same places and times – then, no matter how much those two worlds differed otherwise, the proposition would be true in both worlds or false in both. Conversely, if some proposition never differed in truth value between two worlds that were just alike with respect to their styrofoam, then that proposition would have to be entirely about styrofoam.

What, in general, *is* a subject matter? The answer is anything that somehow encodes the distinction between pairs of worlds that are just alike with respect to the subject matter in question and pairs that are not. A partition of the possible worlds would do, or equivalently an equivalence relation on worlds.

The present conception of aboutness and subject matters, following Lewis (1988), is intensional, not hyperintensional. It does not apply usefully to aboutness in mathematics or philosophy. The truth values of necessary and impossible propositions, regardless of whether they are expressed by sentences that speak of sines and cosines or by sentences that speak of the marital status of bachelors, turn out to supervene trivially on every subject matter.

Styrofoam is one kind of plastic. Therefore two worlds exactly alike with respect to plastic would a fortiori be exactly alike with respect to styrofoam. A proposition entirely about styrofoam is a fortiori entirely about plastic. The subject matter styrofoam is part of the more inclusive subject matter plastic. But plastic may in turn be part of other, still more inclusive, subject matters.

There is a most inclusive subject matter: *being*. Differences in being come in two sorts. There are differences in whether something is, and there are differences in how something is. Two worlds are alike with respect to being if they have no differences of either sort. Nothing exists in one but not in the other. Nothing has a fundamental property in one that it lacks in the other. No two (or more) things stand in a fundamental relation in one but not the other. And, since less-than-fundamental properties (and relations) supervene upon

fundamental properties and relations, nothing has any less-than-fundamental property (and no two or more things stand in any less-than-fundamental relation) in one but not the other. Since being is the most inclusive subject matter, two worlds that are just alike with respect to being are just alike *simpliciter*, and just alike with respect to every less inclusive subject matter. They are just alike with respect to plastic, with respect to styrofoam, … . And every proposition, no matter what lesser subject matter it may also have, is entirely about being. It never has different truth values in two worlds that are just alike with respect to being. In John Bigelow's (1988: 132–3) phrase 'its truth is supervenient on being'.

You might object that if there were two worlds just alike with respect to being, then there would be miscellaneous classes of worlds containing one of the two without the other. For any such class we have the proposition that is true at all and only the worlds in that class; so here we have propositions whose truth does not supervene on being. There are two replies.

(1)   A miscellaneous class of worlds does not determine a proposition at all – or, at any rate, it does not determine what we might call a *qualitative* proposition. The principle that truth supervenes on being applies to qualitative propositions only. Non-qualitative 'propositions', if we may call them that, may be ignored. Indeed, qualitative propositions are exactly those whose truth does supervene on being. Our principle has become true by definition – and none the worse for that. (Likewise when we said that less-than-fundamental properties of things supervened on the fundamental properties and relations of things, we meant the less-than-fundamental *qualitative* properties. Again our supervenience thesis was not meant to apply to non-qualitative 'properties' determined by miscellaneous classes of possible individuals. Again, what at first seemed to be a substantive supervenience thesis turns into a definition, this time of 'qualitative property'.)

(2)   The problem never arises, because indiscernibility with respect to being implies identity. No two worlds are ever exactly alike with respect to being. Therefore there are no miscellaneous classes that contain one but not the other of some such pair of worlds. Neither are there any non-qualitative 'propositions' that are true at one but not the other of some such pair. Our principle that truth supervenes on being is now not a definition, but rather a substantive thesis of identity of indiscernible worlds. (Likewise if possible individuals obey a suitable principle of identity of indiscernibles, there will be no non-qualitative 'properties' determined by miscellaneous classes of possible individuals. But identity of indiscernibles is far less plausible for individuals than it is for worlds, because it would rule out, for instance, the indiscernibilities found in a world of two-way eternal recurrence.)

Let me remain agnostic about whether there are indiscernible worlds and

non-qualitative 'propositions'. (And even, so far as this chapter goes, about whether there are indiscernible possible individuals and non-qualitative 'properties' and 'relations'.) But if it matters, let me impose a tacit restriction to qualitative propositions (and properties and relations).

2   Counterparts

I said that two worlds are alike with respect to being only if there is nothing that exists in one but not the other. But strictly speaking I say that this is never true. Nothing is (wholly) in two different worlds. (Unless it is a universal. But since no world is inhabited by universals alone, it still cannot happen that exactly the same things exist in two worlds.) What is true, rather, is that things have counterparts in other worlds, united with them not by identity but by some sort of intrinsic or extrinsic resemblance (see Lewis 1968). What I meant, then, was that two worlds are exactly alike with respect to being just in case their inhabitants correspond one–one in such a way that corresponding things have exactly the same fundamental properties and corresponding pairs (or triples or …) stand in exactly the same fundamental relations. The correspondence is not always unique: between two indiscernible worlds of two-way eternal recurrence, for instance, there are infinitely many admissible correspondences.

However, if we do have a unique one–one correspondence such that corresponding things match perfectly with respect to all the fundamental properties and relations, then it is completely unproblematic which things are counterparts of which. Then it would scarcely matter if we mistook the counterpart relation for genuine identity. But of course this is an especially easy case. In the general case we will have many counterpart relations – or, you might prefer to say, many alternative precisifications of 'the' counterpart relation. These relations will weigh different respects of intrinsic or extrinsic similarity differently (or sometimes not at all), and so they will pair things off differently with their otherworldly partners. And sometimes the price to be paid for respecting (the appropriate sorts of) similarity and dissimilarity, and avoiding arbitrary choices, will be that the counterpart relation is no longer a neat one–one correspondence. One thing in this world may have one counterpart in that world, or two, or even more, or none; and two in this world may share a common counterpart in that world.

Counterpart theory makes a kind of sense of essentialism: *a* is essentially *F* just in case all of *a*'s counterparts (including *a* itself) are *F*. But this is a half-hearted and flexible essentialism. The truth of (all but the most trivial) essentialist judgements is relative to the counterpart relation. Indeed Quine (1976) once formulated his well-known misgivings about essentialism exactly as a complaint that we have no determinate counterpart relation. Such flexibility is all to the good. Our essentialist judgements *are* flexible. (Except in the case of those who follow where philosophical fashion leads, and imagine that some interesting essentialistic judgements have been established once

and for all.) Today, thinking of Saul Kripke as essentially the occupant of a distinguished role in contemporary philosophy, I can truly say that he might have been brought by a stork. Tomorrow, thinking of him as essentially the man who came from whatever sperm and egg he actually came from, I can truly say that he might never have had a philosophical thought in his life. I would be right both times, but relative to different, equally admissible, counterpart relations.

Lumpl the lump was created in the shape of a statue of Goliath, and remained in that shape until destroyed (Gibbard 1975). Lumpl *is* Goliath. Yet what might have happened to Lumpl differs from what might have happened to Goliath. Lumpl could have survived squashing. Goliath could not. How so, if indeed Lumpl and Goliath are one and the same? In another world there is something that does survive squashing. Is it a counterpart of Lumpl/Goliath? Yes and no. It is a counterpart under the counterpart relation that is called to mind when we describe Lumpl/Goliath as a lump, but not under the different counterpart relation that is called to mind when we describe the very same thing as a statue. Even the two names, when introduced in the way I did, are evocative. 'Lumpl' evokes a counterpart relation on which Lumpl/Goliath does have counterparts that survive squashing. 'Goliath' evokes a counterpart relation on which it does not. Thanks to the multiplicity of counterpart relations, we have no need to multiply entities.

Likewise, since I have no immaterial soul, I am my body. Yet my body could, and I could not, survive the complete erasure of my mental life; but I could, and my body could not, survive the transcription of my mental life into the previously blank brain of a different body, while at the same time my original body was destroyed. The solution is the same (Lewis 1971). One identical thing can have different potentialities and different essences if it has them relative to different counterpart relations. The one identical thing is both a person and a body, but these different descriptions evoke different counterpart relations. Thus we have the illusion that there are two different things.

3   Truthmaking

One way for the truth of a proposition to supervene on being is for that proposition to be made true, in any world where it is true, by a truthmaker. If *a* is a possible individual and *P* is a proposition, call *a* a *truthmaker* for *P* just in case every world where *a* exists is a world where *P* is true. By 'world where *a* exists' I mean, of course, 'world where *a* has a counterpart'. (Otherwise, anything that exists in just one world would trivially count as a truthmaker for all propositions true in its world.) Note that a proposition may have different truthmakers in different worlds; and that it may have many truthmakers in a single world, any one of which would have sufficed to make it true. Note also that finding a truthmaker need not afford an informative explanation of why a proposition is true. Take the proposition that there is a cat. It is true

because it has a truthmaker. And what are its truthmakers? Cats. So it is true because there is a cat.

Call a proposition *positive existential* – for short, *positive* – just in case it has a truthmaker in any world where it is true. Some philosophers hold the *Truthmaker Principle*: they say that every truth must have a truthmaker. That is, all propositions are positive. In recent times, the Truthmaker Principle has been advocated by C. B. Martin, then by D. M. Armstrong, then (either in its original form or in revised versions) by many others (Fox 1987; Bigelow 1988; Armstrong 1989; Martin 1996; Mellor 1998: 19–28; Lewis forthcoming). But it had appeared often before, under different names in different traditions (Mulligan *et al*. 1984).

Even if the Truthmaker Principle is false, the supervenience of truth on being is unscathed. There are more ways than one for the truth of a proposition to supervene on being. Call possible individual *a* a *falsemaker* for proposition *P* just in case every world where *a* exists – or, rather, has a counterpart – is a world where *P* is false. For instance (assuming that if any possible individual is a unicorn, it is so essentially) a unicorn is a falsemaker for the proposition that there are no unicorns. That proposition is true in this world because it has no falsemakers. (Again, this is not an informative explanation.) Call a proposition *negative existential* – for short, *negative* – just in case it has a falsemaker in any world where it is false. A falsemaker for *P* is a truthmaker for not-*P*, and vice versa. So if the Truthmaker Principle is correct and, necessarily, every truth has a truthmaker, then also, necessarily, every falsehood has a falsemaker. Further, necessarily, every truth lacks falsemakers and every falsehood lacks truthmakers. In short, every proposition is both positive and negative.

But if the Truthmaker Principle is incorrect, then many more cases may be possible. A proposition may be positive, or negative, or both, or neither. If proposition *P* is true in world W_1 and false in W_2, *P* might have truthmakers in W_1 but not in W_2, or falsemakers in W_2 but not in W_1, or both, or neither. And if it is neither, something in W_1 might have some fundamental property that its counterpart in W_2 lacks, or vice versa or both. Or some pair (or triple, or …) of things in W_1 might stand in some fundamental relation, but the pair (or …) of their counterparts in W_2 might not, or vice versa or both. In each case W_1 and W_2 differ somehow with respect to being. So each case respects the requirement that whether *P* is true must supervene on being.

4   Making predications true

The principle that truth supervenes on being is a safe fallback. Nevertheless, it is interesting to see how far we can get with the Truthmaker Principle itself. I once doubted that there were truthmakers for negative existential truths, such as the truth that there are no unicorns. I also doubted that there were truthmakers for predications, such as the truth that cat Long is black. For the time being I retain my doubt about negative existentials (Rosen and

I reconsider that question in our postscript to the present chapter), but I withdraw my doubt about truthmakers for predications.

When I doubted that there were truthmakers for predications, I was trying to remain entirely neutral about the metaphysics of modality (Lewis 2001). Under that constraint, I still do not see how a satisfactory theory of truthmaking for predications can be found. But when I abandon neutrality, and work within counterpart theory (or some alternative that matches the flexibility of counterpart theory; see Lewis 1986: 259–63), I think I can do better.

We shall consider predications of intrinsic properties. But if intrinsic predications always have truthmakers, then many extrinsic predications do too. For things have many of their extrinsic properties in virtue of the intrinsic properties of more inclusive things – perhaps the entire universe, perhaps something less. Where F is one of these extrinsic properties, the proposition *Fa* is implied by some *Gb,* where G is intrinsic and a is part of b (provided that, relative to our counterpart relation, any counterpart of b includes a counterpart of a and any counterpart of a is included in a counterpart of b). So a truthmaker for *Gb* is a truthmaker for *Fa* as well. But not all extrinsic predications are covered in this way: things have some of their extrinsic properties at least partly in virtue of negative existentials.

Consider the proposition that cat Long is black. Is there a truthmaker for this intrinsic predication? We might be tempted to redefine truthmaking so as to make it easy to find 'truthmakers' for intrinsic predications. Call a a *truthmaker** for P just in case every world where a exists with no change in its intrinsic properties is a world where P is true, in other words just in case every world where a has a counterpart that is also an intrinsic duplicate of a is a world where P is true (Parsons 1999). Long himself is a truthmaker* for the truth that Long is black, and for every other true intrinsic predication with Long as subject.

Truthmaking* is all very well. But what would it take to give us truthmakers for predications without having recourse to redefinition?

Imagine something, call it *Long qua black*, that is very like Long in most ways, but differs from him in essence. Long is accidentally black, and might have been striped, orange all over, or even green. Long qua black, however, is essentially black. Long has counterparts of many colours, whereas all counterparts of Long qua black are black. Indeed, the counterparts of Long qua black are all and only the black counterparts of Long. Long qua black, if there were such a thing, would be a truthmaker for the truth that Long is black. Every world where Long qua black had a counterpart would be a world where Long is black.

Better still, imagine something, call it *Long qua just as he is*, that is very like Long but having all of Long's intrinsic properties essentially. Its counterparts are all and only those of Long's counterparts that are also intrinsic duplicates of Long. Long qua just as he is, if there were such a thing, would be a truthmaker for the truth that Long is black, and for every other true intrinsic predica-

tion with Long as subject, in very much the same way that Long himself is a truthmaker* for these same truths.

If wishes were horses, we would believe in these qua-versions of things, and they would serve nicely as truthmakers for intrinsic predications. Since wishes are not horses, what reason have we to believe in these novel and peculiar entities we have just imagined?

One bad reason to believe in them is that we have suitable names for them: 'Long qua black', 'Long qua just as he is', and the like. But

(1) The existence of a suitable name is no guarantee that there is something for it to name. Presumptive instances of pseudo-reference are legion: 'Sherlock Holmes', 'the average taxpayer', 'a dearth of beer', and so on. Anyway,

(2) It is by no means clear that qua-phrases in ordinary language even purport to name anything.

Given a sentence of the form

> NP qua Adj VP

we have a choice of two parsings

> (NP qua Adj) VP
> NP (qua Adj VP),

and the second parsing, on which the 'qua Adj' is an adverbial modifier of the verb phrase, is prima facie at least as plausible as the first. But if the second parsing is right, 'NP qua Adj' is not a syntactic constituent of the sentence at all, still less an ostensible name (see Kroon 2001). Indeed, we are free to co-opt it as a name, if we already believe in something it could suitably name. But if we do, we cannot claim to be following the lead of ordinary language.

But I deny that Long qua black is a novel and peculiar sort of thing. Long qua black is none other than Long himself. Surely you are willing to believe in a cat – and that is all I ask. Likewise for Long qua just as he is; likewise, *mutatis mutandis*, for all the other qua-versions of things that serve as truthmakers for intrinsic predications.

Long qua black *is* Long, yet the two of them have different essences. How can this one thing, Long qua black/Long, be essentially black and also be only accidentally black? My answer, of course, is that he has different essences under different counterpart relations. The name 'Long' evokes one counterpart relation; the (novel) name 'Long qua black' evokes another. The counterparts of Long qua black/Long under the second counterpart relation are just those of his counterparts under the first counterpart relation that are black. (More precisely, 'Long' evokes one rather indeterminate range of counterpart relations, and 'Long qua black' evokes another. The relations

of the second range are like those of the first except with blackness built in. Thus, the vagueness that infects the question of essentialism of origins, for instance, is unaltered.) Likewise, *mutatis mutandis*, for Long qua just as he is, and all the other qua-versions of things.

Once again, just as in the cases of Lumpl and Goliath and me and my body, the ostensible multiplication of entities is replaced by an innocent multiplicity of counterpart relations. (Compare Yablo 1987, in which the acceptance of a multitude of qua-versions of things – not his term – really is a multiplication of entities.) Once we have decided that Lumpl is Goliath, there is no need to try to understand the strangely intimate relation of 'constitution' that supposedly unites these two different things. Likewise for me and my body. Likewise for Long qua black and Long *simpliciter*.

5   Toil or theft?

The solution I have proposed can be parodied to its discredit. Why not provide truthmakers for negative existential propositions in a similar fashion? Let 'Long qua unaccompanied by unicorns' be still another evocative name for Long, one that evokes a still more peculiar counterpart relation. Under this peculiar counterpart relation, something will be one of Long's counterparts just in case

(1) it is one of his counterparts under the ordinary counterpart relation evoked by the name 'Long' (pretend for simplicity that this is fully determinate); and
(2) it is unaccompanied by unicorns – that is, it is in a world where there are no unicorns.

Then Long qua unaccompanied by unicorns is a truthmaker for the negative existential proposition that there are no unicorns: any world where he exists – that is, any world where he has a counterpart under the counterpart relation evoked by the name I just gave him – is a world where there are no unicorns.

The same trick works for negative existential propositions generally, with the sole exception of the proposition that there is nothing at all.

It should be obvious that this is just a cheap trick, and does not give the friends of the Truthmaker Principle what they wanted. But why is it any worse than my own proposal for truthmakers for predications?

Answer: because the 'peculiar counterpart relation' is so very peculiar as not to be a genuine counterpart relation at all. The 'similarity', if we may call it that, between things that are unaccompanied by unicorns is, in the first place, one that would strike us in almost any context as an utterly unimportant similarity. It is, in the second place, an entirely extrinsic similarity. Two things both unaccompanied by unicorns could be as different as you please

intrinsically. Their surroundings too, both nearby and remote, could differ intrinsically in any respect other than the absence of unicorns.

Satisfactory counterpart relations, on the other hand, rest upon similarities that strike us as having at least some importance; and they rest predominantly upon intrinsic similarity. Not just on intrinsic similarity between the counterparts themselves, although that will often be part of what makes them counterparts. But a satisfactory counterpart relation will often give a lot of weight to intrinsic similarity between the contexts in which the counterparts are embedded in their worlds. For instance, in the case of match of origins, we have the intrinsic similarity of the pasts from which the two counterparts originated. (Indeed, essentialism of origins is at its most plausible when we have divergence between two possible worlds that are exact intrinsic duplicates up to about the time when the counterparts come into existence.) In the case of similarity in philosophical role, we have the intrinsic similarity of the philosophical events in which the two counterparts participate.

The alleged counterpart relation allegedly evoked by 'Long qua unaccompanied by unicorns', as well as failing to heed similarities that we would find important, also fails to heed intrinsic similarity. But the counterpart relations evoked by 'Long qua black' or, still more, by 'Long qua just as he is' place *more* weight on intrinsic similarity than the counterpart relation evoked just by 'Long'. And that is how my proposal for predications differs from the cheap trick.

6   States of affairs

Armstrong (1997) says that the truthmakers for predications are *states of affairs*, or *facts*. I want to compare this with my proposal that the truthmakers are qua-versions of the things which are the subjects of the predications. My conclusion will not be that my proposal is preferable, but rather that there is less difference between the two proposals than meets the eye – and maybe none at all.

But first we need to clear up a troublesome ambiguity. Long is black; we have the state of affairs of Long's being black, and the fact that Long is black. What would become of these entities if Long were not black? What does become of them in a world where Long's only counterpart is not black, or where he has no counterpart? What Armstrong calls a state of affairs, or a fact, is something that would not exist at all if Long were not black, and this is the conception I want to discuss.

But there is another conception, on which the state of affairs of Long's being black would still *exist* if Long were not black, but would in that case be a state of affairs that did not *obtain* (see, *inter alia*, Plantinga 1974: 44–6). It is as if 'state of affairs' meant 'proposition' and 'obtain' meant 'true'. And there is a conception on which the fact that Long is black is something that would still exist if Long were not black, but would in that case be not a fact but a

falsehood. It is as if 'fact' meant 'true proposition'. It is hard to see why 'states of affairs' or 'facts', so conceived, are anything other than propositions. They are useless as truthmakers for predications, since they would exist regardless of whether the subject did or did not have the predicated property. (The same goes for ersatz facts or states of affairs constructed set-theoretically or mere-ologically out of the subject and the property–thing–property pairs, or the like, at least if we are operating under a counterpart relation that makes the set-theoretical or mereological constitution of such a construction essential to it.) Here, let us follow Armstrong and understand the state of affairs of Long's being black to be something that would not exist at all if Long were not black, and therefore something suited to serve as a truthmaker for the truth that Long is black.

It would be nice to borrow Mellor's (1995: 161–2) unambiguous term 'factum', which means almost what Armstrong means by 'state of affairs'. But there is one difference between Armstrong and Mellor that will concern us later, so it seems best to use Mellor's term only when discussing Mellor's own theory.

What does Armstrong tell us about states of affairs, and how do they compare with our qua-versions of things?

(1) States of affairs are particulars, spatio-temporally located and unrepeatable (except for certain higher-order states of affairs that turn out to be universals in their own right and which need not concern us here, such as lawmaking relations of universals). The state of affairs of Long's being black, for instance, is located exactly when and where Long is. The same is true of our qua-versions of things. Since Long qua black is none other than Long himself, of course Long qua black is located exactly where Long is.

(2) Necessarily, the state of affairs of a's being F exists just in case thing a and property F both exist and a has F. For instance, Long's being black exists just in case Long is black. This would be a prima facie mysterious necessary connection between distinct existences, if Long and that state of affairs were distinct existences. Likewise, Long qua black exists just in case Long is black. This is a necessary connection. But it is not between distinct existences, since Long qua black is none other than Long. It is not mysterious and not objectionable. It holds just because blackness is part of what it takes to be Long's counterpart, under the peculiar counterpart relation evoked by the name 'Long qua black'.

(3) The state of affairs of a's being F is said to be composed, but not mereologically, of two constituents: the particular a and the universal F. Prima facie I cannot understand this: mereology is the general theory of composition, so 'unmereological composition' is contradictory. But what cannot be understood literally can perhaps be understood analogically, and the analogy that comes to mind is as follows. If necessary connections between distinct existences are forbidden, then mereological composition

(in which the whole is not distinct from its parts but rather is partially identical to each of them) becomes a licence for necessary connections. Maybe it means to say that a state of affairs that is unmereologically composed of its constituents bears a necessary connection to them: the necessary connection considered in the previous paragraph. If that is what the claim of unmereological composition means, we already have seen that it applies just as well to Long qua black.

(4)   We also have a denial that the state of affairs is mereologically composed of *a* and *F*. (Otherwise, Long's being black would exist if Long and blackness did, regardless of whether Long was black; at least under a counterpart relation that validates mereological essentialism.) Likewise I deny that Long qua black is mereologically composed of Long and blackness. Long, yes: he is part of Long qua black because he is the whole of Long qua black. But blackness, no.

(5)   We are not given a fully general denial that states of affairs are identical to the ordinary particulars that are the subjects of predications. Indeed, in one special case this identity is asserted. Let *F* be the complete intrinsic character of *a*, including all of *a*'s intrinsic properties, or, at any rate, all of them that are genuine universals. (I shall assume, questionably perhaps, that all the rest supervene upon these.) Let *a* be a so-called 'thick' particular, taken to include the whole of *F*. ('Include unmereologically', whatever that means.) Then the state of affairs of *a*'s being *F* is identified with *a* itself. I can match this. 'Thick' Long has the same existence conditions as Long qua *F* – that is, Long qua just as he is. So 'thick' Long, like Long qua just as he is, serves as a truthmaker for all true predications with Long as subject. And Long qua just as he is, like all other qua-versions of Long, is identical to Long himself.

So in the end, the only difference I can find between Armstrong's proposal and mine is that I claim in full generality, and Armstrong claims only in a special case, that the truthmaker for a true predication is identical with the subject of that predication. Should I conclude, therefore, that despite appearances the two proposals are almost the same? I doubt it, despite my failure to articulate the differences. Instead, I am inclined to think that the two proposals come out alike because they are constrained alike by the goal of finding truthmakers for predications.

7   Temporary intrinsics

Cat Long is black all his life. But there are other intrinsic properties, for instance purring, that things have only temporarily. Cat Ajax purrs, perhaps, throughout the three-millionth minute of his life, but not the minute before or the minute after.

Nothing new here, if we accept the hypothesis of temporal parts. There is a temporal part, Ajax throughout his three-millionth minute, for short

Ajax$_{3m}$, that has the intrinsic property of purring; and this intrinsic predication is made true in just the way that other true intrinsic predications are. I could say that the truthmaker is a qua-version of the temporal part: Ajax$_{3m}$ qua purring. Armstrong, who accepts the hypothesis of temporal parts, could say that the truthmaker is a state of affairs, Ajax$_{3m}$'s purring. Either way, the same truthmaker that makes it true that Ajax$_{3m}$ purrs, also makes it true that Ajax, a persisting cat composed of many temporal parts, purrs throughout his three-millionth minute. (Let descriptions like 'Ajax's three-millionth minute' be read as rigidified, designating in any world the time that fits that description in actuality.)

Mellor, however, does not believe that Ajax has temporal parts. He rather thinks that Ajax *endures identically*: he in his entirety is located at all the different times when he is alive, much as a saint practising bilocation, or a universal is said to be wholly present at multiple locations in space. Mellor therefore needs an account of truthmaking for temporary intrinsic predications that avoids any commitment to temporal parts. Further, he needs an account of intrinsic change that does not implicitly deny persistence altogether; that does not represent change as contradictory; that does not misrepresent temporary intrinsic properties as relations to moments of time; and that does not trade in the changing temporary intrinsic properties for the permanent intrinsic property of having such-and-such history of change. (The final option has been suggested by Parsons 2000). Mellor's ingenious solution does indeed avoid all these pitfalls, but I think it is nevertheless unsatisfactory.

Mellor (1998: 26, 91–5) gives us a theory of indiscernible facta. As previously noted, Mellor's facta are very like Armstrong's states of affairs. However, Armstrong's states of affairs are located exactly when and where their particular constituents are. Not so for Mellor's facta, in the case where the particular constituent endures identically. In that case, the factum shares only one, not all, of the many temporal locations of its particular constituent. Suppose Ajax purrs throughout his three-millionth minute. Call this time, for short, t_{3m}. (Perhaps t_{3m} should really be an instant, not a minute; but for simplicity I pretend that minutes are the smallest divisions of time.) There is a factum, Ajax's purring. This factum has two constituents, Ajax and the property of purring, but it does not have t_{3m} as a third constituent. Rather, it is located at t_{3m}. Assume that it is *essentially* located at t_{3m}. (Mellor does not say this, but it seems to be required by what he does say. It seems a safe enough assumption: similarity of temporal location is one similarity that could unite this factum with its counterpart facta in other worlds, and what countervailing differences could there be?) Then this factum is a truthmaker for the truth that Ajax purrs at t_{3m}. Necessarily, if it exists it is located at t_{3m}. (If it has a counterpart, that counterpart is located at t_{3m}, or at a counterpart of t_{3m}.) Necessarily, if it exists and is located at t_{3m} then Ajax purrs at t_{3m}.

Now suppose that Ajax purrs again at a later time, say his four-millionth minute (or some instant therein), for short t_{4m}. Again there is a factum with Ajax and the property of purring as its constituents, but this is a different

factum. It is uniquely and essentially located at $t_{4\mathrm{m}}$ rather than $t_{3\mathrm{m}}$. Yet despite their difference in location, these two facta differ not at all with respect to their constituents. In that respect, they are indiscernible. Just as the factum located at $t_{3\mathrm{m}}$ is a truthmaker for the truth that Ajax purrs at $t_{3\mathrm{m}}$, so likewise the factum located at $t_{4\mathrm{m}}$ is a truthmaker for the truth that Ajax purrs at $t_{4\mathrm{m}}$. Doubtless Ajax purrs at still other times, so we have still other facta indiscernible from these two.

These indiscernible facta are temporary, just as temporal parts would be. But they are not temporal parts, and they do not have temporal parts as constituents. Rather, they have identically enduring Ajax as their common particular constituent. It is because Ajax purrs more than once, and we need different truthmakers for different truths about when he purrs, that we need different facta with different locations but exactly the same constituents.

Is that a problem? I said that I did not understand the 'unmereological composition' of Armstrong's states of affairs; no more do I understand it in the case of Mellor's facta. Since I do not understand 'unmereological composition', I do not know what rules it ought to follow. Therefore, I know of no reason why different facta should not have the very same constituents.

The difference between Armstrong's states of affairs and Mellor's facta is slight. We should have expected some such difference given that Armstrong accepts the hypothesis of temporal parts and Mellor does not. Yet this slight difference means that I cannot use qua-versions of things to imitate Mellor's indiscernible facta in the same way that I used them to imitate Armstrong's states of affairs, or not without having recourse to the temporal parts that Mellor rejects. Ajax qua whatever you please is identical to Ajax. If Ajax endures identically, so does Ajax qua whatever you please. Helping myself to peculiar counterpart relations is not a way to conjure up temporary entities without benefit of temporal parts.

Is there such a thing as Ajax qua purring, if Ajax endures identically? Well, there is such a thing as Ajax qua permanently purring – but not in this world, and not in any world very close to this world. And perhaps there is such a thing as Ajax qua purring at $t_{3\mathrm{m}}$. Whether there is any such qua-version of an identically enduring Ajax depends on whether the hypothesis of identical endurance affords any satisfactory account of temporary intrinsic properties, something I still doubt. But if there is such a qua-version, then it is a truthmaker for the proposition that Ajax is purring at $t_{3\mathrm{m}}$. Every world where this qua-version of Ajax exists – has a counterpart – is a world where Ajax is purring at $t_{3\mathrm{m}}$. Under the same proviso, there is another qua-version of identically enduring Ajax, Ajax qua purring just when he does, that can serve as a truthmaker for all truths about when he is purring and when he is not.

I said against Armstrong's states of affairs that they prima facie involved mysteries of unmereological composition and of necessary connection between mereologically distinct existences. (Perhaps these two mysteries are one and the same.) The same complaint applies against Mellor's facta. In Armstrong's case the complaint can be dodged if we interpret states of affairs as qua-

versions of their particular constituents. (Most likely this interpretation is unintended.) In the case of Mellor's indiscernible facta, there is no parallel way to dodge. My complaint stands; and that is why I doubt that Mellor has given us a fully satisfactory treatment of temporary intrinsic predications under the hypothesis of identical endurance.

But maybe there is another way to dodge the complaint. Mellor does not reject the hypothesis of temporal parts across the board; rather, he thinks that events have temporal parts and things – cats, for instance – do not. So perhaps we can use the temporal parts Mellor accepts as proxies, so to speak, for those he rejects. Ajax's life is one prolonged event, and presumably does have temporal parts. One of these life-parts, call it 'life-part$_{3m}$' occupies the three-millionth minute of Ajax's life. It has a property we can call purring*. (Not purring – it is Ajax himself, not a part of his life, that purrs – but a property that is somehow necessarily connected with purring.) 'Life-part$_{3m}$ qua purring*' can be taken as a name for life-part$_{3m}$ that evokes a peculiar counterpart relation with purring* built in; life-part$_{3m}$ qua purring* is a truthmaker for the truth that life-part$_{3m}$ is purring*; and that somehow – how? – implies that Ajax himself purrs at t$_{3m}$. The idea is that qua-versions of parts of lives (more generally, of histories, since not all things are alive) might imitate Mellor's indiscernible facta in roughly the way that qua-versions of things imitated Armstrong's states of affairs. I find this solution unsatisfying:

(1)   because, just as Mellor fears, I have some difficulty understanding the supposed distinction between Ajax's life and Ajax himself;

(2)   consequently, I have some difficulty understanding the distinction and the connection between purring and purring*; and

(3)   it is disappointing that a way of rejecting the hypothesis of temporal parts should succeed only because the rejection is not whole-hearted.

Postscript to 'Things qua truthmakers': negative existentials

Gideon Rosen and David Lewis

So far, Lewis has granted that true predications do after all have truthmakers. But he does not yet accept the Truthmaker Principle in full generality, because he still doubts that true negative existentials have truthmakers. But if Lewis's proposal to take qua-versions of things as truthmakers will work at all – in other words, if we are entitled to take ordinary things as truthmakers by supposing that they make propositions true relative to the peculiar counterpart relations that are evoked by peculiar names for those ordinary things – then his proposal can be extended to the case of negative existentials.

We should not take cat Long qua unaccompanied by unicorns as a truthmaker for the truth that there are no unicorns. That was indeed a cheap trick, for the reason Lewis said: the requisite 'peculiar counterpart relation' is no genuine counterpart relation at all, being founded on an unimportant and unduly extrinsic respect of similarity. But if we take a qua-version of a better-chosen thing, we can use a much more satisfactory counterpart relation.

Begin with an easy case: restricted negative existentials, such as the truth that there are no unicorns in this room. (In this room now, but let that restriction remain tacit.) Let this room$^+$ consist of this room together with everything in it: the air, the furniture, the unicorns if any, … . This room$^+$ qua including no unicorns is a truthmaker for the truth that there are no unicorns in this room. This time, the peculiar counterpart relation evoked is founded on an entirely intrinsic and salient respect of similarity. But we could instead have used this room qua containing no unicorns; the counterpart relation is still satisfactory, being founded on intrinsic similarity not between the counterparts themselves – the rooms – but between more inclusive things – rooms$^+$ – that are saliently related to the counterpart rooms.

Likewise, *mutatis mutandis*, for the less restricted negative existential truth that there are no unicorns on this planet; or even the truth that there are no unicorns in this galaxy; or even the truth that there are no unicorns in this galaxy throughout its history.

For unrestricted negative existentials, such as the truth that there are no unicorns anywhere, ever, we can take as truthmaker a qua-version of the entire world: the totality of everything there actually is. That way, our counterpart relation can again be founded on intrinsic similarity.

What is a counterpart of the world? Must it be an entire possible world, the totality of all there is in its world? (In that case, a counterpart of the actual world in a world *W* would have to be the world *W* itself, nothing less.) Or might it be just a proper part of a world? For instance, might our four-dimensional world have as a counterpart a four-dimensional slice of some five-dimensional world? We suppose this is one of those questions about 'the' counterpart relation that has no determinate answer; in other words, there are counterpart relations under which the world is essentially total, and there are counterpart relations under which it is not. But for present purposes, we need to consider counterpart relations under which the world is essentially total. 'The entire world' or 'the world qua total', or 'the world qua unaccompanied' can be taken as names for the world that evoke such counterpart relations.

Is the counterpart relation evoked by such names a satisfactory one? We think so. Being unaccompanied is an extrinsic property, to be sure (Lewis 1983; Langton and Lewis 1998). So similarity in respect of being unaccompanied is an extrinsic respect of similarity. However, the property of being completely unaccompanied (unlike Long's property of being unaccompanied by a unicorn) does seem quite important to the character of anything that has it. Further, it is nomologically linked to quite an important intrinsic property: being, at least ostensibly, self-contained. Because the world is completely unaccompanied it will never, short of a miracle, be affected by signals or visitors suddenly arriving as if from elsewhere.

Besides making the world essentially total, we can impose further conditions on the evoked counterpart relation by adding further qua-phrases in our usual way. For instance, the entire world qua lacking unicorns, under the counterpart relation evoked by the name we just gave it, is (1) essentially total and (2) essentially without unicorns. If indeed the world does lack unicorns, this evocative name is just another name for the world. We propose that the entire world qua lacking unicorns is a truthmaker for the negative existential truth that there are no unicorns anywhere, ever.

The proposal can be repeated for other negative existential propositions, with one exception: the proposition that there are no contingent things at all, not even the world. If indeed that proposition could be true, it would have to be a truth without a truthmaker – for if it were true in virtue of some truthmaker, never mind what, never mind under what counterpart relation, then there would be something and not nothing.

Another truthmaker for the truth that there are no unicorns, and indeed for all other negative existential truths, and indeed for all truths without exception, is the entire world qua just as it is. The counterparts of the world under the peculiar counterpart relation evoked by this name are just those entire worlds that are intrinsic duplicates of the actual world.

Recall that Lewis left open the question of whether there are indiscernible worlds. If there are not, then the actual world itself is the only counterpart of the entire world qua just as it is. So we may well suspect that the Truthmaker

Principle has been trivialized in an unintended way: the proposition – *any* proposition – is true in all worlds where the truthmaker exists because (1) it is true in this world and (2) we have chosen the truthmaker so as to make sure that there are no other worlds where it exists! If, on the other hand, there are indiscernible worlds, then the evoked counterpart relation is not identity but indiscernibility, and so our sense of trivialization should diminish. Anyhow, no parallel suspicion can arise against our first proposal that the truthmaker for the truth that there are no unicorns is the entire world qua lacking unicorns. In that case, the counterparts of the world under the evoked counterpart relation are many and varied.

In Armstrong's (1997: 134–5, 196–201) scheme of things, the truthmakers for negative existential propositions are *totality facts*. These are special states of affairs of the form $T(X)$, where T is a property (perhaps higher order) of totality and X is something (perhaps not a particular) that has this property because it is exhaustive, all there is. Or they may have the form $T(X,Y)$, where T is a totality relation and X and Y stand in this relation because X exhausts Y. We need only consider the easiest case: $T(a)$, where a is a particular and $T(a)$ is the state of affairs of a's being exhaustive.

Now if a is going to be exhaustive, a had better be an especially big particular: the entire world. And it must be the world considered as a 'concrete' particular, the cosmos, not some sort of 'abstract' entity, such as a linguistic or mathematical or propositional representation of the cosmos, or a structural property instantiated by the cosmos. [It does not matter for present purposes whether we believe, with Lewis (1986), that unactualized *cosmoi* exist, or whether we believe, with Rosen (1990; 1995), that they are fictitious.] And let a be the world as a 'thick' particular, identified with the state of affairs $F(a)$, where F gives the complete intrinsic character of a. The totality fact $T(a)$ is a citizen in good standing of Armstrong's world of states of affairs; and by his lights, it should be a truthmaker for all negative existential truths, all true predications having the world or its parts as subjects and all other truths as well. We note that $T(a)$ has just the same existence conditions as the entire world qua just as it is: necessarily, it exists (it has a counterpart) just in case an exact intrinsic duplicate of the actual world both exists and is exhaustive. So Armstrong, at any rate, dare not say that it trivializes the Truthmaker Principle to take the entire world qua just as it is as a truthmaker for all truths. The parallel with $T(a)$ would be too close for comfort.[1]

Note

References

Armstrong, D. M. (1989) 'C. B. Martin, counterfactuals, causality, and conditionals', in J. Heil (ed.) *Cause, Mind, and Reality: Essays Honoring C. B. Martin*, Dordrecht: Kluwer.

—— (1997) *A World of States of Affairs*, Cambridge, UK: Cambridge University Press.

Bigelow, J. (1988) *The Reality of Numbers: a Physicalist's Philosophy of Mathematics*, Oxford: Oxford University Press.

Bricker, P. (1999) 'The world: facts or things?' lecture presented at New York University, February 1999.

Fox, J. (1987) 'Truthmaker', *Australasian Journal of Philosophy* 65: 188–207.

Gibbard, A. (1975) 'Contingent identity', *Journal of Philosophical Logic* 4: 187–221.

Kroon, F. (2001) 'Parts and pretense', *Philosophy and Phenomenological Research* 61: 543–60.

Langton, R. and Lewis, D. (1998) 'Defining "intrinsic"', *Philosophy and Phenomenological Research* 58: 333–45.

Lewis, D. (1968) 'Counterpart theory and quantified modal logic', *Journal of Philosophy* 65: 113–26.

—— (1971) 'Counterparts of persons and their bodies', *Journal of Philosophy* 68: 203–11.

—— (1983) 'Extrinsic properties', *Philosophical Studies* 44: 197–200.

—— (1986) *On The Plurality of Worlds*, Oxford: Basil Blackwell.

—— (1988) 'Statements partly about observation', *Philosophical Papers* 7: 1–31.

—— (2001) 'Truthmaking and difference-making', *Noûs* 35: 602–15.

Martin, C. B. (1996) 'How it is: entities, absences, and voids', *Australasian Journal of Philosophy* 74: 57–65.

Mellor, D. H. (1995) *The Facts of Causation*, London: Routledge.

—— (1998) *Real Time II*, London: Routledge.

Mulligan, K., Simons, P. and Smith, B. (1984) 'Truth-makers', *Philosophy and Phenomenological Research* 44: 287–321.

Parsons, J. (1999) 'Is there a "truthmaker" argument against nominalism?', *Australasian Journal of Philosophy* 77: 325–34.

—— (2000) 'Must a four-dimensionalist believe in temporal parts?', *The Monist* 83: 399–418.

Plantinga, A. (1974) *The Nature of Necessity*, Oxford: Oxford University Press.

Quine, W. V. (1976) 'Worlds away', *Journal of Philosophy* 73: 859–63.

Rosen, G. (1990) 'Modal fictionalism', *Mind* 99: 327–54.

—— (1995) 'Modal fictionalism fixed', *Analysis* 55: 67–73.

Yablo, S. (1987) 'Identity, essence, and indiscernibility', *Journal of Philosophy* 84: 293–314.

3 Deflationism

The facts

Peter Smith

1

Ramsey, many of us think, is on to something about truth. What it takes for it to be true that Caesar is dead is no more than that Caesar is dead. What it takes for it to be true that Gwyneth is beautiful is no more than that Gwyneth is beautiful. What it takes for it to be true that 7 is prime is no more than that 7 is prime. And so it goes. There are stories to be told about the metaphysical commitments of our temporal talk, aesthetic talk, arithmetical talk (and we might well expect that these will be interestingly and importantly different stories). But there is no additional, overarching story to be told about the further metaphysical commitment we take on when we say that it is *true* that Caesar is dead, or *true* that Gwyneth is beautiful, or *true* that 7 is prime. There just is no metaphysical weight to the concept of truth.

Indeed, the concept of truth arguably carries little weight of any other kind either. Say, if you will, that truth is a norm of assertion. But that is just compendiously to endorse each instance of a schema such as:

> One should, *ceteris paribus*, assert *p* only if *p*.

The normativity here attaches to the instances of the schema (and those do not involve the concept of truth). Say, if you will, that truth is a norm distinct from warranted assertibility. But that just combines the previous compendious endorsement with a reminder that there can be correct instances of the schema:

> *p* is warrantedly assertible but, even so, *not-p*.

Say too, if you will, that the success of a theory is (often) explained by its truth. But again that says no more than that there are many instances of the schema:

> (Belief in) the theory that *p* works well because *p*.

The concept of truth, in sum, carries no normative or explanatory weight of

its own, at least according to thorough-going minimalists. Still, these further deflationary claims perhaps go beyond the initial rejection of a distinctive metaphysical loading to the concept of truth; and it is that widely shared metaphysical deflationism about truth which is my concern here.

Mellor, many of us think, is on to something about facts. Serious metaphysics means taking facts seriously – where facts are not mere true propositions (at least if those are conceived of as entities in the domain of sense, the abstract correlates of true sentences). Rather, they are complexes whose constituents are worldly objects together with worldly properties and relations. Note that not just any old gerrymandered scattered sum of things counts as an object in the sense that matters here (a sense that needs explication but commands intuitive allegiance). Likewise, not just any old gerrymandered extension is the extension of a real property in the sense that matters (again, a sense that needs explication but commands intuitive allegiance). We *need* a distinction between real objects and (say) arbitrary mereological sums if we are to talk sense about identity and change and other central themes of metaphysics. We *need* a distinction between real (elite, sparse) properties and the multitude of gruesomely disjunctive second-raters if we are to talk sense about similarity, change, causation and some more central themes of metaphysics. Objects arguably do not come bare, *sans* properties; and, according to some, properties, the actual as opposed to possible ones, do not come uninstantiated. But be that as it may: whether or not there are bare particulars and uninstantiated but existent properties, it is of the nature of genuine objects and properties to be apt to combine into existent states of affairs, into worldly facts. Taking sparse objects and sparse properties seriously means taking sparse facts seriously.

On the face of it, it looks as if Ramsey and Mellor are pulling in different directions here. For does not metaphysical deflationism about truth require rejecting talk of facts (if that comes to any more than anodyne talk of truths, i.e. true propositions)? Conversely, if our metaphysics countenances worldly facts, then we have items in our ontology that are *truthmakers*, items whose existence makes what we say true, when it is. These truthmakers do not line up one-to-one with the truths we utter. For example, the fact that Caesar is dead – if that will do as a sample – not only makes it true that Caesar is dead but also makes it true that either Caesar is dead or I am a Dutchman. We do not need a disjunctive fact to make the disjunction true. So the modern enthusiast for sparse facts will not want to reinstate a traditional one-to-one correspondence theory. But in the wider scheme of things, that is not a big deal. The old-style correspondence theorist loses that battle but (you might suppose) has won the war, once we countenance facts.

So, still on the face of it, we are faced with an uncomfortable choice. This is not just uncomfortable for those of us who admire both Ramsey and Mellor (and learnt to admire the one from the other). For the apparent tension is, of course, between a whole raft of broadly deflationist views about truth, and a whole cluster of positions in modern metaphysics. We can pick and mix various popular views from the theory-of-truth side with various popular views from

the metaphysics side, and we still end up with the same problem. Agreed, the issue does not arise for those immune to the attractions of metaphysics or to the attractions of some species of deflationism – and so it is that some do write about truth never mentioning 'facts', and there are metaphysicians who just assume that deflationists are not taking truth seriously. Neither position will seem much better than point-missing to those whose insights (as they take them to be) are being ignored.

However, perhaps we can do better. Perhaps, despite those first appearances, we *can* give due acknowledgement to the arguments on both sides, and allow both the deflationist about truth and the enthusiast for facts what they want. This chapter is about the prospects of pulling off the balancing act.

2

Three preliminary points. First, there are superficially similar cases where a kind of minimalism and a co-ordinate substantive theory can peaceably coexist. For example, it is plausible to say that a grasp of the concept *red* requires little more than an ability to use the concept in appropriate recognitional judgements (and, for example, to accept colour attributions on the basis of memory and testimony too, thus distinguishing *it is red* from *it is currently looking red*). But such a minimal theory of the sense of colour words is surely consistent with a much more substantive account of what colours are (and this account will feature in an explanation of why it is apt for us to have thin, basically recognitional, concepts of colour properties, as well as the more articulated concepts of the same properties embedded in our substantive theory). Can we take a parallel line here, and argue similarly that a minimalist account of the *sense* of 'true' is consistent with a non-minimal account of the *reference*, i.e. a substantive theory of what truth in general consists in?

For various reasons, the model of *red* is not at all a promising one. For a start, the reason we need an account of what redness consists in is because we need a causal story – in fact, it is a collection of disparate causal stories – about what our visual system is tracking when we successfully make the experiential judgement that something is red (for there is no magic here). But where is the need for any analogous story to explain what in general we are tracking when we judge that something is true? So long as we have (say) the disposition to pass from one side to the other of any instance of disquotational biconditionals, no deeper causal story is needed to explain our competence with 'true', and certainly no story about some cognitive engagement with a distinctive property of truth.

Second, taking sparse facts seriously does not mean taking them to be *basic*. Maybe there is something to be said for a Tractarian ontology, according to which the world is, ultimately, all that is the case (the totality of sparse facts), and objects and properties alike are in some sense abstractions from the facts. That view has the merits of sidestepping a familiar putative difficulty with views that take objects and universals as the two basic kinds of entity – for how, it is

asked, do they get combined into a fact? Not, for familiar reasons, by standing in a *relation*: so we have to postulate some kind of irreducible non-relational tie between objects and the universals they instantiate. And this, some say, is difficult to understand. Others will feel that taking facts as basic and treating objects as some kind of abstraction from the facts in which they feature does not give chunky physical objects the right kind of status (as if, perhaps, we are in danger of assimilating the being of all objects to that of other abstracta like Frege's directions). But we just do not need to tangle with that kind of debate now. Which is fortunate, as the rules of engagement for debates about what is metaphysically 'basic' are obscure, to say the least.

Thus, suppose we agree with Mellor that causation involves facts (the sparse, worldly facts – 'facta' as he calls them), and that it will not do, for example, to treat causation as relating tropes, for tropes do not have enough structure. That in itself does not rule out analysing facts as complexes of (sparse) objects and properties, and then treating those in turn as each constituted in different ways by suitably structured collections of tropes. Maybe, then, tropes are the alphabet of being, and the facta are (so to speak) rather long paragraphs. Or changing tack, perhaps some will prefer to treat sparse properties as elite sets of objects drawn from many possible worlds, and ultimately construct facta from cross-world collections of objects. For the present, it really does not matter what our favoured metaphysical story is, so long as it has the resources to make sense of talk of sparse worldly objects and properties (if only by construction), and thus make sense of talk of sparse facts (if only by further construction). Because once we do have facts in our story of the world, however they are to be further analysed, surely they should enter into our story about truth? *That* is the basic challenge to the metaphysical deflationist about truth.

Third, it certainly is not settled how best to frame a theory of truth that is both formally competent yet also uncontentiously deserving of the label 'metaphysically deflationist'. Again, it turns out that the details mostly do not matter for our problem, but it is worth pausing to say more about this.

To help fix ideas, take the theory *PA*, i.e. first-order Peano Arithmetic, whose language is *L*. Extend *L* to *L*⁺ by adding both a construction $\langle\ldots\rangle$, which forms terms from wffs (well-formed formulae) of *L* (with the intended interpretation that $\ulcorner\langle\varphi\rangle\urcorner$ denotes φ), and also a new predicate 'Tr'. Let *MT* be the set of instances of the T-schema

$$\mathrm{Tr}\langle\varphi\rangle \equiv \varphi,$$

where φ is a closed wff of the original *L*. *PA* + *MT* might thus be advertised as arithmetic plus a theory of truth that captures in a minimal way the thought that there is no more to the idea of (arithmetic) truth than is given by the requirement that arithmetic instances of the T-schema hold.

PA + *MT* indeed involves a very modest theory of truth. For example, as you would expect if the truth-predicate really is just akin to a disquotational device, *PA* + *MT* is conservative over *PA* (i.e. no *L*-wff, not already provable

from *PA*, is provable from *PA* + *MT*). The trouble is that *MT* looks *too* modest. For any closed *L*-wff φ, we have both

$$MT \vdash \mathrm{Tr}\langle\varphi\rangle \equiv \varphi$$

$$MT \vdash \mathrm{Tr}\langle\neg\varphi\rangle \equiv \neg\varphi,$$

and hence

$$MT \vdash \mathrm{Tr}\langle\neg\varphi\rangle \equiv \neg\,\mathrm{Tr}\langle\varphi\rangle.$$

However, while we can prove each instance of $\mathrm{Tr}\langle\neg\varphi\rangle \equiv \neg\,\mathrm{Tr}\langle\varphi\rangle$, we cannot yet even express, let alone prove, the generalization that a negated wff of *L* is true if and only if the original wff is not true. Now, the expressive lack is easily repaired. Extend L^+ by adding the functor 'neg', where, for each wff φ of *L*, we have as a syntactic axiom

$$\mathrm{neg}\langle\varphi\rangle = \langle\neg\varphi\rangle$$

and add too, perhaps, a predicate 'sen', where for each closed wff φ of *L* we have the syntactic axiom

$$\mathrm{sen}\langle\varphi\rangle.$$

And we can now, in this extended language, frame a generalization *N* about negation thus:

$$\forall x(\mathrm{sen}\,x \rightarrow (\mathrm{Tr}\,\mathrm{neg}\,x \equiv \neg\,\mathrm{Tr}\,x)).$$

But even with the syntactic axioms *S* in play, we do *not* have

$$PA + MT + S \vdash N.$$

Why so? The basic idea is to take a 'natural' model for *PA* + *MT* + *S*, add a rogue element α to the domain and extend the interpretations in the natural model so that α is in the new extension of 'sen' while the new interpretation of 'neg' maps α to itself, and this model will still satisfy *PA* + *MT* + *S* while falsifying *N*.

It is sometimes said that the truth-predicate is just a formal device of disquotation *and* that a major point of having such a truth-predicate is to be able to frame generalizations (such as that a negation is true just so long as its un-negated counterpart is not) which it would otherwise need infinite conjunctions to express. But now we can see that the two halves of this claim do not quite chime together. For the minimal rules governing a mere disquotational

device (even given the needed syntactic resources) do not by themselves entitle us to make the desired generalizations.

This shortcoming of *MT* was long ago noticed by Tarski, and we have learnt from him one way of doing better – namely replace *MT* (plus the syntactic extras) with a full Tarskian theory of truth, *TT*. This certainly allows us to derive the laws of truth like *N*. But, from a deflationist perspective, the price is high. To take a dramatic example, it is familiar that

Not $[PA \vdash G]$,

where *G* is a standardly constructed Gödel-sentence for the given version of *PA*. However, we also have[1]

$PA + TT \vdash G.$

So *TT* is *not* conservative over *PA*. But a theory that enables us to deduce new truths in an old domain can hardly be said to be unsubstantial, minimal or fully deflationary.

Still, it might perhaps be said that the Tarskian theory remains *metaphysically* deflationary, even if not maximally deflationary in other ways. But is that entirely right? To be sure, a Tarskian truth-theory is blind to any metaphysical difference between the truth-conditions for 'Caesar is dead' and 'Gwyneth is beautiful' and or '7 is prime' (thus, the base clauses for the predicates '… is dead', '… is beautiful', '… is prime' treat them exactly on a par). And the truth-theory does not balk either at delivering, in the same indiscriminate way, T-biconditionals for 'This emerald is grue' and 'Caenyth is gappy' (where Caenyth is the mereological sum of Caesar and Gwyneth). But still, being metaphysically quite undiscriminating is not the same as carrying no metaphysical baggage at all. There is the non-trivial additional set-theoretic apparatus of the Tarskian theory for a start.

In sum, *MT* is certainly deflationary, but is too weak to establish generalizations like *N*. By contrast, *TT* will prove *N*, but arguably rather too much else besides, and is arguably not fully deflationary. Can we steer between? Maybe, to revert to an idea that Tarski considers and dismisses, we could try allowing an ω-rule, so that *N* holds given each instance of $(\mathrm{sen}\, \varphi \rightarrow (\mathrm{Tr}\, \mathrm{neg}\, \varphi \equiv \neg\, \mathrm{Tr}\, \varphi))$. Logicians, interested in finitary proofs, are generally dismissive of invoking ω-rules – in arithmetic as well as in truth-theories. Metaphysicians, interested in what fixes what, need to be a lot more tolerant of infinitary determination relations anyway, so perhaps they could look more kindly upon ω-rules in either case. I speculate, at any rate, that they are the best hope for the theorist of truth who wants to be maximally deflationary.

But again, let that fall out as it may (I shall not say any more here). For even if we after all buy the familiar, full-blown Tarskian works, that theory, as just remarked, falls far short of any kind of metaphysical commitment to discriminating genuine (as opposed to falsified) objects, sparse (as opposed

to abundant) properties, or the facta they compose. So we would *still* be faced with the apparent tension between the relatively thin and undiscriminating commitments of our truth-theory, and any serious metaphysics of sparse facts, properly so called. And let our ontology include what items it may; if none are especially connected to truthmaking, then none, surely, deserves the appellation 'facts'. So, should not real facts (if such entities there be) matter for real truth, *pace* the deflationist?

3

Ramsey famously remarks that there is 'no … problem of truth', that is to say no separate problem once we have 'analysed judgement'. But analysing judgement – or, as we would now put it, giving a theory of content – is, of course, highly non-trivial. And here, perhaps, there *is* ample room for sparse facts to feature centrally in a plausible causal–naturalist theory – or, better, to feature in distinctly different ways in theories of different types of content. And that opens up the possibility that it is the theory of content rather than the theory of truth which gets the facta into the picture, in an account of what makes certain true judgements true.

Here is another thought. It is, on the Armstrong–Mellor view, the business of science, not of a priori reflection, to determine what objects there are, what sparse properties there are and hence what the facta are. In particular, the sparse properties are those that feature in the contingent laws of nature, and science is how we get at these laws. So it looks as if the sparse facts – whose ingredients on the current view exist contingently – will at most be apt to make true the contingent empirical truths. For what has, for example, the necessary primeness of 7 got to do with which concrete facta do or do not exist? Perhaps, then, the so-called Truthmaker Principle (that truths need the existence of something worldly to make them true) is better construed as reflecting a view about what it is to be a *contingent* 'brute fact' rather than a general view about truth per se. (And although the facta perhaps fix the cast of Gwyneth's features, we might also wonder – if we are good Humeans – whether they fix that she is beautiful. Which is not to deny her beauty, but to wonder whether it is appropriate to think in terms of the facta entailing the aesthetic value we find here.)

These two lines of thought – each hinting at a resolution of the tension we located – can be happily brought together following a thought already to be found in Ramsey.[2] For certain beliefs, the content of the belief is that *p* just if, for any appropriate desire, actions caused by that belief combined with a desire will be successful in realizing the desire's object just in case that *p*. And of course, there is no magic about the relation between its being the case that *p* and successful action: it will be a causal condition for success. Thus, a belief is the belief that the ice-cream is in the freezer, if actions caused by that belief in combination with an appropriate desire are successful (get me ice-cream if that is what I want; let me avoid ice-cream if *that* is what I want, etc.) just

in case the ice-cream is indeed in the freezer. The ice-cream's being in the freezer will be a causal condition of getting the ice-cream by the action-path I take (or of avoiding the ice-cream, if that is what I want). Now, this kind of Ramseyian 'success semantics' may succeed for many types of singular factual content. It could work for general beliefs too. To have a general belief *All As are Bs* is to have a disposition to believe *x in a B* if you believe *x in a A*. And actions which that disposition (in company with other singular beliefs and a desire) causes will be successful – given satisfaction of the conditions associated with the singular beliefs – just so long as all *A*s are indeed *B*s. This kind of story looks a good deal less promising, however, for (say) arithmetical truths. The number 7 is prime in just the same situations that 6 is a perfect number, i.e. in every possible situation: and neither is a causal condition. So just how can the condition that 7 is prime (as opposed to the condition that 6 is perfect) enter differentially into the conditions for success for actions generated by the belief that 7 is prime? And, depending on how we modalize the rule for content, it is not clear either that the success semantics strategy works for the belief that Gwyneth is beautiful (it is enough for successful action, e.g. when I aim to pick a beautiful actress in the actual and near worlds, that Gwyneth has one from some disjunctive range of casts of features: but the content of the belief – if it is simply a belief – is not, or is not just, that she has such features).

However, take a case where success semantics *does* apply. Suppose my belief B is a state such that it generates successful action just if it is the case that p; and suppose also that this success condition obtains, i.e. suppose p. Then (by the rule for content) B is the belief that p; and so it is the case both that I believe that p and that p; and hence (now invoking nothing more than a minimalist theory of truth) my belief is *true*. The condition that gives the content would be, should I act on the belief, a causal condition for success. But causal conditions involve facta (genuine objects instantiating genuine properties). So the obtaining of the causal condition p requires the existence of relevant facta. Or to put it summarily the other way about, relevant facta must obtain if B is to be true. Call that a Truthmaker Principle by all means. But it is warranted by the specific theory of content for empirical beliefs, *not* by the general theory of truth, which can remain fully deflationary.

Of course, while the story goes particularly smoothly for success semantics, other naturalistic causal theories of content can potentially deliver the same result. We just need a story about content that associates what it is for a belief-state to be a state of believing that p to the causal condition that p. And once that link is in place, both facta (because it is a causally salient condition) and truth (because if we have p and the belief that p, then we have a true belief) have entered the story. In sum, just as the initially noted tension was insensitive to the fine details of our preferred version of deflationism and of the metaphysics of facts, the resolution of the tension is in key respects insensitive to the fine details of the theory of content.

4

An apparently happy and easy reconciliation, then: it seems that we can cleave to deflationism, but still allow the sparse facts to play a key role in the story about empirical (we might say 'factual'!) truth.

Compare the kind of pluralism about truth that Crispin Wright has articulated.[3] He argues that being a concept of truth is a matter of satisfying various constraints – and there are various concepts that apply in different domains but which variously satisfy the constraints. I want to acknowledge a pluralism, but a more familiar one, a pluralism about types of content. Then a single concept of truth applies across the different types of content. Why prefer this way of putting it? For Ramsey's reason, i.e. because after the theory of content has done its work, there is no further substantive task of explaining, in general terms, what it is for the contents to be true. (This would be a cheat if, in giving a theory of content, we always smuggle in the notion of truth again: naturalistic theories like success semantics aim to show why this is not so.)

What are the problems? Or rather – since this is not the place (a) to further defend deflationism or (b) to take on those who do not see why science gets to be the privileged arbiter of ontology – what problems arise, assuming the dual framework, (a) plus (b)?

There are problems about further developing the story. We would, for example, need to articulate a theory of content for value propositions, so that we better understand how it can be acceptable to judge that Gwyneth is beautiful (and hence to judge that it is *true* that Gwyneth is beautiful) although – plausibly – it is not a factum that she is beautiful, nor is it straightforwardly entailed by the facta. A Blackburnian projectivism would nicely fit the dual framework: but it is not without difficulties.

There are worse problems in trying to articulate a theory of content for arithmetical propositions, so that we better understand how it can be acceptable to judge that 7 is prime (and hence to judge that it is *true* that 7 is prime) although – plausibly – there are no arithmetical facta. Neo-logicism retains its independent attractions as a story about the necessity of arithmetic; but the metaphysical underpinnings of that approach do not seem compatible with Mellorian metaphysics. In a slogan, for the neo-Fregean, the existence of numbers and numerical properties consists in the truth of various arithmetical propositions, and truth precedes being: for those who see metaphysics the antipodean way, that is get things upside down. We can alternatively spin stories about numbers as higher-order universals, etc.; but it then becomes unclear why arithmetic should be necessary. (Given that, at the first order, what universals there are and what relations they stand in are contingent, where does the necessity of relations at higher orders come from?)

But these problems are further down the road. There are more immediate worries, about just what properties – and hence, just what facta – there are. Take again the condition needed for the success of the actions caused by the

belief that the ice-cream is in the freezer, namely the ice-cream's being in the freezer. Is it a *factum* that the ice-cream is in the freezer? Surely not. Neither being ice-cream nor being a freezer is a sparse property featuring in some law of nature (at least, not if 'law of nature' is understood in anything like its normal sense). But then, what *is* the relation between the facta and the ice-cream's being in the freezer? I cheerily said before that the condition, being causal, 'requires the existence of relevant facta'. But that was arm-waving: we must do better!

Some will invoke talk of a supervenience relation between the gross condition and the facta, as if that answers anything. But I share Mellor's impatience with this – 'supervenience' labels the problem, not the solution. Elsewhere, when discussing the relation between the physical properties and those recognized by the special sciences, Mellor holds that the ties (e.g. between brain-properties and being in pain) are just more contingent laws. But that cannot be the move to make here, not (at any rate) if we were right that there are no laws of nature about ice-cream and freezers. And the relation we want is not contingent anyway (if God fixes *all* the facta, then surely there isn't something else left to do, namely fix that the ice-cream is in the freezer).

A hopeful thought is that something *definitional* will do the trick. Take some platitudes about what freezers do: something is, analytically, a freezer if it has features such that the platitudes hold. At a first pass, we might locate suitable functional features still at some level above the sparse properties (as when we might speak, for example, of a 'moderate-size compartment' – i.e. one larger than a pin-head, smaller than a planet, etc.). But these features too we might hope to be able to analyse in some way, in terms of existential quantifications over some more basic features, and so on down until we do get quantifications instantiated by sparse properties. There is a familiar and not untroublesome programme for analysis here, but it seems we need a promissory note that it can be pulled off – for it is difficult to see what less will do the trick of tying common-or-garden causal conditions to the facta in a perspicuous way.

But at least Mellor's anti-physicalism makes the programme look feasible. For consider that familiar line of thought that the 'laws' of a special science like biochemistry are not strictly so called, because there is too much that is not in the purview of biochemistry that can interfere to disrupt the general correlations that hold *ceteris paribus*. The 'laws' of biochemistry, the story continues, depend on more basic chemical 'laws', which in turn hold *ceteris paribus* in virtue of physical 'laws' governing molecules, which are made true by … And the reasoning that sets this regress going will only bottom out with the ultimate laws of fundamental physics (assuming there are such). But now it only seems proper caution to be rather pessimistic about whether we have yet got a very firm handle on the really fundamental laws (if those are conceived at the level of, say, quantum field theory and general relativity, let alone quantum gravity). And while we should be chary about taking lessons in philosophy from physicists, it is notable that they routinely conceive themselves, when things get to this level, as in the enterprise of building idealized mathematical

models which in some way capture the essential physics, and would perhaps be pretty surprised to have their model-building activities given the ultimate say in ontology. Those tempted by the kind of physicalism that yields the thought 'fundamental physics is metaphysically fundamental' are perhaps perilously close to making the facta noumenal – we-know-not-quite-whats that yet supposedly make everything else factual true.

For Mellor, by contrast, the facta start much closer to home, with the things and properties recognized by much more humdrum science – science whose epistemology is a good deal more secure (by the standards of a naturalistic reliabilism) than fundamental physics. The regress to the ill-understood foundations of physics is blocked early, and the programme of analysis we sketched above has a much shorter route to travel. That is not a knock-down consideration: but it is one reason for suspecting that friends of facts should find comfort in Mellor's anti-physicalist metaphysics. Now this does indeed require getting clear – and perhaps clearer than yet we are – about the nature and status of the humdrum laws and their metaphysical commitments to the capacities of complex things: but that is another story.[4]

Notes

1 Here is a proof sketch (in a very summary form due to Jeffrey Ketland). We know that a provability predicate 'Prov$_{PA}$' is expressible in PA. By the fundamental property of the Tarskian theory, $PA + TT \vdash \text{Prov}_{PA} \langle \varphi \rangle \rightarrow \text{Tr} \langle \varphi \rangle$ for each closed φ of L. Hence, in particular, $PA + TT \vdash \text{Prov}_{PA} \langle G \rangle \rightarrow \text{Tr} \langle G \rangle$; hence $PA + TT \vdash \text{Prov}_{PA} \langle G \rangle \rightarrow G$. But by construction, if we have linked $\langle \dots \rangle$ to Gödel numbering, we will have $PA \vdash G \leftrightarrow \neg \text{Prov}_{PA} \langle G \rangle$. Whence, by simple logic, $PA + TT \vdash G$.
2 Developed in Whyte (1990).
3 In Wright (1992).
4 I would like to thank Hugh Mellor immensely for, among so many things, his warm support and encouragement throughout my dozen years editing *Analysis* (a journal that aims, indeed, to promote the characteristic Mellorian virtues of concision, clarity and straight talking).

References

Whyte, J. T. (1990) 'Success semantics', *Analysis* 50: 149–57.
Wright, C. (1992) *Truth and Objectivity*, Cambridge, MA: Harvard University Press.

4 Truth and the theory of communication

Chris Daly

1 Introduction

On one view of language use, language is principally a means of communication. A speaker uses language principally to communicate the contents of his mental states, and especially the contents of his beliefs. One proponent of this view is D. H. Mellor (1990).

2 Mellor's method

Mellor prefers to practise philosophy than to preach about how to do it:

> In the sense in which astronomers are interested not in astronomy but in the stars, I am interested not in philosophy but in the various philosophical topics dealt with in this book [Mellor 1991a] – topics on which I find discussions of what philosophy is and how to do it shed very little light. I think the proof of our methods lies rather in the results of our applying them, and my case for my method, such as it is, rests on the contents of the ensuing chapters.
>
> (Mellor 1991a: xv)

Distinguish two questions. (Q1) What is Mellor's method? (Q2) Is it a good method? Mellor answers (Q2) by inviting us to assess the results of the application of his method. These results are his theories of various philosophical topics, and the degree of their success. Fair enough, but this does not answer (Q1). If Mellor's method has good results, it is a good method. But we still need to know what his method is. So (Q1) deserves an answer. Here is an outline of such an answer. In making explicit what is largely implicit, the following occasionally goes beyond what Mellor has said in print.[1]

Take a monadic term 'F'. (The following outline carries over to polyadic terms.) A concept *C* is suitably associated with 'F' so that *C* provides the meaning of 'F'. Given what 'F' means (the universal closure of), the open sentence 'Fx' entails certain propositions. These are the connotations of 'F'. They constrain what *F-ness*, the property which 'F' expresses, is. Some connotations are informative, others not. Some are obvious, others not. Philosophical

analysis says, *inter alia*, what the connotations of 'F' are.[2] A connotation of 'F' is true (false) at a world w if and only if the proposition expressed by that connotation is true (false) at w. Science and metaphysics say which connotations are true, which false. For example, Mellor claims that one connotation of an event's happening now is that its happening now is always a matter of fact (Mellor 1995: 59). Following Reichenbach, Mellor thinks that special relativity shows that connotation to be false (at the actual world). Mellor also thinks that 'event x is now' has the connotation that time flows (i.e. events move from the past to the present). Following McTaggart, Mellor (1995: 1–2) thinks that that connotation is false (at every world). If a connotation of 'F' is false at a world w, nothing at w falls under the concept C. For any world w, the objects at w that fall under C are exactly those objects at w which have *F-ness*. It follows that if a connotation of 'F' is false at w, nothing at w has *F-ness*.

Suppose we resist this conclusion because we want to say that something is F at w. One option is to replace the concept C associated with 'F'. There is nothing untoward in this procedure since our concern is with what *F-ness* is in the world, not with what the ordinary concept of it is.[3] Accordingly, we can replace C with a new concept, C^*, where C^* has all the connotations of 'F' that are known to be true, but none of the connotations of 'F' that are known to be false. If C^* is similar enough to C, then C^* can be suitably associated with 'F' to provide its (reformulated) meaning.

What of the claims made by those who continued to use C? Unphilosophical folk have never obviously replaced the concept *now* with *now**. Does Mellor think that the folk speak falsely when they use the word 'now' because of the false connotations the concept *now* has? Is he an error-theorist about such folk utterances? Mellor might reinterpret folk utterances in line with his own metaphysics. Thus, he might reinterpret the folk's use of 'now' as being associated with the concept *now**. Yet even if Mellor revises which concept *he* associates with that word, it is unclear how this bears on the conceptual practices of the folk. The folk are ignorant of these philosophical views and revisions. If they were in error before, they remain in error. So even if the replacement procedure enables clued-up philosophers not to speak falsehoods when they use such words as 'now', 'later', 'soon' and the like, it leaves the folk in the lurch.

Lastly, a complication: some connotations admit of degree. 'C causes E' has the connotations that C is evidence for E, that C explains E and that C is a means to E. Causation can be probabilistic, but 'the closer C comes to determining E … the stronger the evidence is that C provides for E, the better C explains E and the more useful C is as a means to E' (Mellor 1995: 93). Can this be reconciled with the fact that the connotations of causation are semantic entailments of 'C causes E'? Here is one way. Probability admits of degree. So too do evidence, explanation and the usefulness of a means to an end. The degree of probability is commensurate with these other degrees: C's being a cause with probability p of E entails that C provides evidence for E with a degree commensurate to p, that C explains E to a commensurate degree, and so on.

Having outlined Mellor's method, I turn to its application to the issue of communication.

3   Mellor's theory of communication

A theory of communication tells us what communication is, and thereby what we need to do to communicate with others. Mellor offers a theory of communication of which linguistic communication is a special case. He makes a truth-theoretic component central to his theory. Mellor (1990: 81) offers 'two important truisms' about truth. The first draws on Aristotle's *Metaphysics* (1928: 7, Book IV). It is that 'to believe or say truly is to believe or say, of what is, that it is, or of what is not, that it is not' (Mellor 1990: 82). The second draws on Ramsey (1990b).[4] It is that 'truth is that property of our beliefs which ensures that the actions they make us perform will succeed' (Mellor 1990: 82).

A connotation may be obvious or unobvious. Presumably, 'truism' is Mellor's term for an obvious connotation. I assume that Mellor takes the above truisms to express necessary truths. Mellor (1991c: 275) elsewhere says that 'I expect truth itself to be defined as the property of full beliefs that guarantees the success of actions based on them, probability providing the weaker assurance of success that expected utility spells out'. Philosophical orthodoxy takes (correct) philosophical definitions to state necessary truths. So I take Mellor to be stating a (purported) necessary truth about the property of *truth*. Mellor also takes his theory of communication to provide the truth conditions of beliefs:

> we can't equate a belief's truth conditions with those in which every action it helps to cause succeeds. But we can if we restrict the actions to those caused just by it and some desire. Then its truth conditions are what I shall call its 'utility conditions': those in which all such actions would achieve the desired end.
>
> (Mellor 1991d: 23)

Specifying that a belief's truth conditions are its utility conditions is, presumably, stating a necessary truth about those truth conditions.

Now the theory. Mellor takes communication to be a form of observation. Suppose you observe some fact, and so acquire a belief about it. What makes your observation a good one? Mellor's answer is that there has to be a causal link between the fact observed and your belief about it, such that your belief is true because the fact which makes that belief true has caused you to get that belief. Some observations are direct, others are indirect. To make an indirect observation of something, ϕ, a learnable correlation between ϕ and something else that we can observe directly is needed. This is a sign of ϕ. 'We make an indirect observation by first making a direct observation of a sign, and then making an inference from that to what we believe the sign signifies' (Mellor 1990: 86).

Suppose Sam Spade hears the doorbell ring (i.e. directly observes it). If his

observation is a good one, then his belief that the doorbell is ringing is true. Suppose he believes that the ringing is a sign of a client, and that a client *is* ringing the doorbell. Then Sam indirectly observes a client. For Sam's indirect observation to be a good one, his direct observation of the doorbell ringing must be good, and the inference from the premises (his beliefs that the doorbell has rung and that the doorbell's ringing is a sign of a client, and the fact that the doorbell's ringing is a sign of a client) to the conclusion (that there is a client) must be good. That is, the inference must transmit truth from its premises to its conclusion. If Sam's indirect observation of a client is a good one, then his belief that there is a client is true.

The case of communication is similar if more complicated. Suppose Sam hears his secretary saying 'there's a client'. Suppose his observation is a good one. Then his belief that the secretary said 'there's a client' is true. Suppose Sam believes that what she said is a sign that she believes that there is a client. Sam then infers that she believes that there is a client. Lastly, suppose Sam believes that her believing that there is a client is a sign that there is a client. Sam then infers that there is a client.

What distinguishes communication from the doorbell case is that X gets the belief that *p* from what Y believes (namely, from Y's belief that *p*). But X does not infer that *p* directly from what Y says. X infers *p* indirectly via what X believes Y believes. Suppose Rabbit wants to tell Pooh the truth about honey, and believes the truth to be that there is honey in the pot. Then:

> Rabbit doesn't just want to say what's true: he wants to make Pooh believe it. And as an experienced [informant], he knows that Pooh will only believe what he says if Pooh believes that he believes it too. So Rabbit's *immediate* desire is to give Pooh a true belief about what he, Rabbit, believes. So what Rabbit will tell Pooh is not necessarily what he actually believes, but what he *believes* he believes.
>
> (Mellor 1990: 92)

Communication, then, is the production in the audience of beliefs about what the speaker believes he believes (i.e. to produce in the audience beliefs about some of the speaker's second-order beliefs).

Lastly, we want *true* beliefs because:

> What is generally and inherently good about getting true beliefs is that they're useful, in the following sense: *truth is that property of our beliefs which ensures that the actions they make us perform will succeed.*
>
> (Mellor 1990: 82)

An action is the effect of a belief plus a desire. The action is successful if it fulfils the desire in question – if it achieves the object of the desire. That is 'what the truth of our beliefs ensures: that the actions they combine with our desires to cause will succeed in fulfilling those desires' (Mellor 1990: 83).

4   Difficulties with Mellor's theory

The following four difficulties concern full beliefs: those beliefs which, if true, supposedly guarantee the success of actions based on them.

Mellor has elsewhere argued that there can be no simultaneous causation (Mellor 1995: 220–4). Given those arguments, it follows that if at time t Toad desires honey, and believes that there is honey in the pot, and this causes him to eat what is in the pot, then his eating it cannot occur at t. It must occur at some distinct time t^*. And, given Mellor's arguments that causes must precede their effects, it follows that t^* must be later than t (Mellor 1995: 234–7). Therefore, Toad's belief and desire at t cause him to eat what is in the pot at a later time t^*, and cannot cause him to eat what is in the pot at any time at, or earlier than, t.

First difficulty: the time-lag between cause and effect

Given Mellor's arguments that there cannot be simultaneous causation, Toad's belief and desire at t cannot cause Toad to act at t, but only at a later time t^*. But between t and t^* relevant changes may occur in Toad's environment. At t Toad desires honey and has the true belief that there is honey in the pot. He acts at t^*. But suppose that between t and t^* the weasels replace the honey in the pot with gravel. When Toad eats the contents of the pot at t^* his action of eating is the effect of a desire for honey and of a true belief – namely, his belief at t that there is honey in the pot. But his action does not fulfil his desire despite being caused by a true belief.

Evidently, we need to specify the time that Toad's beliefs are about. As noted, Toad's belief that there is honey in the pot at t may not cause an action which fulfils a desire for honey. But consider his belief that there will be honey in the pot at the later time t^*. If that belief is true, there will be honey in the pot at t^*. The time-lag difficulty has no force against this belief. In general, there are two (exclusive but non-exhaustive) classes of beliefs. There are those beliefs that concern a time t, when t occurs before the time of any action which those beliefs can cause. There are also those beliefs that concern a time t, when t is not earlier than the time of any action which those beliefs can cause. The time-lag difficulty faces beliefs of the first sort, but not the second. But since Mellor's account is intended to apply to all true beliefs, and a fortiori to beliefs of the first sort, this first difficulty remains.

Second difficulty: the modality of causal connections

Mellor (1995: 31) believes that causal connections are contingent. If factum C causes factum E in some world, there is another world in which C exists, but in which C does not cause E. ('Facta' is Mellor's term for the relata of causation). Let B and D be a token belief and desire respectively. It follows

from the above that where a true belief B and desire D cause an action A which fulfils D, the causal connection between B and D, on the one hand, and A, on the other, is contingent. That is, there is a world in which B and D exist (or counterparts thereof), but do not cause A. In that world, B and D do not cause an action which fulfils D. Now, for every world w, consider all the token beliefs in w with the same content as B. Call these the B-beliefs in w. Consider too all the token desires in w with the same content as D. Call these the D-desires in w. Lastly, call all token actions of the same type as A A-actions. Given the contingency of causation, there is a world w^* in which B-beliefs and D-desires are jointly held, but no B-belief and D-desire cause an A-action, and so do not cause an action which fulfils a token D-desire. So Mellor seems committed to claiming that no B-belief is true in w^*. That is a surprising consequence. Alternatively, he could reply: 'Granted the causal powers of beliefs vary across worlds. But consider, for each world, the causal powers beliefs have at that world. In particular, consider whatever actions those beliefs, in conjunction with desires, cause at that world. My view is that the truth of those beliefs, in conjunction with those desires, causes actions which guarantee the fulfilment of those desires.' This reply confronts the third difficulty.

Third difficulty: probabilistic causation

Some causation is irreducibly probabilistic (Mellor 1995: 52–8). Consider a time t at which Toad has the desire to eat honey, and the true belief that there will be honey in the pot at the later time t^*. Suppose that the causal connection between belief and desire (cause) and action (effect) is irreducibly probabilistic. Then Toad's having that (true) belief and desire need not cause him to perform an action which fulfils that desire. Given that the causal connection is only probabilistic, Toad may perform no action. Or, for the same reason, his belief and desire may cause an action other than the one which would fulfil his desire. Given probabilistic causation, a belief and desire are not guaranteed to cause an action of a certain given type, still less an action which fulfils that desire. A belief and desire pair assign different probabilities to various possible actions. Some of these actions will fulfil the desire, others will not. Therefore, a desire and a true belief may cause an action although that action fails to fulfil that desire.

Moreover, even if in the actual world beliefs and desires deterministically cause actions, that is a contingent truth. Whether the causal connection between belief–desire pairs and actions is deterministic or indeterministic depends – as Mellor agrees – on what psychological laws hold between belief–desire pairs and the actions they cause (see Crane and Mellor 1991: 93–100). And laws of nature – as Mellor (1991e) agrees – are contingent. So, in some possible worlds, desires and true beliefs cause actions where those actions do not fulfil those desires.

Fourth difficulty: unknown features of the environment

Suppose that at *t* Toad desires honey and believes (truly) that there will be honey in the pot at *t**. But suppose that the pot contains both honey and – unknown to Toad – a booby-trap. The action supposed to fulfil the desire for honey triggers the bomb, and Toad dies, his desire for honey unfulfilled. In this case, a desire and a true belief cause an action, but the action has unexpected consequences. These consequences prevent the desire being fulfilled. Again, a true belief and a desire need not cause an action which fulfils that desire.

The upshot of each of these difficulties is that a belief's truth conditions are not its utility conditions.

5   The appeal to 'no impediments'

Can the fourth difficulty be met by adding to Toad's set of (relevant) beliefs a belief that there are 'no impediments' to his action's fulfilling his desire?[5] The suggestion is that if, for instance, the pot is booby-trapped, this counts as an impediment to Toad's action fulfilling his desire. Moreover, Toad's belief that there are no impediments to his action's fulfilling his desire is false. In general, if there are impediments to Toad's actions fulfilling his desires, Toad's belief that there are no such impediments will be false. Therefore, not all Toad's (relevant) beliefs will be true.

There is a problem, however, about how the 'no impediments' clause is to be specified. I cannot find a satisfactory specification. Here are three candidates:

(1)   There is no true proposition which entails the proposition that Toad's desire remains unfulfilled.
(2)   There is nothing which makes it physically impossible for Toad's desire to be fulfilled.
(3)   There is nothing which raises the chance that Toad's desire will be unfulfilled.

Readings (1) and (2) are too weak. Suppose that the pot is on a high shelf. This does not entail that Toad cannot get at the pot, nor does it make it physically impossible for him to get at it. But it may, in the circumstances, prevent him getting at it. Reading (3) is too strong for Mellor's purposes. At every (metaphysically) possible world, there is something that is apt to cause Toad's desire to eat honey to remain unfulfilled. This may, for instance, be that he is mortal and his mortality may cause his death before he eats the honey, or it may be the chance of an earthquake which would shatter the pot. Every world will contain some impediment – some factor whose presence (in contrast to whose absence) lowers the chance of Toad's desire being fulfilled. It follows that, in every world, Toad's belief that there are 'no impediments' is false. Consider, then, the following conditional:

> If all of Toad's beliefs (including his beliefs that there is honey in the pot and that there are no impediments) are true, and he desires to eat honey, then the action his beliefs and desire cause will ensure that his desire is fulfilled.

As noted, at every world, there are impediments to the fulfilment of Toad's desire. It follows that, in every world, Toad's belief that there are no impediments is false. It further follows that the antecedent of the above conditional will be false at every world. Therefore, at every world, the above conditional is trivially true. That is, it is necessarily the case that the conditional is trivially true. Therefore, Mellor's account will be only trivially true.

Clearly, some other reading of the 'no impediments' clause is needed. Whyte (1997) suggests that Toad has a further true belief, namely the true conditional belief that if Toad believes that there is honey in the pot, and he desires to eat honey, then the action these propositional attitudes cause will (if performed) fulfil his desire. According to Whyte (1997: 84–6), adding this true conditional belief to Toad's belief and desire set entails that the action in question (if performed) will fulfil his desire. In general, Toad's having a *true* belief that p is a matter of the action caused by his belief that p and his desire that q ensuring the fulfilment of his desire that q if Toad has the following further true belief: the belief that the action caused by his belief that p and his desire that q ensures the fulfilment of his desire that q.

This suggestion also faces a problem of trivialization. Consider an analogy. Someone claims that it is a conceptual truth that grey skies ensure rain. You deny this by citing cases in which there are grey skies but no rain. The claim is then amended: it is a conceptual truth that grey skies ensure rain under circumstances C. And what are circumstances C? Just those circumstances in which it is true that grey skies ensure rain. But whereas the original claim was interesting but false, the amended claim – that grey skies ensure rain in just those circumstances in which grey skies ensure rain – is only trivially true.

Return to the debate about the connection between true belief and desire-fulfilment. Mellor advanced the thesis that there is a conceptual connection between (1) Toad's having a true belief that there's honey in the pot and (2) a guarantee that Toad's desire for honey is fulfilled. But the thesis faced counterexamples. Whyte added a 'no impediments' clause to exclude these counterexamples. But this clause specifies something about the circumstances in which Toad has a true belief that there is honey in the pot, namely that the circumstances are just those in which the action caused by Toad's having the true belief that there is honey in the pot and his desire for honey ensures the fulfilment of that desire. Of course, Whyte can show that *in circumstances so specified* there is a conceptual connection between Toad's true belief about honey and his desire for honey being fulfilled. For the suggestion is now the following triviality: in certain circumstances (namely those in which the action caused by Toad's true belief that there is honey in the pot and his desire for honey ensures the fulfilment of that desire), the action caused by Toad's true

belief that there is honey in the pot and his desire for honey ensures the fulfilment of that desire.

The difficulties facing Mellor's theory show that it is not the case that an action caused by a desire and a true belief will ensure that that desire is fulfilled.[6] What can be salvaged? Perhaps true belief is belief that is a reliable means of producing the fulfilment of desires. More fully: that a belief with a property F is a reliable means of fulfilling a desire by causing an action if the belief's having F confers a higher chance of the desire's being fulfilled than if the belief lacked F. The idea here is that, in the actual world (and relevantly similar worlds), a true belief is a reliable means of fulfilling a desire by causing an action if, in the actual world, the belief's being true gives the desire a higher chance of being fulfilled than if the belief were not true.[7]

6   Belief and communication

Here I assess two of Mellor's claims about communication.

(1)   Communication is the production by the speaker in his audience of beliefs about certain second-order beliefs of the speaker.
(2)   A speaker cannot tell his audience that p without being conscious that he believes that p.

These claims are logically independent. If communication requires the speaker to get his audience to have beliefs about the speaker's second-order beliefs, it does not follow that communication requires the speaker to *have* any second-order beliefs. Nor does the converse hold. Presumably Mellor makes claim (2) because he makes claim (1) and holds a certain theory about conscious belief. According to that theory, for someone to believe that he believes that p is for him to have a conscious belief that p.[8] Mellor defends his theory against the charge that it makes communication more complicated than it appears to be. His defence is that mental states need not be conscious and that 'our mental life is more complicated than we ourselves are ever aware of at the time' (Mellor 1990: 92). But this defence conflicts with claim (2) – the claim that a speaker cannot tell his audience that p without being conscious that he believes that p. Whereas the defence buries the complexity of Mellor's theory in the unconscious, claim (2) exhumes it.

Claims (1) and (2) seem not to state necessary conditions for a speaker to tell his audience that p. One kind of counterexample concerns speech acts performed automatically. A non-linguistic action may occur without any second-order or conscious belief on the agent's part, as when a cricketer catches a ball that suddenly comes his way. Likewise, it seems that a speech act may occur without any second-order or conscious belief, as when a bank-teller blurts out 'don't shoot!' to a robber. Another kind of counterexample concerns speech acts performed without the attention that second-order or conscious beliefs involve. For instance, when a husband absent-mindedly replies to his

nagging wife, he tells her something, but not necessarily because of any of his second-order or conscious beliefs. Accordingly, just as Mellor (1990: 91) grants that people's beliefs may cause them to perform non-linguistic acts without their being aware that they have those beliefs, so too it seems that people's beliefs may cause them to perform linguistic acts without their being aware that they have those beliefs.

The upshot is that (1) and (2) are false. We should accept a simpler account of communication that does not make either claim.

7   Is there a property of truth?

Mellor (1991c: 275) thinks that 'truth itself [is] to be defined as the property of full beliefs that guarantees the success of actions based on them.' Likewise, Whyte (1990: 149) thinks that 'truth just *is* the property of a belief that suffices for your getting what you want when you act on it'. Now Mellor distinguishes between concepts and properties; the concept of *F* is not to be identified with the property *F-ness*.[9] Accordingly, we can distinguish between the concept of truth and the (putative) property *truth*. One option is to take Mellor's theory as a theory of the concept of truth, but not also of the property *truth*. The description 'the property of a belief that suffices for your getting what you want when you act on it' may fix the reference of the term 'truth', but leave it open whether there is some other description which picks out the property *truth* and tells us what that property is.[10] Thus, even if Mellor and Whyte provide a correct and comprehensive characterization of the concept of truth, it remains open whether they provide a correct and comprehensive characterization of the property of *truth*, and in particular whether they tell us 'the most interesting fact' about it (Whyte 1990: 149).

Mellor has not (to date) stated whether or not *truth* is a *bona fide* property alongside *being-a-mass* and *being-a-temperature*. Perhaps his talk about the property *truth* is intended as a *façon de parler*. According to Mellor, there may be true propositions involving the concept of *F*, although there is no property *F-ness* that makes them true.[11] Consequently, that there are truths does not entail that there is a property *truth*. Nevertheless, I will argue that Mellor's views about what a property is, and which properties exist, commit him to there being such a property.

Mellor (1995: 190–2) holds what he calls the 'Ramsey's test' for the existence of properties. This states that a property *F* exists if and only if there exists some law of nature of which *F* is a constituent. Mellor admits the existence of psychophysical laws, such as the law of nature stated by sentence (S1) below (1995:174):

> (S1)   For all particulars *x*, if *x* has a certain credence assigning a degree of objective probability *p* that it is raining and *x* goes out, then the chance that *x* will take a coat is *p*.

Now consider *credence* (i.e. *degree of belief*) as an example of a putative property. We proceed in three steps.

Step (1)   *Credence* is defined as the (putative) property such that, among other things, (S1) states a law of nature.

Step (2)   To say that (S1) states a law of nature is to make an existential claim. It is to claim that there exists a property of *credence* such that (S1) states a law of nature of which that property is a constituent.

Step (3)   Given that (S1) states a law of nature, it follows that there exists a property of *credence* (Mellor 1995: 174).

Now presumably, Mellor would take sentence (S2) to state a psychophysical law:

(S2)   For all objects x, if x has a true belief that there is honey in the pot, and x desires to eat honey, then x's belief and desire will cause an action which has probability p (where p is a degree of objective probability) of fulfilling x's desire.

We proceed in three steps as before:

Step (1)   *Truth* is defined as the (putative) property of a belief such that, among other things, (S2) states a law of nature.

Step (2)   To say that (S2) states a law of nature is to make an existential claim. It is to claim that there exists a property of *truth* such that (S2) states a law of nature of which that property is a constituent.

Step (3)   Given that (S2) states a law of nature, it follows that there exists a property of *truth*.

Truth passes Ramsey's test for the existence of properties. Given his views about what it takes for a property to exist, Mellor is committed to there being a property *truth* (see Putnam 1978: 101–2).

Some philosophers would explore whether (S2) can be paraphrased by an equivalent sentence, (S2*), which does not involve explicit reference to a property of *truth*.[12] (S2*) might be:

(S2*)   For all objects x, if x believes that there is honey in the pot, and x desires to eat honey, and there is honey in the pot, then x's belief and desire will cause an action which has probability p of fulfilling x's desire.

If (S2) can be so paraphrased, these philosophers would claim, the above case for admitting a property of *truth* collapses. This reasoning, however,

requires that passing Ramsey's test is not a sufficient condition for a putative property being a real property. Mellor can either reject that line of thought or accept it and amend Ramsey's test, perhaps as follows: a putative property F exists if and only if, for some law of nature, reference to F is ineliminable in fully stating that law. I assume that Mellor would take the first option. For he denies that if entities of a putative kind K do not need to be referred to, it follows that Ks do not exist (Mellor 1973: 113).[13] Accordingly, the fact that (S2) apparently involves reference to a property of *truth* but a paraphrase of (S2), (S2*), apparently lacks this reference does not entail that (S2) does not involve genuine reference to a property of *truth*. [Perhaps both (S2) and (S2*) involve genuine reference to the property of *truth* because (S2) involves such reference, and (S2*) is equivalent to (S2)].[14] Therefore, Mellor apparently remains committed to there being a property of *truth*.

Summary

Mellor says that, where a desire plus a true belief cause an action, that action ensures the desire's fulfilment. I say that the action makes the fulfilment at most probable. Mellor says that for a speaker to communicate to his audience is for him to get his audience to believe that he has a second-order belief about what a certain sign signifies. I say that, to be communicated to, the audience needs believe only that the speaker has a first-order belief about what a sign signifies. Mellor does not say whether or not there is a property of *truth*. I say that, by his lights, there is.[15]

Notes

1   The most explicit statement of his methodology is in (Mellor 1995: 58–61). Its source is Ramsey (1990a).
2   For the role of philosophical analysis, see Mellor (1991b), especially Section 1.
3   Mellor (1995: 5) writes: 'what concerns us is what causation really is, not what those who never think about the matter think it is, i.e. not the everyday concept of causation but what causation is in the world.' See Mackie (1973: 11–13).
4   Ramsey's work has subsequently also been developed in Appiah (1986) and Whyte (1990; 1991).
5   This suggestion and phrase is introduced by Whyte (1990: 151).
6   Chapter 5 of Stich (1990) presents a more extreme view. According to him, 'the instrumental value of true beliefs is far from obvious, and thus those who think that true beliefs are instrumentally valuable owe us an argument that is not going to be easy to provide' (Stich 1990: 124). This is not my view. For replies to Stich, see Lycan (1991), Alston (1996: 258–61) and Goldman (1999: 72–4).
7   For more along these lines, see Stalnaker (1986: 117–19).
8   See Mellor (1991f). Mellor uses the term 'assent' to refer to conscious belief (1991f: 33) and claims that '*assenting to a proposition is believing one believes it*' (1991f: 36).
9   Here a concept is to be understood as the kind of mental representation which is typically expressed by a general term. Mellor claims that '… properties like masses and temperatures could and would exist even if we have no corresponding

concepts or predicates', and that 'the existence of properties in any sense relevant to causation cannot depend on the existence of concepts or predicates' (Mellor 1995: 185, 186).

10   The distinction between the concept of truth and the property of *truth* is drawn by Alston (1996: 37–8). He also claims that different concepts may pick out the property of *truth*, some informatively, others uninformatively. See also O'Leary-Hawthorne and Oppy (1997: 172, 175–6).

11   For Mellor's views on these matters, see Mellor (1991g,h; 1995: 172–5, 185–99).

12   See, for example, Leeds (1995).

13   See also the further references to Mellor's work given there.

14   See Alston (1958; 1996: 238–9). Alston's 1958 view is endorsed by Mellor and Oliver (1997: 15).

15   I am grateful to André Gallois and Rosanna Keefe for comments.

References

Alston, W. P. (1958) 'Ontological commitments', *Philosophical Studies* 9: 8–17

—— (1996) *A Realist Conception of Truth*, Ithaca, NY: Cornell University Press.

Appiah, A. (1986) 'Truth conditions: a causal account', in J. Butterfield (ed.) *Language, Mind and Logic*, Cambridge, UK: Cambridge University Press.

Aristotle (1928) Metaphysics, in W. D. Ross (ed.) *The Complete Works of Aristotle Translated into English*, Vol. 8, Oxford: Clarendon Press.

Crane, T. and Mellor, D. H. (1991) There is no question of physicalism, in D. H. Mellor (ed.) *Matters of Metaphysics*, Cambridge, UK: Cambridge University Press.

Goldman, A. I. (1999) *Knowledge in a Social World*, Oxford: Oxford University Press.

Leeds, S. (1995) 'Truth, correspondence, and success', *Philosophical Studies* 79: 1–36.

Lycan, W. G. (1991) 'Why we should care whether our beliefs are true', *Philosophy and Phenomenological Research* 51: 201–5.

Mackie, J. L. (1973) *Truth, Probability, and Paradox*, Oxford: Oxford University Press.

Mellor, D. H. (1973) 'Materialism and phenomenal qualities II', *Proceedings of the Aristotelian Society* 47 (Suppl.): 107–19.

—— (1990) 'Telling the truth', in *Ways of Communicating*, Cambridge, UK: Cambridge University Press.

—— (1991a) *Matters of Metaphysics*, Cambridge, UK: Cambridge University Press.

—— (1991b) 'Analytic philosophy and the self', in *Matters of Metaphysics*, Cambridge, UK: Cambridge University Press.

—— (1991c) 'Objective decision making', in *Matters of Metaphysics*, Cambridge, UK: Cambridge University Press.

—— (1991d) 'I and now', in *Matters of Metaphysics*, Cambridge, UK: Cambridge University Press.

—— (1991e) 'Necessities and universals in laws of nature', in *Matters of Metaphysics*, Cambridge, UK: Cambridge University Press.

—— (1991f) 'Consciousness and degrees of belief', in *Matters of Metaphysics*, Cambridge, UK: Cambridge University Press.

—— (1991g) 'Laws, chances, and properties', in *Matters of Metaphysics*, Cambridge, UK: Cambridge University Press.

—— (1991h) 'Properties and predicates', in *Matters of Metaphysics*, Cambridge, UK: Cambridge University Press.

—— (1995) *The Facts of Causation*, London: Routledge

Mellor, D. H. and Oliver, A. (1997) (eds) *Properties*, Oxford: Oxford University Press.

O'Leary-Hawthorne, J. and Oppy, G. (1997) 'Minimalism and truth', *Noûs* 31: 170–96.

Putnam, H. (1978) *Meaning and the Moral Sciences*, London: Routledge and Kegan Paul.

Ramsey, F. P. (1990a) 'Theories', in D. H. Mellor (ed.) *F. P. Ramsey: Philosophical Papers*, Cambridge, UK: Cambridge University Press.

—— (1990b) 'Facts and propositions', in D. H. Mellor (ed.) *F. P. Ramsey: Philosophical Papers*, Cambridge, UK: Cambridge University Press.

Stalnaker, R. (1986) 'Replies to Schiffer and Field', *Pacific Philosophical Quarterly* 67: 113–23.

Stich, S. (1990) *The Fragmentation of Reason*, Cambridge, MA: MIT Press.

Whyte, J. T. (1990) 'Success semantics', *Analysis* 50: 149–57

—— (1991) 'The normal rewards of success', *Analysis* 51: 65–74.

—— (1997) 'Success again: reply to Brandom and Godfrey-Smith', *Analysis* 57: 84–8.

5   Subjective facts[1]

Tim Crane

> It is obvious that a man who can see knows things which a blind man cannot
> know; but a blind man can know the whole of physics. Thus the knowledge
> which other men have and he has not is not a part of physics.
>
> (Bertrand Russell 1927: 389)

1   Mellor's objectivism and subjective facts

An important theme running through D. H. Mellor's work is his realism or, as
I shall call it, his *objectivism*: the idea that reality as such is how it is, regard-
less of the way we represent it, and that philosophical error often arises from
confusing aspects of our subjective representation of the world with aspects of
the world itself. Thus central to Mellor's work on time has been the claim that
the temporal A-series (previously called 'tense') is unreal while the B-series
(the series of 'dates') is real. The A-series is something that is a product of our
representation of the world, but not a feature of reality itself. And in other,
less central areas of his work, this kind of theme has been repeated: 'Objec-
tive decision making' (Mellor 1991a) argues that the right way to understand
decision theory is as a theory of what is the objectively correct decision, the
one that will actually as a matter of fact achieve your intended goal, rather
than the one that is justified purely in terms of what you believe, regardless of
whether the belief is true or false. 'I and now' (Mellor 1991b) argues against a
substantial subjective conception of the self, using analogies between subjective
and objective ways of thinking about time and subjective and objective ways
of thinking about the self. And in the paper which shall be the focus of my
attention here, 'Nothing like experience', Mellor (1991c) contests arguments
that try and derive anti-physicalist conclusions from reflections on the subjec-
tive character of experience. A common injunction is detectable: when doing
metaphysics, keep the subjective where it belongs, that is inside the subject's
representation of the world.

   Mellor's objectivism agrees with the Australian metaphysics, which he
admires. Australian metaphysics is, however, characteristically physicalist
in letter and in spirit. But Mellor has rejected physicalism in a number of
places, in most detail in a paper we wrote together, 'There is no question of

physicalism' (Crane and Mellor 1991). One view which is implicit in this paper is that each area of investigation should be answerable to its own standards and should not be required to justify itself in terms of how it relates to physics. The facts discovered by the various sciences can all be as objective as the facts discovered by physics. Objectivism, therefore, is not physicalism, since the former does not entail that all objective reality is physical, whereas the latter does.

Yet I shall argue here that in the case of the subjective quality of experience, Mellor has adopted ideas from physicalism which are implausible, and arguments which are mistaken, and that he would be better off without them. In his various discussions of the problem of the subjective character of experience, Mellor has expressed his view by denying that there are any 'subjective facts'. In 'I and now' he writes:

> Many philosophers overrate the present subject. *Pace* Nagel, there are no subjective facts or selves; nor … does our ability to think and talk about our present selves, and the world as seen from our present point of view, pose any special metaphysical, semantic or epistemic problems.
>
> (Mellor 1991b: 17)

And in 'Nothing like experience', he says that 'there are, I believe, no subjective facts about anything: they have all been falsely inferred from certain kinds of knowledge' (Mellor 1993: 1). The inferences he is talking about are made most lucidly in Frank Jackson's famous 'knowledge argument', which is designed to show, from apparently uncontroversial premises and simple reasoning, that the physicalist conception of the world is false. Mellor rightly points out that, if sound, the argument would show more than that: it would show that some facts are subjective, and thus that a view which says that all facts are objective would be false. (Since a lot of what follows depends on what 'objective' and 'subjective' mean, the reader will have to wait for a more precise statement.) So Mellor thinks that he has to show that the knowledge argument is unsound, since he thinks that he cannot accept its conclusion. He therefore adopts the ability hypothesis of Lewis (1990) and Nemirow (1990), which is intended to show that the knowledge argument is fallacious, resting on an equivocation on 'knowledge'.[2]

I shall argue here, against Lewis, Nemirow and Mellor, that the ability hypothesis is mistaken, and that all the other physicalist attempts to reject the argument (either as invalid or as unsound) are equally mistaken. The knowledge argument is a sound argument for the conclusion that there are subjective facts: facts about the subjective character of experience. However, unlike some defenders of the argument,[3] I do not think that this conclusion threatens any plausible version of physicalism, nor should it threaten the most plausible understanding of Mellor's views. Mellor and the physicalist should both accept that there are subjective facts, and they should both deny, therefore, that all facts are objective, in the sense that I shall explain.

2   The knowledge argument

Jackson's famous knowledge argument does not move from a claim about the existence of experience to the denial of physicalism; it moves from a claim about how we *know* about experience to the denial of physicalism, hence its name.[4] The argument starts with a thought-experiment about Mary, who has spent all her life in a black-and-white room and has never seen any colours other than black and white. Now imagine that Mary has made an intensive study of the science of colour in all its aspects – physics, physiology, psychology and so on. In fact, let us suppose that she knows all the physical facts about colour. Now suppose that one day Mary leaves her black-and-white room, and the first thing that she sees is a red tomato. It is natural to say that she now knows something which she did not know in the black-and-white room: what it is like to see red. Yet this thing she now knows is not a physical fact, since by hypothesis she knew all the physical facts in the black-and-white room. So if a new piece of knowledge is a new fact, then Mary learns a new fact when she leaves the black-and-white room. If physicalism is (as seems plausible enough) the view that *all facts are physical facts*, then it appears that physicalism is false.

The knowledge argument does not beg the question against physicalism. This is clear if we represent its premises and conclusion as follows:

(1)   In the room, Mary knows all the physical facts about colour.
(2)   Having left the room, Mary learns something new about colour.
(3)   Therefore: not all facts are physical facts.

That, in essence, is the argument – although some extra assumptions are needed to demonstrate its validity properly. But it is clear that neither premise (1) nor premise (2) obviously begs any questions against physicalism. A physicalist could hardly object that the idea of someone learning all the physical facts begs the question against physicalism. And (2) seems an irresistible and simple thing to say about the story as described above. Maybe, when these premises are scrutinized, they will come to show some deep incoherence – but the argument as stated does not obviously beg the question.

Physicalists have tried to resist the conclusion by impugning either the validity of the argument or the truth of the premises. I think they are wrong. I think that the argument is valid, and that physicalists should accept its premises. So they should accept its conclusion. Yet I shall argue too that they should not worry about this conclusion; so this conclusion cannot be that physicalism, properly understood, is false.

In Section 3, I will assess the objection that the argument is invalid, and in Section 4 I will assess the objections to the premises. In Section 5 I will bring out what I think the argument really shows: that there are subjective facts. In Section 6 I shall examine the consequences of this conclusion for physicalism and for Mellor's views.

3 Challenging the argument's validity: the 'ability hypothesis'

Those who challenge the argument's validity normally claim that it involves an equivocation on 'know'.[5] In the first premise, 'know' is used to express propositional knowledge, but (they say) in the second premise it is used to express knowledge-how or ability knowledge. We should agree that Mary learns something new, but what she learns when she first sees red is how to recognize red, to imagine red and remember experiences of red things (see Lewis 1990; Nemirow 1990; Mellor 1991c). Having seen something red, she can now recognize the colour of fire engines, she can consider whether she wants to paint her bedroom red and she can remember this decisive encounter with a tomato. These are cognitive abilities, not pieces of propositional knowledge, and it is a widely held view that there is no reduction of ability knowledge to propositional knowledge. So Mary can learn something new – in the sense of gaining an ability – but it is not a new piece of propositional knowledge. Knowing what it is like to see red is know-how. So the knowledge argument is invalid because it involves a fallacy of equivocation: 'know' means something different in the two premises. Since it is only in the case of propositional knowledge that the objects of knowledge are facts – if I know how to ride a bicycle, *how to ride a bicycle* is not a fact – it is concluded that Mary does not come to know any new facts and physicalism is saved.

This response, known as 'the ability hypothesis', presupposes two things:

(1) that knowledge-how is ability knowledge, and it is completely different from, and irreducible to, propositional knowledge; and
(2) that regardless of the abilities she acquires, Mary does not come to know any new propositions whatsoever.

The first claim is a general theoretical claim about the relation between know-how, abilities and propositional knowledge. This claim is actually more dubious than is normally assumed; but space does not permit me to examine it here.[6] I shall concentrate rather on the second claim.

The defenders of the ability hypothesis say that Mary learns no new propositional knowledge at all. But this claim is really very implausible. For there is a very natural way for Mary to express her knowledge of what it is like to see red: 'Aha! Red looks like this!'. (Let us suppose, for simplicity, that Mary knows that tomatoes are red, and she knows that she is seeing a tomato; these are innocuous assumptions.) Now 'Red looks like this' is an indicative sentence; in a given context, it surely expresses a proposition; and in the context described, the proposition is true. (It could have been false. Suppose Mary were shown a joke tomato, painted blue. The proposition expressed by 'Red looks like this' would be false; red does not look like that.) And it is a proposition that Mary did not know before. This all assumes that a sentence containing a demonstrative can be used to express a proposition; but this assumption is

innocuous and should be accepted by all participants in the debate (we shall see its full relevance later). So *even if* Mary did acquire lots of know-how, and *even if* know-how is essentially different from propositional knowledge, then there is still something that she learns that she could not have known before. And that is enough for the argument to succeed.

Further support for the view that there is a proposition which is learned is provided by Brian Loar's (1997) observation that someone can reason using the sentence 'Red looks like this': they could embed it in a conditional, for example 'If red looks like this, then either it looks like this to dogs or it does not'. On the face of it, this is a conditional of the form 'If P then Q'; the substituends for P and Q are bearers of truth values and therefore possible objects of propositional knowledge (Loar 1997: 607).[7] The ability hypothesis has to explain this away if it is to support its conclusion that nothing propositional is learned. I doubt whether this can be done. For all these reasons, I reject the ability hypothesis.

An alternative way to question the validity of the argument is to say that the knowledge gained is knowledge by acquaintance.[8] Mary is acquainted with some feature of redness (what it looks like) or with some feature of her experience (qualia, as it may be). Acquaintance knowledge is not reducible to propositional knowledge; but these features (of redness or of experiences) may nonetheless be physical. To this objection, my response is essentially the same as my response to the ability hypothesis: unless the objector can show that Mary does *not* learn any propositional knowledge too, then the fact that she does gain acquaintance knowledge is irrelevant to the argument's conclusion. And we have a perfectly clear example of the kind of proposition Mary learns: the proposition expressed by the sentence 'red looks like this'.

Mellor thinks that the ability hypothesis refutes the knowledge argument; he also says it explains why Nagel is wrong about the limits of objective knowledge:

> These are not the only otherwise mysterious facts which the know-how theory explains. It also explains science's mysterious inability, which so impresses Nagel, to tell us what a bat's sonar experiences are like. But on the know-how theory this is no mystery, nor a limitation on the factual scope of objective science. For the only knowledge any science ever gives us is knowledge of facts. And even if many abilities depend on knowing facts, there is always more to having those abilities than knowing those facts.
> (Mellor 1991c: 7)

But if the ability hypothesis is false, then it cannot explain why Nagel is wrong about the 'factual scope of objective science'. Indeed, it seems rather that there are facts about the bat's experience (assuming it has experiences) which are beyond the scope of objective science: the facts which would be truly expressed (*per impossibile*) by saying 'Experiencing the world from a sonar point of view is like this'. Or, to take a more everyday example, the fact

that I can express when I say 'red looks like this' is a fact that a blind person cannot know. Yet, as Russell (1927) points out, a blind person can know the whole of physics. And there is nothing relevant to this debate which stops the blind person learning the whole of objective science. True enough, the sighted person has abilities that the blind person does not have, and Mellor is right that no amount of science can give you these abilities. But this is irrelevant. The important point is not that there are these abilities which someone who knows what it is like has; the important point is that someone who knows what it is like knows *that certain things are the case*. This is the propositional knowledge which the sighted have and the blind lack, in addition to whatever abilities they may also have.

4   Challenging the premises

I therefore reject these attempts to dispute the validity of the argument; the argument is valid. But what about the premises? Few physicalists wish to challenge the first premise, that in the story as told Mary knows all the physical facts about colour vision.[9] For suppose a physicalist did deny this. Then her or she would have to accept that there are some *physical facts* which in principle cannot be known without having certain experiences. Physics, the science which states the *physical facts*, is in principle *incompletable* until certain very specific experiences are had. Now it may be true that having knowledge in general requires having experiences of some kind. Yet how can physicalism, which bases its epistemological outlook on physical science, require that science demands us to have certain *specific* experiences? The suggestion has little plausibility.

So most responses to the argument have challenged the second premise instead, and claimed that Mary does not learn any new fact. In a recent survey, Güven Güzeldere describes this character of this dominant response as follows:

> The pivotal issue here is whether the having of an experience constitutes a special class of irreducible 'first-person facts' or whether what is lacking in Mary has to do with her experiential 'mode of access' to facts that she is already acquainted with.
>
> (Güzeldere 1997: 38)

The idea seems to be that Mary already knows all the facts in question; she simply gains a new 'mode of access' (whatever that is) to a fact she already knew. If this response were right, then certainly the argument would be undermined. But it seems to me that, despite its popularity, the response cannot be correct.

The central idea is that Mary apprehends or encounters in a new way something she already knew. The phrase 'mode of access' is often used to describe what this encountering in a new way is. But what are 'modes of access'? One

way to understand them is in terms of new *Fregean modes of presentation* of the objects and properties already known under other modes of presentation. On this interpretation, the puzzle about the argument is of a piece with other puzzles about intensionality, and many authors have explicitly drawn this comparison. Vladimir might know that Hesperus shines in the evening but not know that Phosphorus shines in the evening. We do not conclude from this that Hesperus is not Phosphorus since, as is well known, 'X knows that *p*' is not an extensional context. According to this view, the fact that Hesperus shines in the evening is the same fact as the fact that Phosphorus shines in the evening – after all, they are the same star, the same shining, the same evening! So although Mary knows that red looks like this, this is not a new fact that she has learned but, analogously, a new mode of presentation of a fact she knew before.

But which fact is this? We need to identify something that can be referred to in more than one way, the relevant fact concerning which can be learned about in the black and white room. One way of putting it might be like this. When she leaves the black and white room, Mary judges that *seeing red is like this*. The physicalist says that seeing red is being in brain state B, so let us suppose Mary knew this in the black and white room. Mary can therefore infer that being in brain state B is like this. We therefore have two terms, 'seeing red' and 'being in brain state B', that pick out the same thing, and a predicate 'like this' which can only be used when one is having the experience. But nonetheless, the experience is the brain state for all that.

So far so good. But remember that the distinction between different modes of presentation of the same thing is supposed to show that the second premise of the argument is false: *Mary does not learn anything new*. But it cannot show this. For if this construal of Mary's case and the case of Hesperus and Phosphorus are really parallel, then this entails that someone who comes to believe that Phosphorus shines in the evening because of their belief that Hesperus is Phosphorus does not learn anything new, but only comes to appreciate a previously known fact under a new mode of presentation. And this cannot be right: the original point of the distinction between sense and reference was to do justice to the fact that the discovery that Hesperus is Phosphorus can be a significant advance in someone's knowledge. It was a *discovery* about the heavens that Hesperus is Phosphorus, it was a new piece of knowledge that the ancients gained. So, similarly, the knowledge that Phosphorus shines in the evening is a new piece of knowledge. If facts are what you learn when you gain knowledge, then the normal approach to the distinction between sense and reference entails that what the ancient astronomers learned when they learned that Hesperus is Phosphorus is a new fact.

Of course, there is *something* that is the same before and after this particular discovery: how things are in the world, the reference of the terms, the entities. No-one disputes this about the Hesperus–Phosphorus case. So one could say: 'in a sense the facts are the same, in a sense they are different'. But the relevant question is whether anything is *learned* when someone acquires the belief that

Hesperus is Phosphorus, whether there is any new knowledge at all. And if there is a sense in which the fact learned *is* a new fact (even if there is a sense in which things are the same too) then there is new knowledge. This surely cannot be denied. Note that if you do deny this, you have to deny *at the very least* that there is new knowledge in the following sense: the knowledge that the two modes of presentation are modes of presentation of the same thing.[10] But this makes it impossible to even state what it is that the ancients learned.

Since they introduced the parallel, it would be fruitless for physicalists to try and draw some principled difference between the case of Mary and the case of Hesperus and Phosphorus. So either physicalism says that nothing new is learned in either case – which is a hopeless thing to say – or it says that something is learned in both cases. This is the only plausible thing to say. But then Mary does learn something new, the argument's premises are true, and we already decided it was valid. So is physicalism refuted?

5   Physical facts and subjective facts

This depends, of course, on what physicalism is. What is refuted is the doctrine that all facts are physical facts – given a certain understanding of 'physical' and 'fact'. The argument assumes a certain understanding of what 'physical facts' are.

What are facts? Philosophers have disagreed over the nature of facts, and over whether there are such things. Some say that facts are true propositions, others that they correspond one-to-one with true propositions, and others say that since they are what make true propositions true (they are truthmakers) they need not correspond one-to-one with true propositions.[11] What conception of fact does the knowledge argument assume? It is obvious, I think, that the knowledge argument has to assume that facts are *objects of propositional knowledge* – where a state of propositional knowledge is one described in claims of the form 'X knows that *p*' where X is a knower and '*p*' is replaced by a sentence. So for something to be a new fact is at least for it to be a new piece of knowledge, an advance in someone's knowledge, some piece of knowledge that he or she did not have before.

Does this mean that the knowledge argument covertly begs the question against physicalism by assuming a conception of *fact* which physicalism would reject? No. Whether or not physicalism decides to call objects of propositional knowledge 'facts', physicalism should certainly accept that there are objects of propositional knowledge, and that knowledge states are individuated partly by their objects. Everyone accepts that there are such objects of propositional knowledge, *whether or not* they also accept that there are facts in some other sense. So I think it is a mistake to say that we need to establish which theory of facts is correct before settling whether the knowledge argument works. This would be to claim that the argument had to have as a hidden premise that one particular theory of facts is the right one. But this is not so; everyone has to accept that there are objects of propositional knowledge.

The knowledge argument's conception of fact does not beg any questions. What it says is that a distinct piece of propositional knowledge is knowledge of a distinct fact. This is surely a very natural and uncontroversial idea. We can learn skills or pieces of information; when we learn pieces of information, we learn facts. But it is sometimes said that there are two notions of pieces of information (or fact): a coarse-grained notion and a fine-grained notion (see, for example, Van Gulick 1997: 562–3). According to the fine-grained notion, facts are individuated at the level of sense; for the coarse-grained notion, facts are individuated at the level of reference. Note that this point is sometimes put in service of the mistaken idea (dismissed above) that Mary learns nothing new, but only gains a new 'mode of access' to what she knew already. If one uses the distinction between coarse- and fine-grained facts to support this mistaken idea, then one is forced to say that *only* the coarse-grained notion is relevant to the individuation of knowledge. But this is clearly false, and not something a physicalist should appeal to, for all the reasons given in the previous section.

In *The Facts of Causation* (Mellor 1995), written after the essays in philosophy of mind under discussion here, Mellor makes a distinction between facts and what he calls *facta*. Facts are the 'shadows' of truths – if it is true that *p* it is a fact that *p*. Facta are the truthmakers for truths; it is an empirical question which facts there are, just as it is an empirical question which properties there are. So we should not infer difference of facta from difference of facts; facta and facts do not stand in one–one correspondence. This distinction, which for present purposes corresponds to the distinction between fine-grained and coarse-grained facts, marks a terminological departure from his earlier work, in which (as we saw) Mellor claimed that there were no subjective facts. In the earlier work, the term 'fact' was reserved for truthmakers only. The terminological change is welcome, since without it Mellor would have to deny that the ancients learned a new fact when they learned that Hesperus is Phosphorus – he would have to express what is new about the ancients' condition in a different way. But the terminological change cannot help Mellor in his campaign against the knowledge argument, as we shall see below.

I think that we should agree with Mellor that both notions of fact (or the notions of fact and factum) have their place. This is consistent with what I said above, namely that the objects of knowledge are normally individuated in a fine-grained way. Maybe sometimes we individuate the objects of knowledge in a coarse-grained way. That is perfectly acceptable too. But so long as we do also individuate objects of knowledge in a fine-grained way, then we should accept the conclusion that Mary learns a new fact.

Having said what the argument means by 'fact' we can now turn to 'physical'. What we are asked to imagine is that the knowledge which one acquires about colours inside Jackson's black-and-white room is stated in the language of physics. But it would not help Mary if she learned things in the room which were in the language of psychology and physiology. Nor would it help her if she learned a fully developed dualist psychology (if there were such a thing)

talking about states of consciousness while explicitly acknowledging their utterly non-physical nature. None of these theories would help tell her what it is like to see red. The point is not that the kind of knowledge she gains in the black-and-white room is physical knowledge; rather, the point is that it is the sort of knowledge that can be stated in some form or another: it is 'book-learning'. As David Lewis puts it, the 'intuitive starting point wasn't just that *physics* lessons couldn't help the inexperienced to know what it is like. It was that *lessons* couldn't help' (Lewis 1990: 281; see also Mellor 1991c).

So although physicalism – understood as the view that all facts are physical facts – is one of the targets of the argument, it is really an instance of a more general target: the view that all knowledge of the world is the kind that can be imparted in lessons, without presupposing any particular kind of experience. Thus any view which was committed to this view of knowledge would come within the knowledge argument's range. Likewise with Cartesian dualism – one could not know what it is like to see red, the argument implies, even if one learned the complete Cartesian theory of the mind.

Paul Churchland has argued that this feature of the argument shows that it proves too much.[12] He thinks that Jackson's argument involves a 'logical pathology': it 'makes any scientific account of our sensory experience entirely impossible, no matter what the ontology employed'. But this is plainly a non sequitur: all that follows from the knowledge argument is that if one *knew* the full scientific account of our sensory experience, it would not follow that one knew what it was like to have the experience. This entails nothing about whether such a full scientific account of the workings of our senses can be given. Now Churchland himself identifies this as the main issue at one point:

> if it works at all, Jackson's argument works against physicalism not because of some defect that is unique to physicalism; *it works because no amount of discursive knowledge, on any topic, will constitute the nondiscursive knowledge that Mary lacks.*
>
> (Churchland 1997: 574)

But he takes this to be connected to the claim that any scientific account of experience must be impossible. This, I think, is a mistake, for the reason just given. (Note that since I think Mary gains propositional knowledge, I would not identify 'discursive' with 'propositional'.)

It is true that what Mellor calls 'the factual scope of objective science' is shown to be restricted by the knowledge argument. For no scientific account of vision will tell the blind what it is like to see, and I have argued that what the blind lack here is (in addition to ability knowledge and acquaintance knowledge) propositional knowledge. These pieces of propositional knowledge – these kinds of fact – are what objective science cannot express. But no-one should expect it to; this should not be seen as a mysterious 'restriction' on the powers of science.

I conclude that there is no fallacy in the knowledge argument; but perhaps

now we are beginning to see that its conclusion is stated rather misleadingly, i.e. as an objection to *physicalism*. For even if physicalism is the view that all facts are physical facts, the knowledge argument is an objection to more than this (so far, Churchland is right). It is really an objection to the view that all facts are, so to speak, 'book-learning' facts: *facts the learning of which does not require you to have a certain kind of experience or occupy a certain position in the world*. As Jackson (1997: 569) says, 'you do not need colour television to learn physics or functionalist psychology'. 'Objective' would be a good name for these facts. And 'subjective' would therefore be a good name for *those facts the learning of which requires that one has certain kinds of experience*, *or occupies a certain position in the world*, etc. This is why I say that the knowledge argument is an argument for the view that there are *subjective facts*. It is an argument which shows that in order to gain new knowledge of a certain sort – to learn new facts – you have to have experiences of a certain sort.

That there are subjective facts in this sense should not really come as a surprise. For another example of a fact whose apprehension depends on the subject's specific location in space and time, consider the case of indexical knowledge. Consider, for example, Vladimir lost in the forest; he consults his compass and a map and remarks with relief 'I am here!', pointing to a place on the map. When Vladimir exclaims 'I am here!', pointing at the map, this is something he learned. He now knows where he is, and he did not before. In a classic paper, John Perry (1979) describes himself following a trail of sugar around a supermarket, intending to tell the shopper from whom it came that he was making a mess. When Perry realized that *he* was making a mess he learned something, which he expresses by saying 'It is me! I am making a mess!'. And this piece of knowledge is distinct from the knowledge he would express by saying 'The shopper with the leaking sugar bag is making a mess'. Both examples of new pieces of knowledge require one to have a certain position in the world: Vladimir and Perry cannot learn what they learn without occupying certain positions, or being the people that they are. In particular, they cannot learn these pieces of knowledge, these facts, from books. How could they? (Some writers have noted here the analogy with the knowledge argument. I will discuss this further below.[13]) What Mary, Vladimir and Perry have all learned are *subjective facts*.

Mellor may try to neutralize this conclusion at this point by appealing to the distinction between facts and facta. Perhaps he may admit that there are subjective facts in the sense of subjective *truths*, or in the sense of objects of knowledge (so long as objects of knowledge are individuated by sense rather than solely by reference). After all, Mellor will not want to deny that Vladimir and Perry learn something new, since his account of time requires that indexical propositions are genuinely distinct propositions from their non-indexical truthmakers (see Mellor 1998). But he may say that the original denial of subjective facts should now be interpreted (in the terminology of *The Facts of Causation*) as a denial of subjective *facta*, or truthmakers. That is, even if Mellor

were persuaded by my argument that Mary does learn a new fact, and that her situation is relevantly like the indexical case, he may nonetheless say that this is just another way of saying that there are subjective *truths*. What really matters is the denial of subjective *facta*. And this, as the indexical analogy shows, is untouched by the knowledge argument.

But what would a subjective factum be? A subjective fact, as I defined it above, is a fact the learning of which requires that the learner has a certain kind of experience or occupies a certain position in the world. Facta, by contrast, are not learned: they are what *make true* the truths that are learned. So maybe we could say this: a subjective factum is the truthmaker for a subjective truth or fact. Or a subjective factum is what has to exist in order for a subjective fact to be learned. (This is approximate, but nothing here depends on its being more precise.) So what needs to be the case for Mary to learn that red looks like this? An obvious part of the answer is: a visual experience of red. Mary's visual experience of red needs to exist if she is to learn that red looks like this. Now if a subjective factum is an experience, then no-one should deny the existence of subjective facta; for the issue is not about the existence of *experiences*. Experiences are subjective in the sense that they depend on the existence of experiencing *subjects*; but Mellor does not deny the existence of experiencing subjects (e.g. Mary) either. So what could Mellor be denying if he were to deny that there are subjective facta?

The objective–subjective distinction I drew above was between different kinds of knowledge. Admittedly, it is hard to see how it clearly applies to kinds of entity. Mellor should certainly say that one of the facta which constitute the truthmaker for Mary's knowledge that *red looks like this* is Mary's experience of the tomato. And this experience might be called a subjective entity in the sense that it is an entity which is dependent on a subject of experience. The experience could be called a subjective factum, then. So it seems that Mellor must accept that there are subjective facts and that (in so far as the idea makes sense) there are subjective facta too, since there are experiences. The fact–facta distinction does not help Mellor to sustain his earlier denial that there are subjective facts.

I have argued that Mellor and the physicalist should accept that there are subjective facts. The question now is how this can be made compatible with more plausible versions of physicalism and Mellor's objectivism; that is, versions which do not say that all facts are physical or objective.

6   Physicalism and objectivism revisited and redescribed

The knowledge argument takes physicalism to be the view that all facts are physical. Given what it means by 'fact', this means that all propositional knowledge is physical. And given what is meant by 'physical', this means that all knowledge is the kind of knowledge which can be learned inside a

scenario such as the black and white room – that is, without having to have any particular kind of experience. So the target of the argument is that all facts are 'objective facts'. And this is the view that the knowledge argument refutes conclusively.

But, why should physicalists have to say that all *knowledge* is physical in this sense? Indeed, why should physicalism be a thesis about knowledge at all? Physicalism is a view about what there is, and only derivatively about how we know it. The strongest and clearest motivation for physicalism, I have argued, comes from its claim to explain mental causation.[14] In order to do this, physicalism need not be committed to the view that all knowledge must be expressible without the expresser having to have any particular experiences. It just needs to be committed to the idea that physics is *causally closed*, not even to the view that physics is *explanatorily adequate*.[15] Therefore, physicalism does not need to say that *physics must state all the facts*. (The idea that it must may derive from the image of the book of the world, with all the truths written down in the one true story of reality. But the image is misleading; if what I say here is right, there could never be such a book. For the book cannot express the proposition that Vladimir expresses when he says 'I am here!' and that Mary expresses when she says 'red looks like this!'.)

It is at this point – rather than in the mistaken attempt to dispute the argument's second premise – that the physicalist should appeal to the parallel with indexicality. The idea that Vladimir and Perry gain new knowledge – knowledge of new facts – is compatible with every object and property involved in these stories being physical, in the sense of *the subject matter of physical science*. And it is compatible with every object and property being objective, in the sense relevant to Mellor's objectivism: *the subject matter of objective science*. The fact that these pieces of knowledge are only available from certain perspectives does not entail that there are some further non-physical/non-objective objects and properties involved in the these situations. What is subjective are the facts.

Now many have made the connection between indexicality and the knowledge argument. But it is important to emphasize that, to appreciate it, we do not need to enter the debate about what is the correct theory of facts or resolve the question of how to individuate propositions.[16] And we do not have to make the implausible move that Mary learns nothing that is really new. All we need is to recognize that there is knowledge which can only be had from certain points of view: knowledge of subjective facts. This knowledge will not be physical knowledge in the knowledge argument's sense. And it will not be objective knowledge in Mellor's sense. But this should not worry Mellor or the physicalist. Surprising as it may seem, a physicalist can (and should) sensibly deny that all knowledge is (in the relevant sense) physical knowledge.[17] And he or she should therefore deny that all facts are physical facts. And Mellor should deny that all knowledge is (in the relevant sense) objective knowledge – that is, knowledge of objective facts. He should therefore deny that all facts are objective facts.

A number of writers have drawn attention to the fact that the argument

moves from epistemological premises to a metaphysical conclusion.[18] Mellor says that the existence of subjective facts has 'been falsely inferred from certain kinds of knowledge' (Mellor 1991c: 1). In so considering the matter, Mellor and others have tried to find something wrong with the argument. But, as I have tried to show, there is nothing wrong with the argument, there is no false inference. Indeed, demonstrating exactly what the argument achieves should in itself tell us why we should not be worried by it. So long as Mellor and the physicalist do not hold that all knowledge is physical or objective, that all facts are physical or objective, or that physics must be 'explanatorily adequate' – or that objective science can state all the facts – then the knowledge argument poses no objection to Mellor or to the physicalist. It tells us, rather, something important about our knowledge, something even physicalists and hard-headed objectivists like Mellor must accept.

At the beginning of this chapter, I said that a common theme in Mellor's work is that we should not confuse aspects of the subject's representation of reality with aspects of reality itself. Saying that there are subjective facts in the sense I have defended here is not to make any such confusion. For subjective facts are simply facts about our subjectivity. And these facts are, if you like, facts about the subject's representation of reality. Putting it this way, we can see that there should be nothing out of keeping with the fundamental spirit of Mellor's metaphysics in allowing facts about our subjectivity to be facts about the subject. For what else, after all, should we expect them to be?

Notes

1 An earlier version of this chapter was presented at the Philosophy of Science seminar at the Eötvös Loránd University, Budapest, at the Universities of Birmingham, Oslo and Wales (Swansea), and at the conference Mind and Action III at the Institute for Philosophy of Language, Lisbon. Many thanks to Brian McLaughlin, my commentator at the Lisbon conference, for his comments there (and for the Russell quote); to Hallvard Lillehammer for his excellent editorial advice; and to Katalin Farkas, Carsten Hansen, Penelope Mackie, Greg McCulloch, Harold Noonan, Alex Oliver and James Tartaglia for discussion and criticism. And special thanks to Hugh Mellor, without whom I would never have learned enough to realize why he is wrong about subjective facts.

2 Lewis (1990) acknowledges a debt to Nemirow (1990).

3 See Robinson (1982) and Jackson (1982). It should be noted that Jackson has changed his mind about what the knowledge argument shows (see Jackson 1995). For Jackson's physicalism, see Chapters 1 and 2 of Jackson (1998). If I am right in what I say here, he did not need to change his mind about the soundness of the argument, even after his conversion to physicalism; he just needed to redescribe the conclusion. Of the many discussions of Jackson to which I am indebted, I must single out Horgan (1984).

4 See the references to the statements of the argument by Jackson and Robinson in Note 3. In its essence, the argument has a longer history than this, of course. Earlier twentieth-century sources are Feigl (1958: 68) and Broad (1926: 71).

5 For a useful catalogue of responses to the knowledge argument, see Van Gulick (1997: 559–63).

6 For excellent discussion of this, see Chapter 8 and especially p. 171 of Moore (1997) and Snowdon (forthcoming).

7   I must ignore here the bearing this point has on the famous Frege–Geach problem.
8   This is the line taken by Churchland (1985).
9   But see Churchland (1985). Dennett (1991) launches a general attack on the methodology of thought-experiments as a way of learning about consciousness.
10 See Chalmers (1996).
11 For these views, see Austin (1961), Frege (1967), Davidson (1984) and Mellor (1995).
12 In Churchland (1985); see also Churchland (1997: 574). Jackson (1997) attempts to answer this criticism, but on the implausible grounds that there is a difference between the kind of knowledge a dualist psychology would give and the kind a physicalist theory would give.
13 For the use of the parallel with indexicals as a response to the knowledge argument, see Rey (1997).
14 See Papineau (2001) and Loewer (2001).
15 Lewis (1966) argues that physics has 'explanatory adequacy'; but the argument from mental causation to physicalism only needs the claim that physics is casually closed, not that it is explanatorily adequate (see Crane 2001: §12).
16 So I disagree with Van Gulick (1997: 562–3) that this is the most fruitful line to pursue.
17 Here I agree with Tye (1995).
18 See Jackson (1995), Lewis (1990), Levine (1993), Horgan (1984), among many others.

References

Austin, J. L. (1961) 'Unfair to facts', in *Philosophical Papers*, Oxford: Clarendon Press.
Broad, C. D. (1926) *The Mind and its Place in Nature*, London: Routledge and Kegan Paul.
Chalmers, D. (1996) *The Conscious Mind*, Oxford: Oxford University Press.
Churchland, P. (1985) 'Reduction, qualia and the direct introspection of brain states', *Journal of Philosophy* 82: 8–28.
—— (1997) 'Knowing qualia: a reply to Jackson', in N. Block, O. Flanagan and G. Güzeldere (eds) *The Nature of Consciousness*, Cambridge, MA: MIT Press.
Crane, T. (2001) *Elements of Mind*, Oxford: Oxford University Press.
Crane, T. and Mellor, D. H. (1991) 'There is no question of physicalism', in D. H. Mellor (ed.) *Matters of Metaphysics*, Cambridge, UK: Cambridge University Press.
Davidson, D. (1984) 'True to the facts', in *Inquiries into Truth and Interpretation*, Oxford: Oxford University Press.
Dennett, D. (1991) *Consciousness Explained*, London: Allen Lane.
Feigl, H. (1958) *The 'Mental' and the 'Physical'*, Minneapolis: University of Minnesota Press.
Frege, G. (1967) 'The thought: a logical inquiry', in P. F. Strawson (ed.) *Philosophical Logic*, Oxford: Oxford University Press.
Güzeldere, G. (1997) 'Approaching consciousness', in N. Block, O. Flanagan and G. Güzeldere (eds) *The Nature of Consciousness*: Cambridge, MA: MIT Press.
Horgan, T. (1984) 'Jackson on physical information and qualia', *Philosophical Quarterly* 34: 147–52.
Jackson, F. (1982) 'Epiphenomenal qualia', *Philosophical Quarterly* 32: 127–36.
—— (1995) '"Postscript" to "What Mary did not know" ', in P. Moser and J. D. Trout (eds) *Contemporary Materialism*, London: Routledge.

—— (1997) 'What Mary did not know', in N. Block, O. Flanagan and G. Güzeldere (eds) *The Nature of Consciousness*: Cambridge, MA: MIT Press.

—— (1998) *From Metaphysics to Ethics*, Oxford: Oxford University Press.

Levine, J. (1993) 'On leaving out what it's like', in M. Davies and G. Humphreys (eds) *Consciousness*, Oxford: Basil Blackwell.

Lewis, D. (1966) 'An argument for the identity theory', *The Journal of Philosophy* 63: 17–25.

—— (1990) 'What experience teaches', in W. G. Lycan (ed.) *Mind and Cognition*, Oxford: Basil Blackwell.

Loar, B. (1997) 'Phenomenal states', in N. Block, O. Flanagan and G. Güzeldere (eds) *The Nature of Consciousness*: Cambridge, MA: MIT Press.

Loewer, B. (2001) 'From physics to physicalism', in C. Gillett and B. Loewer (eds) *Physicalism and its Discontents*, Cambridge, UK: Cambridge University Press.

Mellor, D. H. (1991a) 'Objective decision making', in D. H. Mellor (ed.) *Matters of Metaphysics*, Cambridge, UK: Cambridge University Press.

—— (1991b) 'I and now', in D. H. Mellor (ed.) *Matters of Metaphysics*, Cambridge, UK: Cambridge University Press.

—— (1991c) 'Nothing like experience', in D. H. Mellor (ed.) *Matters of Metaphysics*, Cambridge, UK: Cambridge University Press.

—— (1995) *The Facts of Causation*, London: Routledge.

—— (1998) *Real Time II*, London: Routledge.

Moore, A. W. (1997) *Points of View*, Oxford: Oxford University Press.

Nemirow, L. (1990) 'Physicalism and the subjective quality of experience', in W. G. Lycan (ed.) *Mind and Cognition*, Oxford: Basil Blackwell.

Papineau, D. (2001) 'The rise of physicalism', in C. Gillett and B. Loewer (eds) *Physicalism and its Discontents*, Cambridge, UK: Cambridge University Press.

Perry, J. (1979) 'The problem of the essential indexical', *Nous* 13: 3–21.

Rey, G. (1997) 'Sensational sentences', in N. Block, O. Flanagan and G. Güzeldere (eds) *The Nature of Consciousness*, Cambridge, MA: MIT Press.

Robinson, H. (1982) *Matter and Sense*, Cambridge, UK: Cambridge University Press.

Russell, B. (1927) *The Analysis of Matter*, London: George Allen and Unwin.

Snowdon, P. (forthcoming) 'Knowing how and knowing that: a distinction and its uses reconsidered'.

Tye, M. (1995) *Ten Problems of Consciousness*, Cambridge, MA: MIT Press.

Van Gulick, R. (1997) 'Understanding the phenomenal mind', in N. Block, O. Flanagan and G. Güzeldere (eds) *The Nature of Consciousness*, Cambridge, MA: MIT Press.

6   From H$_2$O to water

The relevance to a priori passage

Frank Jackson

1   Background

*Physicalist*s are committed to holding that the physical necessitates everything else which is the case. If physicalism is true of our world, the physical nature of our world fully determines where the shopping centres are and which ones are the biggest, what you and I are feeling and thinking, the current rate of inflation, and so on. The details are controversial but the basic idea is not.

What is controversial – in principle and not just detail of formulation – is whether physicalists are committed to holding that some suitably rich, true account of the physical way things are a priori entails the psychological, political, social, weather, etc., way things are. Are physicalists committed to what I call the a priori passage principle, the view that for each true statement concerning our world, there is a statement in physical terms that a priori entails that statement?[1]

Deniers of a priori passage point out following Kripke (1980) and Putnam (1975) that there are conditionals of the form 'if things are thus and so H$_2$O-wise, then things are such and such water-wise', and 'if things are thus and so molecular kinetic energy-wise, then things are such and such heat-wise in gases', which are necessary a posteriori truths that go from truth to truth but which are not a priori.

I think the example of H$_2$O and water is a bad one for the deniers of a priori passage. My argument (in Jackson 1992; 1994; 1998: 80–3) for this conclusion has a simple structure. I take a putative inference from some truth about how things are framed in terms of H$_2$O to something about how things are framed in terms of water, which is valid in the sense of being necessarily truth-preserving but which is not valid in the sense of being a priori. I then provide an additional, true, contingent, a posteriori premise about how things are framed in terms of H$_2$O, which, I argue, makes the inference valid in the a priori sense. Although I turn the trick for only a very small number of cases, once you have seen one, you have seen them all. It is obvious how to generalize to other necessarily truth-preserving inferences from H$_2$O to water, and how to extend to similar examples involving gold, heat, etc. and, indeed, all the examples that arise from the standard examples of necessary a posteriori

truths. This means that, although the deniers of a priori passage have correctly pointed to the existence of some inferences of the kind in question that are not a priori, they have given no reason to hold that all are – quite the contrary in fact. And this is the issue on the table – the a priori passage principle says that for each sentence that says how things in our worlds are, there exists *some* set of physical premises that leads a priori to it.

Ned Block and Robert Stalnaker (1999) have recently made a number of criticisms of my argument.[2] This reply addresses what I take to be the parts of their attack that are of most interest and which have been seen by others as most calling for a response. Although my focus will be on defending my argument, I will do so in the context of its role as part of the overall defence of a priori passage.

2   Some preliminaries

The issue before us is whether or not physicalists are committed to holding that a rich enough account of the physical nature of the world a priori entails all the truths. What does 'physical' mean here? A short answer is nature as revealed by, and framed in terms of, completed physics and physical chemistry. This is not the place to try and spell this out further.[3] I am going to assume, as do Block and Stalnaker, that we have some reasonable grasp of what is meant, but it will be important later to address what might be meant by the *purely* physical, or physics *itself*. Block and Stalnaker link their criticisms of me with criticisms of one of my allies, David Chalmers.[4] Chalmers talks mostly of microphysics rather than physics more generally, and Block and Stalnaker often phrase their points in terms of microphysics. In these terms, their target is the view that physicalists are committed to holding that a rich enough account of the world's microphysical nature a priori entails all the truths. Nothing here turns on this and I will slide between the two ways of characterizing the issue. Also, I will take it for granted that physical or microphysical accounts include the relevant laws of nature framed in physical or microphysical terms.

A more important issue is the difference between the claim that a rich enough physical or microphysical account of the world a priori entails all the truths of psychology, shopping centres, wars, etc., and the claim that physicalists are committed to the possibility of conceptual analyses of psychology, shopping centres, wars, etc., in terms of microphysics. At places, Block and Stalnaker characterize their target in these terms. However, we know that the latter is impossible. The conceptual possibility of multiple realizability tells us this. Many, myself included, hold that multiple realizability in the sense of realizability in non-physical stuff – ectoplasm – of psychology, shopping centres, wars, etc., is metaphysically possible, but I know of no-one who denies the conceptual possibility of realization in ectoplasm. This means that we can rule out immediately the possibility that any biconditional linking physical or microphysical nature with, say, psychological nature could be a priori. The

point is equally obvious for biconditionals linking the microphysical with shopping centres.

The final preliminary concerns the well-known 'stop clause'. Does enough information about individual heights a priori entail what the average height is? The answer depends on how we understand 'enough information about individual heights'. If we restrict ourselves to information of the form A is 6′, B is 5′ 7″, … , that is, to a list of all the distinct people that exist along with their heights, the answer is no. For we need to be given that this list is the complete list; we need something like 'and that's the lot' – the stop clause. This kind of point has a long history under the heading of the debate over whether to admit general facts, or states of affairs of totality, into our ontology.[5] But although the ontological issue is controversial, the deduction issue is not: it is agreed that often we need stop clauses to make deductions. What, in consequence, should we say about the claim that enough information about individual heights a priori entails what the average height is? The wrong thing to say is that we have discovered a major issue that requires adjudication. Rather, we have a terminological issue that requires us to draw a distinction: if information about individual heights excludes the stop information, you cannot a priori deduce average heights from such information; if it includes it, you can.

A similar situation applies in the case of the a priori passage principle. Rich enough physical information might be read so as to include a stop clause, or it might be read so as to exclude a stop clause. Read without the stop clause, the principle is certainly false. I will make the point first for a toy example. Consider a world w_3 that contains three electrons whose nature is as conceived in current physics, and nothing further. Whatever may be true of our world, we can all agree that physicalism is true at w_3, and we can all agree that w_3 does not contain any shopping centres. Can we a priori deduce that fact from a rich enough account of the physical nature of w_3? The obvious answer is yes – someone who thinks that three electrons might make up a shopping centre does not have our concept of a shopping centre. However the answer is no, unless the stop clause, carried in this case by the words 'and nothing further', is included as part of the rich enough physical story about w_3. A world with three electrons might have much else besides, including shopping centres.

We defenders of a priori passage, or at least all the ones I know, are explicit that the physical information which we claim a priori entails where the shopping centres are, who is thinking what, when inflation peaked, etc., must in general include the stop clause. The situation is as follows. Exclude the stop clause in what is meant by a rich enough physical account and you get a principle that is certainly false and which no-one defends; include the stop clause and you get a principle that has some chance of being true and is the one, give or take points of precise formulation, that some people, including me, Chalmers (1996) and David Lewis (1994), accept. This point will be important later.

3   The target argument

Consider [under the assumption that we have the percentage correct, it will be important to suppose that (a) below is true]:

(a) Sixty per cent of the earth is covered by H$_2$O.

Therefore,

(d) Sixty per cent of the earth is covered by water.

The passage from (a) to (d) is not a priori, although it is necessarily truth preserving. However, many have supposed that something like (formulations vary)

(b) Water is the stuff that plays the water role.

is a priori, where the water role is spelt out in terms of being potable, odourless, falling from the sky, being the stuff that makes up various bodies of liquid of our acquaintance or in some ostended set of samples, etc. In short, the water role is spelt out in terms of the reference fixers for 'water', and the case for (b)'s a priori status rests on the general thesis that 'N = the F' is a priori when 'F' specifies the reference fixers for 'N'.[6] If (b) is a priori, then the conjunction of (a) with the empirical truth that

(c) H$_2$O is the stuff that plays the water role.

means that we have two 'H$_2$O truths' that together a priori entail (d).[7]

What is the significance of this result? It tells us that it is a mistake to infer from the fact that 'Any water is H$_2$O' is necessary a posteriori that there is no a priori passage from the way things are framed in terms of H$_2$O to the way they are framed in terms of water. Of course, this presupposes that we have to hand a way of spelling out 'the water role' in (b) which plausibly both makes (b), or something suitably like it, a priori and does not contain the term 'water' or an equivalent. I will follow the practice of using the term 'water role', or 'waterish' (and 'heatish' and 'heat role' when discussing the case of heat in gases), but it is important that such expressions be viewed as shorthand for longer expressions that do not contain 'water' (or 'heat').[8]

I will address the following criticisms that Block and Stalnaker make of the argument.

(1) They argue that the definite description in (b) makes trouble for the claim that (b) is a priori and that this matters.
(2) They argue that the definite description in (c) means that the argument

   is, as a matter of principle, unsuitable to be a suggestive model for the
   discussion of the a priori passage principle in general.
(3)   They argue that the example of scientific reductions gives no reason for
   holding that (b) is a priori.

   I will address these criticisms more or less in that order. I will conclude
with a short statement of the positive reason for holding that something like
(b) has to be a priori. When you hear the reason, you will understand why I
keep on saying that *something like* (b) is a priori. I sometimes (understandably)
meet the complaint that I should be able to say *exactly* what is a priori – I can
hardly plead lack of empirical data to excuse my vagueness – but we will see
why it has to be 'something like (b)' that is a priori.

4   Is it a priori that water is *the* stuff that plays the water role?

Block and Stalnaker point out, correctly, that the uniqueness of the stuff that
plays the water role is important to the target argument. For example, from
the premises that H_2O is one of the kinds that plays the water role, and that
water is one of the kinds that plays the water role, nothing follows a priori
about the distribution of water from the distribution of H_2O. Consider the
following analogy: from the fact that drug X is *a* cure for malaria and drug Y
is *a* cure for malaria, it does not follow that X = Y.
   Block and Stalnaker see the need for uniqueness as making serious trouble
for the target argument in two different ways. One way arises from doubts
about its being a priori that water is *the* waterish stuff, that is the unique
waterish stuff. Block and Stalnaker point out that there is a case to be made
that there might have been, in the sense that it is conceptually possible,
more than one kind of water. Water might have turned out to be like jade,
something that comes in two kinds.[9] In fact, they go further and suggest that
it is conceptually possible that water is the role property. On the face of it,
this cannot be right: it is the occupants of the water role that do the things
that we all agree water does, but I take it they mean that it is conceptually
possible that there be indefinitely many different kinds that play the water
role consistently with each being water.
   I think they are right that it is conceptually possible that water is like jade,
although I think they go too far when they suggest that it is conceptually pos-
sible that there might be an awful lot of different kinds that are water. I think
it is part of our concept of water that there are at most only a few natural kinds
that are water. But we do not need to debate the issue here; it is irrelevant in
the present context. Provided only that

   (b*)   Given there is a unique stuff that plays the water role, it is water.

is a priori, the target argument is a priori valid. That is, if (b*) is a priori, then

the conjunction of (a) and (c) a priori entails (d). The reason is the uniqueness built into the empirical premise (c) – 'H$_2$O is *the* stuff that plays the water role'. If (b*) is a priori, it follows a priori from (c) that the distributions of water and H$_2$O go together. Hence, the a priori nature of (b*) shows that (a) and (c) a priori entail (d). To put the matter in terms of our earlier analogy: it does not follow from the fact that drug X is a cure for malaria and drug Y is a cure for malaria that X = Y. But it does follow from the fact that, given there is a unique cure for malaria, it is drug X and drug Y is the unique cure for malaria that X = Y

Is it plausible that (b*) is a priori? I will return to the general question of why and how some statement like (b*) is a priori later. For now, let us simply note that although the literature is full of plausible cases where, in various *counterfactual* worlds, the unique waterish stuff in those worlds fails to be water – by being XYZ as it might be – there are none where the *actual* unique water-ish stuff fails to be water. There are no plausible possible cases we describe as showing that it 'might have turned out' that the unique waterish stuff is not water. Indeed, when we describe the metaphysically and conceptually possible case where the unique waterish stuff turns out to be XYZ, we promptly go on to describe it as water turning out to be XYZ.

My reply to Block and Stalnaker above depends on the uniqueness part of the empirical premise (c). It is, therefore, crucial that I address their principled objections to the uniqueness part of (c).

5   Are uniqueness claims part of microphysics and does this matter?

We noted above that the uniqueness part of (c) is crucial. If we replace (c) by

(c*) H$_2$O is a stuff that plays the water role.

the target argument fails. That is, (a) and (c*) together do not a priori entail (d). This is the case regardless of whether or not uniqueness is part of the concept of water, regardless of whether the issue that we have just been noting can be set aside. However, the correct description of why the target argument fails varies depending on how that issue should be resolved. Suppose, first, that it is part of the concept of water that it is the single filler of the water role. On this supposition, the problem is that the conjunction of (a) and (c*) does not a priori entail that there is any water. This is because (c*) leaves open as a conceptual possibility that there is more than one filler of the water role, in which case there would be no water to be co-distributed with H$_2$O. Suppose, secondly, that it is part of the concept of water that there are at most some smallish number of fillers (the view I in fact favour). On this supposition, there are two problems. First, again it is the case that the conjunction of (a) and (c*) does not a priori entail that there is any water. This is because (c*) leaves

open as a conceptual possibility that there are very many fillers of the water role, in which case there would be no water to co-distribute with H_2O on the second supposition. Second, on this second supposition, (c*) leaves open the conceptual possibility that there are a limited number of other fillers which count as water. But this means that it is conceptually possible that water covers more than 60 per cent of the earth by virtue of one of the other fillers of the role covering, say, an extra 20 per cent. Hence, it is conceptually possible that (a) and (c*) be true together when it is 80 per cent, not 60 per cent, of the earth that is covered by water even though H_2O only covers 60 per cent. Suppose, finally, that the concept of water allows indefinitely many fillers to be water. On this supposition, the problem is that it is conceptually possible that there are other fillers which count as water. But this means that it is conceptually possible that water covers say 80 per cent of the earth consistently with H_2O being one of the fillers and covering 60 per cent. In sum, independently of whether or not it is part of the concept of water that it is the only filler, or one of the few fillers, of the water role, or whether it is open slather, the conjunction of (a) and (c*) fails to a priori entail (d).

Now, the target argument does *not* contain (c*); it contains (c), and this was no accident. But Block and Stalnaker see a major problem here. They do not doubt that (c) is true. We do know that there is only one filler of the water role. Their key contention is that the uniqueness part disqualifies (c) from being a statement in physics or microphysics. Of course, (c) is not, and was never supposed to be, a statement in microphysics. But their contention is that the uniqueness part of (c), the fact (c) says that H_2O is *the unique* stuff that fills the water role, is *in itself* enough to stop (c) being a statement of microphysics itself, or a claim purely about microphysics, or something that can be extracted from microphysics itself, as they variously express their contention. So that we have here, in their view, a problem of major principle for my argument, not a detail that can attended to down the track in the way that the presence of the word 'water' in 'water role' is arguably a detail that can be attended to down the track.

Why do they hold this? The point comes up at a number of places. They note, for example, that the issues that arise for water also arise for heat. As H_2O and water are to the water role and being waterish, so molecular kinetic energy and heat (in gases) are to the heat role and being heatish. And they urge that 'the claim that mean molecular kinetic energy = the (unique) heatish stuff around here is not a purely microphysical claim, since it rules out the possibility that ghost heat is also a heatish stuff around here' (Block and Stalnaker 1999: 18). Later, they say '[g]iven the possibility of ghost water that covers part of the earth not covered by physical water, it cannot follow from microphysics that water covers 60 percent of the earth' (Block and Stalnaker 1999: 28) The theme that unites the various presentations of the point is that if a claim excludes ectoplasm or ghost stuff from doing something, then the claim is not purely microphysical (or physical) or extractable from microphysics itself.

From this, Block and Stalnaker infer that it is never legitimate for a

defender of the a priori passage principle to help his- or herself to premises with uniqueness and similar claims in them, for to do that is to go beyond the microphysics. But if this is right, it is not only the target argument that is in trouble. The reason goes back to our earlier discussion of the stop clause. The uniqueness part of (c) is a special case of the stop clause. To make the a priori derivation to water distribution, we need to rule out the conceptual possibility that ectoplasm fills the water role in the same way that in the general case we need the stop clause to rule out shopping centres made of ectoplasm. Block and Stalnaker are essentially making the point we noted earlier – namely, that without stop clauses in the physical or microphysical premises, there will be a great deal about our world that cannot be derived a priori.

My reply follows from what I said earlier in connection with the stop clause. There is no objection to Block and Stalnaker operating with a 'no exclusion of ectoplasm' criterion for being purely microphysical, but if they do they are not addressing the question on the table. The a priori passage principle, on the reading on which it is entertained by me, and has some chance of being true, has a stop clause in its premise set for the reasons canvassed earlier. Block and Stalnaker give a motivation for their 'no exclusion' criterion for the purely microphysical or for being a claim of microphysics itself, as they also put it, in terms of supervenience (Block and Stalnaker 1999: 19). The key idea is that a necessary condition for being a purely microphysical claim or a claim of microphysics itself about our world is being true at all microphysical duplicates of our world. Since microphysical duplicates of our world include worlds with lots of extras made of ghost stuff or ectoplasm, this necessary condition automatically excludes a claim such as 'the unique stuff that does so and so is H$_2$O' from the class of the purely microphysical, because there is a microphysical duplicate of our world where some ectoplasm as well as H$_2$O does so and so.

However, physicists qua physicists make claims that exclude as well as ones that include. The claim that there are four, not five as previously thought, fundamental forces in nature is normally thought of as part of microphysics itself. But it is not a claim that is true at every microphysical duplicate of our world. Some of the microphysical duplicates of our world have fifth or sixth ectoplasmic forces that are every bit as fundamental as our four. The important point, however, is as before. The a priori passage principle is not a claim about what follows a priori from the purely microphysical or from microphysics itself, in Block and Stalnaker's no exclusion sense of the notion of the purely microphysical or of microphysics itself. Indeed, if it were, we could give a *very* quick proof that no physicalist should accept the a priori passage principle. Physicalists hold that a feature of our world is that there is no ectoplasm, but, on the Block–Stalnaker exclusion test, this would mean *ipso facto* that it could not be a priori entailed by the purely microphysical.

Why do Block and Stalnaker think that they are addressing the issue on the table despite the explicit statements by defenders of the principle that it is to be read as containing the stop clause?[10] My best guess is that they think that

there is nothing else good to mean by candidates for the premise set of the a priori passage principle apart from candidates that pass their no exclusion test. But there is. We can allow, as possible members of the premise set, any and every truth about our world expressible using a vocabulary whose descriptive terms are drawn entirely from microphysics. Often, but not always, the premise set will need to include the stop clause, or a stop clause, depending on the conclusion that is in question, but stop clauses never need to contain descriptive terms from outside microphysics. For example, if the conclusion is that there are some solid objects, no stop clause is needed. Enough information about particle aggregations, their shapes and their physical properties will a priori entail that there are solid objects. But if the conclusion concerns where the largest one is, stop clauses that rule out possible ectoplasmic competitors for the title of largest will be needed. But the stop clauses need not contain descriptive terms drawn from outside microphysics. Or suppose we are considering the case where we have a single, huge premise that says everything there is to say in physical terms, on the simplifying assumption that our world is finite and discrete, so that we can think of this as a huge long conjunction. Then the a priori passage principle says *inter alia* that this premise a priori entails how many pains there have been, are or will be, provided that we include in the huge list 'and there is nothing else'. We need the stop clause to close off the conceptual possibility of ectoplasmic pains boosting the number. But although the stop clause excludes ectoplasmic pains, it does not do so by using the word 'ectoplasm'. If I say that there is nothing between two stars, I say *inter alia* that there is no ectoplasm between them, but I do not use the word 'ectoplasm' in doing so.

Finally, it is worth noting that physicalists' own statements of what they hold about our world do not pass the Block–Stalnaker no exclusion test. Physicalists do not hold merely that our world has some physical nature; dualist interactionists agree about that. Physicalists hold that our world is entirely, completely, etc., physical; that there is nothing in it over and above what is there in the microphysics. This means that physicalists' own statement of their position is not a purely microphysical claim on the Block–Stalnaker criterion. In consequence, restricting discussion to what can be derived a priori from the purely microphysical in their sense would be to avoid discussing what can be derived a priori from the very way physicalists themselves characterize our world when stating their view. It also means that if their real concern is over the intelligibility of stop clauses, they should be targeting physicalism itself.

6   Why we should hold that something like (b*) or (b) is a priori

A *sign* that 'given there is a unique stuff that plays the water role, it is water' is a priori is the difficulty of making sense of its turning out to be false, as we noted before. But there is a theoretical reason for holding that it, or something like it, is a priori. The reason is that it needs to be a priori if we are to make

sense of what is going on in scientific reduction. In saying this, I am going directly against Block and Stalnaker's discussion of reduction. They argue that the job some philosophers – David Armstrong and David Lewis are examples – give conceptual analysis and the a priori in accounts of reduction rightly belongs to considerations of scientific methodology.[11] This is not the place to write an essay on reduction, but I hope I can say enough to indicate why I think that Armstrong and Lewis got it right, and why reflection on theory reduction in science supports the view that claims along the lines of (b) and (b*) are a priori.

Let us start with some passages from Block and Stalnaker. They focus on the familiar case afforded by the reduction of the thermodynamic theory of gases:

> The supposition that it is a conceptual truth that heat = the actual unique heatish stuff around here is incompatible with the actual practice of scientific reduction. The claim that heat and molecular kinetic energy are dual occupants of the same role is not false because it falls afoul of the concept of heat. The view that heat and molecular kinetic energy are two rather than one is not contradictory or conceptually incoherent. It is false, and can be shown to be false by attention to certain methodological principles … usually invoked with the misleading name 'simplicity'.
>
> (Block and Stalnaker 1999: 23)

> Levine, Jackson, and Chalmers suppose that the gap between descriptions in terms of microphysics and descriptions in terms of, for example, 'water' and 'heat' is filled by conceptual analysis. A deep inadequacy in this view is revealed by the role of methodological considerations in our actual decisions about such matters. Why do we suppose that heat = molecular kinetic energy? Consider the explanation given above [in terms of molecular motion] of why heating water makes it boil. … If we were to accept mere correlations instead of identities, we would only have an account of how something correlated with heating causes something correlated with boiling.
>
> (Block and Stalnaker 1999: 23–4)

Block and Stalnaker are conflating two questions. One is: 'Why do we hold that there is only one thing that is heatish or fills the heat role, not two closely correlated things?' This question is answered by appeal to the methodological principles they mention, not by appeal to the concept of heat. The second question is: 'Why do we hold that the one thing that is heatish is heat?' The methodological principles they mention do not address this question, let alone answer it. Simplicity or some such tells us it would be wrong to hold that there is something in addition to molecular kinetic energy filling the heatish role, but that leaves us with two options. One is to say that there is no heat, and that the reason for believing in heat, namely its putative explanatory role,

has disappeared because we have discovered that all the explanatory work is done by molecular kinetic energy. The other is to say that molecular kinetic energy and heat are the very same thing. Neither option is simpler than the other – they agree precisely in how many different kinds there are playing the heat role, i.e. one: molecular kinetic energy.

We famously took the second, identification, option. By contrast, in the cases of vitalism and phlogiston, we took the first, elimination, option: we discovered what the occupants of the roles associated with life and combustion are; we rejected the hypothesis of dual occupancy; but instead of identifying, we concluded that there is no phlogiston and no vital force. The principles of scientific methodology that Block and Stalnaker mention do not help us make the choice between eliminating and identifying.

How should we approach the issue of whether to identify or eliminate? To say, as I just have, that certain methodological principles do not help us is not to say or to show that conceptual analysis or the a priori, or anything in that general area, will come to the rescue. This is a big topic, but here is the short version of why the correct approach to this question means that statements like (b) and (b*), and the corresponding ones concerning heat, pressure, and so on, come out a priori.

There is a clear, if hard to analyse, sense in which the kinetic theory of gases gives us a complete account of the nature of a gas in terms of the kinetic (and potential) energy of its constituent molecules, their location and motion in space, their impacts, molecular momentum transfers, and so on. And we know that we can fully explain the behaviour of gases in the terms of the various features recognized and named in the kinetic theory of gases. There is, in particular, no extra feature of gases that we need the words 'heat' and 'pressure' for. This makes it very hard to hold that no matter how much information you have framed in the terms of the kinetic theory and in terms of the functional roles played by the properties picked out by the terms of that theory, and no matter how confident you are that the kinetic theory and its future developments provides a complete picture of the essential nature of gases, the passage from this information to whether or not gases are hot is a posteriori. Because everything relevant about gases can be explained in the terms of the kinetic theory, how can you be justified in going further – and it is going further if you insist that the passage is a posteriori – and holding that gases are hot?

Moreover, there is no great mystery about how the passage from what is said in the kinetic theory to whether or not gases are hot, have pressure and temperature, and so on, might be a priori. It is plausible, as an empirical matter of fact, that we use the words 'heat', 'pressure', etc., for features that play certain roles; and the same goes for the word 'water'. These roles include those we have used the words 'heatish' and 'waterish' for. Once upon a time, this was thought to imply that the words 'heat' and 'water' mean 'stuff that plays such and such a role'; post Kripke, we know that there is another option, namely that the roles are reference fixers rather than meaning givers. Either

way, statements like (b) and (b*), and the corresponding ones for heat, come out a priori. Either way we are carrying out a bit of conceptual analysis, for we are teasing out what we use the words 'heat' and 'water' for and that is what conceptual analysis is in my view. And either way, we have a simple explanation of why it was right to eliminate phlogiston when the oxidation theory of combustion came along. The word 'phlogiston' was used for the stuff whose giving off is an essential part of combustion, and the oxidation theory showed that there is no such stuff.

Some insist that there is no need to think of the heat role and the water role here as being reference fixers for 'heat' and 'water' respectively. They urge that we should think of being waterish (to make the argument with this example) as a folk marker or identification intuition which serves to identify items that might possibly be water.[12] This delivers an initial division into water and non-water. We then investigate how well this typing corresponds to that made in terms of the categories of our best science, and it is these categories that settle whether or not some stuff is water. If best science vindicates the 'folk' typing in sufficiently many cases (whatever precisely that comes to), X is water if and only if it belongs to the right category, or one of the right categories, as discerned by best science; if it does not, there is no such stuff as water. Either way, it is our best science, not the folk marker or identification intuition, that settles the issue at the end of the day. Rhetorically, this sounds like an objection, but it is, in fact, a version of the reference fixing view. To say that x is water if and only if (a) the folk typing matches enough the best-science typing and (b) x is in the best-science class for water is the very same as saying that the reference fixing is on the best-science kinds that sufficiently often are waterish if such there be.[13]

Now – finally – I can say why we have to be vague about what it is that is a priori; why we say that *something like* (b) or (b*) is a priori. Putting names to things, except in some highly circumscribed cases in mathematics or where explicit semantic decisions are called for, is a highly context-dependent, vague, accommodating-oneself to one's fellow speakers and writers, and leaving issues unresolved in the expectation that the need for resolution will never arise matter. Neat formulae are not to be expected. But, as we saw above, this does not matter for the target argument. What matters is that the empirical facts as stated in terms of H_2O are enough to ensure that our world contains stuff that counts – semantically counts – as water.

Notes

1   There are two theses to distinguish.

   (a)   There is a true (huge) statement frameable in physical terms that a priori entails every true statement about what our world is like.
   (b)   For every true statement about what our world is like, there is a (sometimes huge) true statement frameable in physical terms that a priori entails it.

The significant differences between these will not concern us here. I will also fudge the difference between S a priori entails T, and S's being such that one can move a priori from S to T.

2   As the title suggests, the paper addresses a series of surrounding issues, but a good part of it is devoted to attacking my argument. The most significant issue I will not be discussing is their Twin Earth objection to a priori deducibility (except by way of passing reference in a note below).

3   For some spelling out, see Jackson (1998). This spelling out is largely motivated by the challenge of Crane and Mellor (1990).

4   Chalmers (1996). He is an ally in the sense of supporting a priori passage; he disagrees with my current self though not a former self over what to infer from a priori passage concerning the truth of physicalism.

5   Russell (1972: 93–4).

6   Because it is in general a posteriori that the F exists, strictly there should be a 'modulo the existence of the F' added here, but, in the present context, there is no need to include this qualification as the empirical premise a priori entails the existence of the relevant unique F.

7   The claim is *not* that were you told of some possibly non-actual world w* that

(a)   60 per cent of the earth is covered by H_2O

and

(c) H_2O is the stuff that plays the water role

are both true at w*, you could infer without further ado

(d)   60 per cent of the earth is covered by water

is true at w*.

You could not. You would need to know that H_2O is the stuff that plays the water role in *our* world, and that is an additional piece of information. The difference is similar to that between 'P, therefore actually P' being a priori valid and 'P is true at w*, therefore "actually P" is true at w*' not being a priori valid. I note the point because if I understand Block and Stalnaker's criticism of Joe Levine's (1993) views about the a priori deducibility of boiling from enough physical information, they conflate the issue of the a priori validity of the style of argument in the text with that in this note.

8   Some prefer to use 'watery' and 'heatish' for the role minus the acquaintance; accordingly, they say that what is a priori is that water (heat) is the watery (heatish) stuff of our acquaintance.

9   The precise sense in which it might have turned out that water is like jade is the same sense in which it might have turned out that water is XYZ, the sense in which this is epistemically possible. What this precise sense is is controversial in view of the fact that water could not have been XYZ! But there had better be some sense, or else there is no sense in which it is a posteriori that water is not XYZ. We holders of the view that the phenomenon of the necessary a posteriori is a linguistic one have our own way of finding the path through this little minefield (see, for example, Jackson 1998: 84–6), but it would beg too many questions to presuppose our path in a reply to objections that take off from a very different perspective on the phenomenon.

10   For two very explicit discussions, see Jackson (1994; 1998: 26). Chalmers (1996) is equally clear on the point.

11   See, for example, Armstrong (1968) and Lewis (1970).
12   What follows for the case of water appears to be what Block and Stalnaker are saying for the example of life. My discussion here (and elsewhere) is indebted to discussions with David Braddon-Mitchell. I take the term 'identification intuition' from Devitt (1996: 73). In his view, the relevant intuitions sometimes are those of the folk but sometimes are those of one or another body of experts.
13   Which is, of course, the usual version when we want to include ice and steam as water.

References

Armstrong, D. M. (1968) *A Materialist Theory of the Mind*, London: Routledge.

Block, N. and Stalnaker, R. (1999) 'Conceptual analysis, dualism and the explanatory gap', *The Philosophical Review* 108: 1–46.

Chalmers, D. (1996) *The Conscious Mind*, New York: Oxford University Press.

Crane, T. and Mellor, D. H. (1990) 'There is no question of physicalism', *Mind* 99: 185–206.

Devitt, M. (1996) *Coming to Our Senses*, Cambridge, UK: Cambridge University Press.

Jackson, F. (1992) 'Critical notice of Susan Hurley, *Natural Reasons*', *Australasian Journal of Philosophy* 70: 475–87.

—— (1994) 'Armchair metaphysics', in J. O'Leary Hawthorne and M. Michael (eds) *Philosophy in Mind*, Dordrecht: Kluwer.

—— (1998) *From Metaphysics to Ethics*, Oxford: Clarendon Press.

Kripke, S. (1980) *Naming and Necessity*, Oxford: Basil Blackwell.

Levine, J. (1993) 'On leaving out what it is like', in M. Davies and G. Humphreys (eds) *Consciousness: Psychological and Philosophical Essays*, Oxford: Basil Blackwell.

Lewis, D. (1970) 'How to define theoretical terms', *Journal of Philosophy* 67: 427–46.

—— (1994) 'Reduction of mind', in S. Guttenplan (ed.) *A Companion to the Philosophy of Mind*, Oxford: Basil Blackwell.

Putnam, H. (1975) 'The meaning of "meaning"', in *Language*, *Mind and Reality*, Cambridge, UK: Cambridge University Press.

Russell, B. (1972) *The Philosophy of Logical Atomism*, in D. Pears (ed.) *Russell's Logical Atomism*, London: Fontana.

7 Epiphenomenalism and causal asymmetry

Paul Noordhof

It is a great pleasure to contribute to a festschrift for Hugh Mellor. In his articles, books and conversation, he has been one of the most significant influences on my thought.

In a rather too frequently discussed paper – as Mellor would be the first to hold – Crane and Mellor inveigh against physicalism. The common argument offered in favour of physicalism is based on the claim that the physical world is causally closed or, more specifically, that nothing non-physical causes something physical. In response, Crane and Mellor write:

> Our mental states, intentional and otherwise, could – and would – affect our brain states and bodily movements even if the laws of physics made them all determined also by earlier brain states. The claim that a system thus constrained by non-mental laws must be closed, in the sense of being unaffectable by its mental states, simply does not follow – and it is not true.
>
> (Crane and Mellor 1990: 100)

In his later work, Mellor is even more explicit about what he has in mind. He writes:

> overdetermination exists … the fact, if it is a fact, that a mental cause C which neither is nor supervenes upon a physical cause C′ of the same effect E would overdetermine E is no reason to deny that C is as effective a cause of E as C′ is.
>
> (Mellor 1995: 104).

Nevertheless, although Crane and Mellor are willing to countenance overdetermination – even systematic overdetermination – in discussing the causal history of mental events and behaviour, many are not. That is the major reason why most philosophers of mind have become physicalists. A doughty few have tried to square their commitment to the causal closure of the physical world with their conviction that some mental properties are non-physical. These are the 'epiphenomenalists' (Campbell 1970: 124–6; Jackson 1998: 58; and, in some moods, Chalmers 1996: 150–1, 191–203).

This chapter is an attempt to provide a new argument against epiphenomenalism. It draws on areas of philosophy to which Mellor's work has been so central. I shall argue that the proponents of this kind of epiphenomenalism face an unpleasant dilemma. In modern parlance, this might be better described as an 'opportunity'. In order to make epiphenomenalism attractive, they need to appeal to the idea that causation involves asymmetric necessitation. In so doing, they incur an obligation to explain how causation is related to the fact that causes usually precede their effects. Unfortunately, the only plausible account of this to which they can appeal is a causal theory of temporal precedence. Given that mental events and facts are epiphenomenal, this presents a difficulty. Either the temporal location of mental events and facts becomes problematic or, in attempting to deal with this problem, we undermine the motivation for epiphenomenalism in the first place.

1   Epiphenomenalism: its characterization and defence

I shall take epiphenomenalism to be the doctrine that non-physical mental items do not cause anything physical or, for that matter, anything mental. They are just epiphenomena, the froth of life. In calling this doctrine 'epiphenomenalism', I do not wish to rule out the possibility that mental items may be physical yet inefficacious. Such a possibility would have as much right to be called 'epiphenomenalism'. In fact, I think that the usual grounds for suggesting that this possibility might be actual lack foundation (Noordhof 1997; 1999b). However, I do not set aside the possibility of physical mental epiphenomena for this reason but rather because my target in what follows is the relative merits of interactionist dualism and epiphenomenal dualism. In effect, I shall argue that, if one is going to be a dualist, one should be an interactionist. It is unlikely that the argument I develop, should it prove promising in the present case, will extend to all forms of epiphenomenal physicalism.

We can make a further distinction with regard to epiphenomenalism, this time concerning the category of the mental items involved. Token epiphenomenalism rejects the efficacy of mental particulars. For the present purposes, this would also include mental facts. Type epiphenomenalism rejects the efficacy of mental properties and, indeed, kinds of facts (Broad 1925: 472). Token epiphenomenalism is stronger than type epiphenomenalism. If mental particulars, and here I include instances of mental properties, are inefficacious, it is hard to see how mental properties could be efficacious. However, denying the efficacy of mental properties is compatible with allowing that their instances are efficacious. The distinction has most use in discussions which include the possibility of epiphenomenal physicalisms. For instance, it might be claimed that mental properties are type epiphenomenal but token efficacious because instances of mental properties are instances of efficacious physical properties. Some have held what appears to be a type epiphenomenalism because of their insistence that, although mental events and states are physical, they have non-physical phenomenal properties. They then deny

that these non-physical properties – for instance, the hurtfulness of pain – are efficacious (Campbell 1970: 126–7). I think that this is really a case of token epiphenomenalism in which the inefficacious mental particular is a case of hurtfulness. I suspect that the appearance of a commitment to type epiphenomenalism arises through a failure to distinguish between the question of whether an instance of a property is efficacious and the question of whether it is efficacious in virtue of being that very property (Noordhof 1999b: 293–7). In any event, I will focus on token epiphenomenalism in what follows with this refinement in mind (hereafter referred to simply as 'epiphenomenalism'). I guess it is possible that one might be a dualistic type epiphenomenalist while being convinced that dualistic token epiphenomenalism is untenable, but I think that this position would have limited appeal.

Epiphenomenalism has come under sustained attack. I will mention just three issues out of many. It is argued that it abandons the intuitive claim that people do things as a result of their beliefs, desires and sensations. Even those who limit their epiphenomenalism to phenomenal properties have to argue that we do not withdraw our hand from the hot kettle because touching it hurts (Campbell 1970: 125–6). While this might be an acceptable consequence when the withdrawal is just a reflex, it is not in circumstances in which you are trying to see how long you can touch the kettle before it becomes too painful, perhaps out of bravado. It is also argued that epiphenomenalism is incompatible with claiming that we have knowledge of our mental states and events. Knowledge of this sort requires causal contact, the thought runs, and that is the very thing denied by the epiphenomenalist. Finally, it is argued that we would be unable to refer to our mental states and events. Reference in this case also requires causal contact.

I think that all of these objections raise significant difficulties for epiphenomenalism. However, they all suffer from a certain dialectical weakness. They rely upon a commitment to a particular conception of the way that mental events and states are explanatory of behaviour, to a particular view of knowledge or to a particular view of reference. This leaves epiphenomenalists in a rather better position than we would hope. They can question these commitments. They can outline what the epiphenomenalist can say on each of these matters and challenge us to defend the claim that we should take up the stronger commitments about explanation, knowledge and reference threatening to epiphenomenalism. This is, in effect, what David Chalmers (1996: 191–203) does in his defence of epiphenomenalism. I am not saying that an approach of this sort is wholly successful. All I am pointing out is that it places the enemies of epiphenomenalism in an unacceptable position: on the defensive.

The objection that I am going to raise has a different structure. The distinctive feature of the epiphenomenalist position is an insistence that the mental particulars do not cause physical particulars but allow, indeed require, that physical particulars cause mental particulars. The question is: What benefits does this give us? Why is it so unattractive to allow that mental particulars

cause physical particulars but perfectly OK – or, at least, not as bad – to allow the reverse?

Bearing in mind what I have already noted, a natural place to look to answer this question is the justification of the causal closure principle itself. One part of the justification is empirical. It is that we have grounds to believe that, if we take a conjunction of all the laws of physics, we can use them to predict without exception what will happen as far as the subject matter of these laws is concerned. Let me dub this subject matter physical entities in the narrow sense. This does not hold for the laws identified by other sciences. It is these claims about the laws of physics and other sciences that provide the empirical basis for the closure claim. As the original passages quoted from Mellor (and Crane and Mellor) bring out, though, the empirical justification goes only so far. The fact about laws does not imply that the physical world is causally closed. There could be systematic overdetermination. Another part of the justification, therefore, relies upon the a priori implausibility of systematic overdetermination.

But that is not all we need. The statements of the laws of physics record how some kind of physical particulars are related to other kinds of physical particulars in a world with mental properties in it. Suppose these mental properties are neither identical to physical properties nor realized by them. I shall take it that this means, by definition, that they are not even broadly physical properties. Instead, suppose that the instantiation of mental properties is determined by the coinstantiation of physical properties in virtue of a fundamental physico-psychological law. Both epiphenomenalists and interactionist property dualists will rely upon such laws in the formulation of their position. If instances of mental properties were efficacious and contribute towards the relations between physical particulars, the statements of the laws of physics or neuroscience would not need to record this fact. They would only need to mention the physical properties or neural properties which determine that a particular mental property is instantiated. So the statements of laws of physics may have partly non-physical mental truthmakers. Of course, I do not mean that the behaviour of atom colliding against atom in the void is partly determined by the presence of non-physical mental properties. Instead, we should focus on the complex behaviour of microphysical properties in the brain. Maybe there the microphysical behaves in the way that it does because of the mental.

What are the grounds for resisting this possibility? Those who contemplate it seem to pass over it rather quickly. For instance, David Chalmers writes that 'interactionist dualism requires that physics will turn out to have gaps that can be filled by the action of a non-physical mind. Current evidence suggests that this is unlikely' (Chalmers 1996: 163). It is not clear what evidence for a gap would be as far as Chalmers is concerned. It is not as if things would judder to a halt in the brain, nothing physical show up and then suddenly things start moving again because of the influence of what we take to be non-physical phenomenal properties. Things need not be like that at all. By Chalmers' own

lights, the physical has ensured that the phenomenal properties would be there at the appropriate time via physico-psychological laws. Things would move on smoothly as a result of the causal influence of the phenomenal properties ushered on the scene.

Instead, evidence of a gap would seem to rest on this. Physicists in their study of physical phenomena outside the brain have found reason to believe that there are no macro-surprises in their interaction inside the brain. The way in which physical particulars inside the brain relate to each other is the way that we would expect them to relate to each other if we identified putative laws of physics just by focus upon extra-cranial interaction. Moreover, there is no reason to question whether the expected might be an illusion ensured by the supportive causal role of mental properties.

I question whether we are indeed in such a position. I suspect that the models that we form – along with the values that we attach to coefficients and the like – make assumptions about what it is proper to consider as exerting some influence. We wilfully make ourselves blinder to the activity of the non-physical mental, if such activity there be, in developing our theories. I do not need to insist upon this though. All I need in the present circumstances is a more modest point against the epiphenomenalist. Epiphenomenalists are ill-equipped to rule out the kind of story I have just mentioned. They allow that mental properties are pretty clearly non-physical. It is not that they raise a doubt about the existence of these properties or suggest that there is no reason to doubt that these properties are broadly physical. They are already committed to there being something spooky going on in the brain. In particular, they must allow that physical properties give rise to macro-surprises. Nothing about the laws of physics suggests that there should be non-physical mental properties when you gather the physical together in brain-like structures. Suppose our interactionist says 'Look, when those physical properties are bunched together they behave in the way they do partly because of the presence of non-physical mental properties.' The epiphenomenalist is then in no position to say 'Ah, but we have no reason to believe that this holds. Think about how things interact extra-cranially'. Such a response would rest upon an unmotivated asymmetry. They are prepared to assert that physical properties give rise to macro-surprises – non-physical phenomenal properties say – but not that when co-instantiated in brain-like structures they would behave in surprising ways if the non-physical phenomenal properties had not been present. Why are some surprises worse than others?

Epiphenomenalists may respond that they are forced to allow for the existence of non-physical mental properties but they need not be forced to allow for any other kind of surprise. That is the difference. But, since they allow their experience of their own mental lives to have the status it has in determining their beliefs about the structure of the world, they can be pressed further. Isn't it equally apparent from our experience of our own mental lives that instances of these allegedly non-physical mental properties are the cause of our behaviour? This seems to be a fundamental part of our experience. In

which case, just as the epiphenomenalists take their experience of their own mental lives to give them reason to suppose that there are macro-surprises in the brain, so we would expect them to take it to revise their views about the causal interactions going on in the brain. This throws into doubt the claim that there is no reason to suppose that the mental properties might be playing a supportive causal role. It is not like we are supposing that there are pixies present to give the physical a hand (as it were). We are supposing that instances of mental properties – things we are in daily contact with by their lights – have a role to play.

Epiphenomenalists may seek to challenge the claim that we have experience of the efficacy of phenomenal properties in our daily mental lives. All we have, it may be charged, is our experience of a constant conjunction of certain phenomenal properties with succeeding mental states or behaviour. We infer a causal connection from that (Chalmers 1996: 159). Now it may be that our experience of the efficacy of phenomenal properties is mistaken. I do not want to rule that out. Nevertheless, it strikes me as a misrepresentation of our experience to say that we do not experience the efficacy (see Mellor 1995: 3, 107–8, for discussion of our experience of efficacy more generally). As I am holding on to the hot plate seeing how long before I am forced to drop it, I do not observe my subsequent dropping as merely an instance of a constant conjunction of burning pain and hot-plate dropping. The burning pain became too much for me and I was forced to drop it. Look behind this experience by all means but, at the same time, adopt a more hearty scepticism about one's experience of phenomenal properties as non-physical. I do not think that one should be credulous about one and frankly dismissive of the other.

Given that epiphenomenalism is in a weak position to rule out the possibility that the causal influence of non-physical mental properties is part of the truthmakers of physical laws, the plausibility of epiphenomenalism does not rest upon the empirical case for the closure principle. Its very character undermines some of the moves needed to bolster that case. As far as the considerations in favour of the closure principle go, it represents an implausible middle position between physicalism and interactionist dualism. It is possible that a justification for epiphenomenalism will stem from some independent insight into the nature of non-physical mental properties, but it is worth noting that this would represent a considerable departure from what has motivated epiphenomenalists hitherto. They have emphasized the empirical considerations in favour of the closure principle, the very ones that seem inadequate. It has usually been thought that, if these considerations were not in play, one could be an interactionist dualist. Of course, matters are different in the case of physicalist epiphenomenalisms. Worries about mental efficacy have stemmed from the kind of physical properties mental properties are thought to be. However, such epiphenomenalisms are not the focus of my discussion.

I think that epiphenomenalists must supplement the empirical considerations mentioned so far with the additional thought that it would be more

unattractive to allow that non-physical mental particulars cause physical particulars than it would be to allow that physical particulars cause non-physical mental particulars. They must, in effect, justify their endorsement of the partial closure principle that nothing non-physical causes something physical together with their corresponding rejection of the sister principle that nothing physical causes something non-physical. Why is the second partial closure principle deemed less attractive than the first?

The commitment of epiphenomenalists to this asymmetry means that they cannot rely upon the thought that there is something unintelligible per se about the interaction between physical and non-physical. Readers of David Hume will be justly suspicious of any such appeal to this idea (Hume 1978: 246–51). But even if causal connections must be intelligible, presumably the unintelligibility would hold both ways. So there is no particular gain in limiting the causal interaction to that of physical upon mental.

The answer to our question must stem from some feature of causal asymmetry, if it lies anywhere. When we understand the nature of causal asymmetry, we will see, so the thought presumably runs, why it is so much worse for non-physical mental particulars to cause physical particulars than vice versa. If this is right, the epiphenomenalist is committed to providing an account of causal asymmetry that makes the following claim plausible.

(A)   To deny that the physical is causally autonomous is more metaphysically unattractive than to deny that mental particulars cause physical particulars.

I shall call this the claim about relative metaphysical unattractiveness. It is quite reasonable to be puzzled about the whole idea of what is metaphysically attractive. Some have felt that monism per se is more attractive than any form of dualism. Part of their commitment seems to have rested on some sense of what is metaphysically attractive. Thus Smart writes of dualistic theories:

> sensations would be 'nomological danglers', to use Feigl's expression (Feigl 1958: 428). It is not often realized how odd would be the laws whereby these nomological danglers would dangle … . I cannot believe that ultimate laws of nature could relate simple constituents to configurations consisting of perhaps billions of neurons … Such ultimate laws would be like nothing so far known in science. They would have a queer 'smell' to them.
>
> (Smart 1987: 190)

Epiphenomenalists are prepared to allow the mental to dangle. So they are not quite as fastidious as Smart. But they draw the line at systematic overdetermination and the idea that the physical lacks causal autonomy. It seems that they feel that the kind of interactionist dualist position I set out earlier would undermine the status of the physical sciences, although obviously not the letter.

My objection to epiphenomenalism now comes down to this. There is no reductive account of causal asymmetry which provides a justification for the claim about relative metaphysical attractiveness. On the other hand, if we adopt a non-reductive approach to causal asymmetry, we are committed to a causal theory of temporal precedence. This presents the epiphenomenalist with further difficulties. I shall attempt to establish these points in what follows.

2   Causal asymmetry and time

Theories of causal asymmetry do not develop in a vacuum. In particular, as Mellor points out, they must explain why

(C)   Causes are usually temporally prior to their effects.

(Mellor 1995: 219)

We might interpret this claim in two ways. On one interpretation, it is metaphysically necessary that causes are usually temporally prior to their effects. On the other, it is a nomological truth. Mellor seems committed to holding that it is metaphysically necessary. He writes

The question for us now is how [we] can explain the correlation between causal and temporal order. The answer is that we cannot unless we take the latter to be entailed by the former. For if we take these two orders to be independent, it should be as conceivable that effects generally precede their causes as that causes generally precede their effects. Yet not even those who think that some causes may be later than their effects think that all or even most could be.

(Mellor 1998: 107)

For the moment, I am not concerned with the causal theory of precedence these remarks endorse. Its time will come. The point for now is that the requirement of entailment between causal order and temporal order and the reference to it not being conceivable that causes generally succeed their effects commits Mellor to the claim that (C) is metaphysically necessary (at least).

We thus have two constraints – (A), the claim about relative metaphysical attractiveness, and (C), the claim about the habitual priority of causes – within which the epiphenomenalist must work. It is not easy to provide a theory which satisfies them. For instance, consider the claim that causes are, by definition, just those events which are temporally prior in a causal connection. This might explain why it is metaphysically necessary that causes are usually temporally prior to their effects. However, we have no explanation of why it is unattractive to deny the causal autonomy of the physical world. If it is just a matter of time difference, things should be on a par. This picks up on a point made by Mellor that the simple temporal priority account of causes cannot capture the explanatory connotations of causation (Mellor 1995: 107, 219–20).

The idea that causes are fixed at times when effects are not is scarcely better placed. For one thing, if determinism is true, then causes and effects cannot be distinguished in this way. For another, there is unclarity about the role of time in such an account. If the claim is just that the time at which an effect is fixed is a subset of the times at which a cause is fixed, we have no account of why causes usually precede effects. If, on the other hand, the claim is that causes are just those which are fixed earlier, then the account seems to collapse into the previous one. It is not the notion of fixity that is doing the work but the timing (Mellor 1991: 198–9). If both ideas are in play, then we still need an explanation of why they are coextensive. This is the very problem we are trying to resolve. So there seems no progress to be made here.

The application of these two constraints also throws into question a range of theories that provide a reductive account of causal asymmetry in terms of certain macro-phenomena which, in turn, may be related to some temporal asymmetry. There are many theories of this type. Let me just mention two. The point I wish to make should generalize.

David Lewis has argued that causal asymmetry rests upon an asymmetry of overdetermination (Lewis 1986a: 49–51). Suppose that a stone drops into the water and a wave propagates outwards. The propagation of the wave and the stone dropping are causally connected. Consider the counterfactuals:

  (c1)   If the stone had not dropped, a segment of the wave front, S, would not be propagating outwards.

  (c2)   If the segment of the wave front, S, were not propagating outwards, the stone would not have been dropped.

Lewis claims that (c1) is true and (c2) false. The distinct verdicts reflect the relative importance of the two key aspects of the similarity weighting of possible worlds for counterfactuals: avoiding widespread departures of law and maximizing perfect match. In order for (c1) to be false, the segment would have to propagate outwards even if the stone had not dropped. However, we would only achieve perfect future match if the stone's failure to drop had all of its consequences covered up by a barrage of miracles in addition to the miracle required for the segment to still propagate outwards. This would be in violation of the most important standard of similarity: that there are no big, widespread, diverse departures of law from our world (Lewis 1986a: 47). In order for (c2) to be true, only small-scale departures of law would be needed to make the stone drop even though the segment S was not propagating outwards. The failure of the segment S to propagate outwards has few consequences at earlier times. By contrast, one could maximize the area of perfect match if we supposed that the stone still dropped. The difference that underpins these verdicts is an asymmetry of overdetermination. If we counted the segment of the wave as the cause and the stone as the effect then the dropping of the stone would be multiply overdetermined by all the various wave segments. Each part of the

wavefront would be sufficient for there to be a stone dropping. By contrast, if we counted the stone dropping as the cause, there is no overdetermination of the segment of the wave's propagation. Lewis's suggestion is that, in effect, the correct attribution of causes and effects minimizes overdetermination.

Others have suggested that the difference between cause and effect can be understood in terms of the probabilities distinctive of a conjunctive fork (Reichenbach 1956: 163). Let A, B, X be propositions to the effect that events a, b or x occurred respectively; P(Y/Z) stands for the probability that Y given Z. The relations distinctive of a fork are:

(1) $P(A \& B/X) = P(A/X) \times P(B/X)$.
(2) $P(A \& B/\!-\!X) = P(A/\!-\!X) \times P(B/\!-\!X)$.
(3) $P(A/X) > P(A/\!-\!X)$.
(4) $P(B/X) > P(B/\!-\!X)$.

(Reichenbach 1956: 159)

Relations (1) and (2) assert that the probability of a and the probability of b occurring are independent given the occurrence or non-occurrence of x. Relations (3) and (4) assert that the probability of a occurring and the probability of b occurring are greater given X than they would be given not-X. (1) to (4) entail (5):

(5) $P(A \& B) > P(A) \times P(B)$

(Reichenbach 1956: 160–1; Salmon 1984: 159–60)

So x's occurrence screens off the probabilistic dependence of a's occurrence on b's occurrence. Relations (1) to (4) may hold for many events whose occurrence is mentioned by propositions taking the place of X. Not all of them are plausibly thought of as causes. Those who favour this particular approach to causal asymmetry usually proceed in one of two ways. Either they appeal to an independent account of causal processes or they appeal to further patterns of probabilities and events related to the various candidate causes and which serve to distinguish genuine causes from other cases (first option, Salmon 1984: 168; second option, Papineau 1985; 1993). Either way, they understand causal asymmetry in terms of macro-phenomena of the kind described by the probabilities.

Theories of the sort just described are susceptible of two readings. On the first reading, causes are just defined by reference to these features, which means that in circumstances where these features are not present, there are no causes. On the second reading, causes *explain* the overdetermination asymmetry or the probability relations constitutive of the fork *among other things*. By pointing out their explanatory role in this area, we are just picking out one important dimension of a richer notion. The proponents of the theories I have mentioned adopt the first interpretation.

The proper interpretation matters for our present purposes. If the first

reading is adopted, it is unclear why it should be of particular importance to deny that mental particulars cause physical particulars but allow that physical particulars cause mental particulars. The apparent causal autonomy of the physical obtained by this insistence seems of insufficient significance to justify the claim about relative metaphysical attractiveness. Causal asymmetry is a macroproperty which does not reflect any local asymmetry of necessitation. This is quite explicit in Lewis's work since he is loath to recognize intraworld necessitation in general. The whole point of his defence of Humean supervenience is to deny that there is any such thing (Lewis 1986b: ix–x).

It is perhaps no surprise, therefore, to find that those drawn to epiphenomenalism are inclined to reject the kind of theory of causation put forward by Lewis and others. Thus, according to Chalmers, 'a causal connection between two events is something over and above a regularity between the two events … There is something irreducible in the existence of laws and causation' (Chalmers 1996: 86). And, when he turns to consider whether one can avoid the arguments in favour of epiphenomenalism by adopting a theory of causation giving centre stage to regularities, laws and counterfactuals, he writes:

> I find … these positions implausible. I have argued against Humean views
> of causation … and even on the non-Humean view it is implausible that
> just any nomic connection suffices for causation.
>
> (Chalmers 1996: 151)

It is true that in the textual material surrounding the second quotation Chalmers does not consider the details of the theories described above. They cannot be characterized as supposing that any nomic connection in Chalmers' sense is sufficient for causation. However, taking the two passages together, I think Chalmers' attitude to the kind of theories which characterize causal asymmetry in one of the ways set out above is clear. Each is part of a reductive analysis of causation which he eschews.

Of course, a sophisticated epiphenomenalist's position need not coincide with Chalmers'. Nevertheless, it is questionable whether one of the differences could lie here. What is the merit in insisting that non-physical mental particulars are not causes if the designation of something as a cause is just determined by its particular place in a set of probabilistic relations or in the minimization of overdetermination? Leastways, the challenge for the epiphenomenalist is to say what the merit is.

The force of the challenge is apt to be underestimated if the first reading of the theories of causal asymmetry set out above is not properly distinguished from the second. According to the latter, causes *explain* the macro-phenomena cited. A consequence of this second reading is that the causes may occur outside the context envisaged. Suppose that there are just two particles in the universe; one strikes the other, which goes off. Then there is neither the overdetermination asymmetry nor a causal fork. Yet, according to this second reading, it still might be the case that the striking particle is a cause and the struck an effect. Nor would the asymmetry threaten to run out at the level of

microphysics in our world. If causal asymmetry is understood in these terms, I can see that there might be a reason for thinking that epiphenomenalism preserved a particular, and significant, kind of causal autonomy for the physical. However, we still need an explanation of how this notion of causal asymmetry can underwrite (C), the temporal claim about causes. It is at this point that we need to turn to causal theories of temporal precedence.

3   Epiphenomenalism and causal theories of temporal precedence

To develop a defensible causal theory of temporal precedence, it is natural to start with the idea that

(1)   If e_1 causes e_2, then e_1 is earlier than e_2 (where e_1 and e_2 are events).

If we appeal only to actual causal relations, though, we fail to generate the pervasive relations of temporal precedence that we require. What about events which are not causally related? How are they related in time? For that reason, we might move from actual to possible causal relations giving us

(2)   If it is possible for e_1 to cause e_2, then e_1 is earlier than e_2.

The question is: How we are to account for the possibility mentioned? Causal theorists cannot claim that it is possible for e_1 to cause e_2 because e_1 is earlier than e_2. They want the explanation the other way round.

To deal with these difficulties, Mellor proposes the following:

(T)   t precedes t′ if there is some fact C at t which causes some fact E at t′.

(Mellor 1998: 113)

He writes:

> For each point of spacetime is the location of many facts e.g. about density, curvature, pressure, temperature, the intensity of gravitational, electromagnetic and other fields, etc., all of them related causally to some other facts at other parts. So all we need, for causation to fix the order of any two spacetime points, and hence of t and t′, is – in this case – that *some* fact C at t causes some fact E at t′, thereby making all other facts at t also precede all other facts at t′, whether they cause those later facts or not.
>
> (Mellor 1998: 113)

Mellor's proposal is a natural way to capture the intent behind the formulation given in modal terms. (T) is meant to be a metaphysically necessary truth. If his theory were defensible, then we could couple taking causal asymmetry

as primitive with accounting for the intuition that most causes temporally precede their effects.

I have strong reservations about whether the epiphenomenalist will be able to appeal to a theory of the kind Mellor recommends. As a result, the epiphenomenalist will naturally look for alternatives. However, I think that these are less plausible and, in particular, less obviously available to someone motivated by the concerns that formed the basis of epiphenomenalism than a version of Mellor's theory.

One consequence of holding that (T) is metaphysically necessary is that we must deny that causation can be simultaneous. This might seem a small price to pay. I am prepared to believe that there is no simultaneous causation in our world for the reasons that Mellor identifies. Let me mention a selection. First, perfectly rigid objects are not physically possible (Mellor 1998: 110). So there can be no simultaneous transmission of movement from one end of a perfectly rigid rod to another. Second, if simultaneous causation across any distance were possible, this would be hard to square with the special theory of relativity (Mellor 1998: 108). Such simultaneous causation would involve infinite speeds of transmission. The acceleration of a particle through the speed-of-light barrier would involve the mass of a particle becoming infinitely large. So the only way that such causation would be possible is if it involved entities which uniformly travelled faster than the speed of light, for instance tachyons. The evidence for their existence is limited. Third, in order to affect the electrostatic field at all points absolutely later than st, a point particle with electrical charge E at spacetime point st will not have to simultaneously affect the electrostatic field precisely where it is to avoid unmediated action at a distance. Spacetime is dense. That means that for any spacetime point st' absolutely later than st, there will always be an intermediate point between st and st' to mediate the interaction (Mellor 1998: 110–11).

All of these points strike me as plausible. However, as Mellor acknowledges, although they may establish that simultaneous causation is physically impossible, they do not establish that it is metaphysically impossible. Couldn't there be possible worlds in which spacetime is not dense? In which perfectly rigid objects move at uniform velocity? Or, indeed, where the special theory of relativity is false?[1] I think that it is possible to provide an analysis of causation without writing in that causes temporally precede their effects. Obviously I cannot defend such an analysis here (Noordhof 1999b). But if I am right, then in the absence of clear reasons for rejecting these putative metaphysical possibilities, the best we can say is that simultaneous causation is nomologically impossible.

I think Mellor seeks to establish that simultaneous causation is metaphysically impossible by emphasizing the following points:[2]

(I)   For any two facts to coincide is for them to be able to interact immediately.

(Mellor 1995: 224)

(II) Loops of causability are not possible (i.e. loops in which, although there need not be actual causation between any two successive facts in the loop, there is the possibility of causation).

(Mellor 1995: 229–30)

(III) If P is the same kind of fact as Q, P causes Q and P is spatio-temporally coincident with Q, then P is Q.

(Mellor 1995: 234)

The argument would then proceed as follows. If it is possible that there are two facts, P and Q, of different kinds, such that P and Q are spatio-temporally coincident and yet, while P interacts with Q, Q does not interact with P, then it is possible that this is the case when P and Q are of the same kind (Mellor 1995: 234). From (III), the consequent is false. In which case, the antecedent is false. Hence (I): two distinct facts cannot be spatio-temporally coincident without loops of causability. Loops of causability are not possible. Since, simultaneous causation is possible only if two facts are spatio-temporally coincident, simultaneous causation is not possible.

Let me express a preliminary worry about this argument before we get to the heart of the matter. The argument partly rests on (III), and (III) is not an obvious consequence of Mellor's identity criterion for facts. Indeed, Mellor's identity criterion for facts seems in tension with (I) and (II) taken together. Mellor suggests that, 'for any two facts D and D′, D = D′ if D and D′ have all the same causes and effects' (Mellor 1995: 113). He later strengthens this to:

(F)   D = D′ iff D and D′ have all the same causes and effects.

(Mellor 1998: 104)

The difference seems to be that he does not endorse, in his later work, the claim that D = D′ iff D and D′ occupy the same region of spacetime.[3]

This seems to be right. To give an example he uses, though not for quite this purpose, the fact that Jim is the shortest man and the fact that Jim is the fittest man could occupy the same region of spacetime yet they are not the same fact because they can have different causal consequences. But if (F) gives us reason to doubt the combination of (I) and (II) and does not imply (III) it is not clear from whence (III) derives its support. The most obvious answer would have been (I) but that is undermined by the combination of (II) and (F).

My more substantial worry rests on Mellor's claim that, if simultaneous causation is possible, it will have to hold between spatio-temporally coincident facts. This brings me to (II). Here is how I see the argument run. The reason why simultaneous causation between the non-coincident is not an option is that, relative to some frames of reference, it would be construed as backward causation and hence involve loops of causability. The support for this claim is drawn from the special theory of relativity. So it is not clear to me that it is metaphysically necessary that, if simultaneous causation between the non-

coincident occurred, it would involve backward causation. That would depend upon whether the special theory of relativity is metaphysically necessary. Suppose it is for the sake of argument. The issue is whether backward causation is metaphysically impossible. It does not seem to me that it is.

I cannot prove this but I do want to raise an issue about Mellor's argument that it is metaphysically impossible. His argument rests on the claim that

> (L)   For no cause C and effect E do E's logically independent chances with and without C entail anything about the equally independent chances of C with and without its causes, or about the chances of E's effects with and without E.
>
> (Mellor 1998: 132)

It proceeds as follows. Backward causation implies a loop of causability. Suppose C-type facts and E-type facts stand in such loops. Given (L) we could assign arbitrary values for the chances which the presence or absence of a C gives an E and for the chances which the presence or absence of an E gives a C. The law of large numbers asserts that, for all p, p(X) entails f_∞ (X) = p. In other words, chances entail hypothetical limiting frequencies. Let the actual distribution of C and not-C be some very large numbers n_c and $n_{not\text{-}c}$. Then there will be an assignment of chances which will make it the case that it is almost certain that C and not-C do not have this distribution even though they do. For instance, if there are 10 million Cs and 10 million not-Cs, suppose that the chance of E if C is 0.6 and the chance of E if not-C is 0.2. This would mean that it was almost certain that there would be around 8 million Es and 12 million not-Es. Suppose that the independent probabilities of 0.5 and 0.25 are assigned to the chance of C given E and the chance of not-C given not-E respectively. That would mean that it was almost certain that there would be 7 million Cs and 12 million not-Cs. Mellor claims that this is a contradiction (I take the figures from Mellor 1998: 134–5). He concludes that, if C and not-C give independent chances of E, then E and not-E do not give chances of C. Hence backward causation is not possible (Mellor 1995: 224–9; 1998: 132–5).

Mellor's position seems to rest on an unmotivated asymmetry. In other places, he suggests that considerations of consistency should structure our understanding of the possibility of certain conjunctions of facts. For instance, he writes that 'since … laws must be compatible to coexist, they cannot impose incompatible [time] orders on the points that instantiate them' (Mellor 1995: 236–7) and that 'all laws [are] instantiated everywhere' (Mellor 1995: 215).

By Mellor's lights, what makes laws hold are facts with the following structure: Nst. N is the instantiation of a law, st a spacetime region. If Nst, then, because of the second consideration, N is instantiated everywhere. Moreover, if Nst, then there can be no N* instantiated such that N* imposes a different time order. Nor, and this of course is obvious, could there be an N' that makes a law statement true which contradicts the law statement that N* makes true. Consider our putative loop. Suppose there is an N_c embodied

in spacetime which makes $(x)(Cx \rightarrow ch(E) = 0.6)$ and an N_{not-c} $(x)(not\text{-}Cx \rightarrow ch(E) = 0.2)$ true. My question is why cannot we then conclude, in the light of the reasoning above, that, since arbitrary assignment of chances would give rise to a contradiction, there can be no combinations of N_e and N_{not-e} embodied in spacetime which, in our very simplified case, fails to make it almost certain that there are 10 million Cs and 10 million not-Cs. One would have thought that it is just as true of our conception of laws that they should not hand out inconsistent probabilities in this kind of case as that they should meet the other constraints (see Noordhof 1998: 873–4, where I originally made this objection). This is not just the point that one person's modus ponens is another person's modus tollens. It is that we need a motivation for the differing attitudes taken in these cases.

The upshot of the discussion is that we still have no reason to deny that simultaneous causation is metaphysically possible. If simultaneous causation is metaphysically possible, then Mellor's proposal (T) is, at best, a nomologically necessary truth. Causal facts do not entail the direction of time. In which case, he cannot explain how (C), the claim that causes usually precede their effects, is a metaphysically necessary truth.

It might be tempting to claim that the fact that (C) is a metaphysically necessary truth plus the independent plausibility of (T) being some kind of necessary truth gives us reason to suppose that (T) is a metaphysically necessary truth. The claims come as a package deal and one kind of package can be more attractive than another even though elements of the more attractive package may lack demonstrative support (for hints on this line of argument, see Mellor 1998: 111).

The problem is that I do not think that a proponent of Mellor's causal theory of temporal precedence is forced to assert the metaphysical necessity of (T) in order to explain (C). There are other elements of Mellor's position that present a plausible alternative. He places agency central to causality. He writes

> Causation's means–end connotation is even more basic than its evidential and explanatory connotations, being to my mind the very core of the concept: causation is essentially the feature of the world that gives ends means. But essential or not, the fact is undeniable: causation is in fact what gives ends means.
>
> (Mellor 1995: 79–80)

I propose that we take Mellor at his strongest on this issue:

(M) Metaphysically necessarily, causes are means for agents to bring about ends.

It seems that we can provide an argument in favour of (C), the claim that causes usually precede their effects, by focusing on agents and two important related features of experiences of the past. The first is that the past is that

portion of spacetime containing facts which are the preponderant causal source of our experiences and recollections. This is not a deep truth. It is just an observation about how we settle what we call 'the past'. The second is that the past is that portion of spacetime which our experiences and recollections present as relatively fixed when compared with the future. I put it in these terms to allow for the possibility of clairvoyance. The thought is that if clairvoyance accounted for the majority of our experiences it would no longer be clairvoyance but instead determine that 'the future' was in fact 'the past'.

Agents conceive of means to ends as ways of bringing about something which they fail to experience as fixed. If agency is central to our understanding of causality, this way of conceiving of means to ends correctly captures an important feature of causes. If causes are means to ends, then an appropriate instance could always be pressed into service to bring about something which is not experienced as fixed. Since the past is actually the preponderant causal source of our experiences and recollections, causes will, in general, be future directed. Of course, this does not imply that causes are always prior to ends, for instance if travelling back in time is possible. It is just that there cannot be too many such cases on pain of swapping over the past with the future (see Mellor 1981: 157).

The remarks about how we determine the past and the nature of agency have reasonably good claim to be metaphysically necessary truths. In which case, we have an explanation of (C) even though (T) is, at best, only nomically necessary. However, suppose that some of the claims about agency and our experiences of the past are not metaphysically necessary. Then it seems to me that we have a very good explanation of why we should be under the mistaken apprehension that (C) is metaphysically necessarily true. Either way, we do not have good support for the metaphysical necessity of (T) from the metaphysical necessity of (C).

With these adjustments to Mellor's theory in place, we are now in the position to assess whether epiphenomenalists are in a position to appeal to such a theory. The problem is that if (T), the causal theory of precedence, is only nomically necessary then the epiphenomenalist seems forced to hold that there are laws which determine that certain facts should hold in the physical world *because* certain facts hold in the non-physical mental world. The epiphenomenalist claims that some physical facts cause non-physical mental facts. If (T) is true, then the non-physical mental fact occurs after the physical fact. The issue is this: How can the mental fact precede any other fact? Since the non-physical mental fact does not cause anything, there is no subsequent fact which, by being the effect of the non-physical mental fact, must be posterior to the mental fact. Yet manifestly mental facts do precede other physical facts and mental facts.

To get round this difficulty, I presume that epiphenomenalists would probably claim that there will be physical facts that occur at the same time as mental facts and these physical facts are part of the causal network. Given (T) is nomologically necessary, it is a law of nature that physical facts are present

at the same time as mental facts. We should consider what this implies about the physical world. There are three alternatives.

First, epiphenomenalists could argue that, in the case of physico-psychological causation, the cause is simultaneous with the effect. Then there would be a fact occurring at the same time as the non-physical mental fact. They would claim that (T) held for all physical facts and for the timing of mental facts relative to the causal order of physical facts. However, they would have to deny that causing of a mental fact by the physical facts was governed by (T). They would argue that the proper relationship between causal precedence and temporal precedence needs a more complex statement. I do not propose to seek to try to do this, because I hope the problem with this option is clear. Epiphenomenalists are now not in the business of merely postulating special physico-psychological laws while retaining the causal autonomy of the physical. They are suggesting that a different nomological relationship holds between causal precedence and temporal precedence purely because of the existence of non-physical mental facts. But the nature of the relationship between causal precedence and temporal precedence is as much the business of the physical sciences as laws between mass and energy. So the autonomy of the physical is undermined. It is hard to see how epiphenomenalists could motivate denying efficacy to the mental on the grounds of the causal autonomy of the physical given that they are prepared to say that certain laws which are the subject matter of the physical sciences take the form they do because of non-physical mental facts. Moreover, the laws in question will have implications for the causal relations which hold in the physical world, since the question of whether something precedes something else can determine whether or not it influences it. Each physical interaction will be tainted by a law whose form is determined by the existence of non-physical mental facts.

The other two options rest on the claim that there will always be other physical facts, posterior to the causes of mental facts, which occur at the same time as the mental facts. These serve to put the mental facts in time. There need be no revision of (T) as a result. One way of developing this suggestion is to claim that, although there will always be physical facts at the same time as non-physical mental facts, there need be no physical cause of these physical facts. They just occur without explanation. This would be baffling and quite unacceptable to those motivated by the kinds of considerations that led up to an endorsement of the causal closure of the physical realm. The presence of gaps of this kind would be precisely the sort of evidence which physicalists conceded would show the efficacy of the non-physical mental. Epiphenomenalists would be in the absurd position of preferring to deny the existence of any causal explanation rather than allow that non-physical mental facts are causes.

The alternative is that the required physical facts are there to put mental facts in time due to physical laws and physical causes. In which case, we once more seem faced with a situation in which there is a conspiracy at work. Physical laws partly take the form they do so that, given the initial conditions of the universe, it is guaranteed that there will be physical facts to put the

mental facts in time. It strikes me that this is even more peculiar than the situation envisaged by the interactionist. Epiphenomenalists might be able to preserve the causal autonomy of the physical world by stipulating that the non-physical mental is inefficacious but at the expense of undermining the status of physical laws themselves. Their structure will be partly explained by the nature of non-physical mental facts and a causal theory of temporal precedence. In fact, matters threaten to be a little worse than that. Although there might be indeterminism in the universe, there is one place where the strictest determinism must reign according to the epiphenomenalist. Whenever there is a non-physical mental fact, there will be a physical fact at the same time. This will be so even if the physical laws themselves are indeterministic. It is hard not to think that this would be evidence of a hidden variable: efficacious mental facts. But if that is the case, we would not need to appeal to the idea that there are physical facts holding at the same time as mental facts. The efficacy of the mental facts would place them in time.

One way out of the problems just sketched would be to suppose that causal relations hold between points in spacetime quite independently of whether there were mental or physical facts at these points. This conflicts with Mellor's own view of the matter (Mellor 1995: 236). He denies that spacetime points can exist independently of the facts that are located at them. I do not think we need to decide that Mellor is right about this to look askance at epiphenomenalists. Are they really asking us to adopt a particular view about the nature of spacetime in order to allow for the viability of their position? This would hardly preserve the autonomy of the physical sciences. Of course, if there were some independent motivation to adopt this position, that would be different. In the absence of that, epiphenomenalists cannot pretend that their view respects what interactionist dualists fail to respect: the autonomy of the physical realm.

Although the problems I have sketched appealed to the nomological reading of the causal theory of temporal precedence, it does not seem that they would be avoided if it was a metaphysically necessary truth after all. The points I made regarding the last two options would still hold. It is only with regard to the first that some respite could be offered. If the connection between causal precedence and temporal precedence were a matter of metaphysical necessity, then it would not be true that one of the laws of nature took the form it did because of the presence of mental facts. The connection is not a matter of law. Instead, the worry is that there is little justification for this particular claim about the nature of temporal precedence apart from its accommodation of epiphenomenalism. Rather than explore this option further I will turn to a final way in which epiphenomenalists may seek to avoid force of the discussion so far. It does not seek to tinker with Mellor's causal theory of temporal precedence. It abandons it altogether for something rather different.

Instead, the idea is to explain temporal precedence in terms of de facto irreversible processes. One development of the idea is to claim that the direction of time is the direction of the majority of open causal forks. It is

compatible with this approach to characterize the forks purely in terms of the probabilistic relations without any appeal to an objective asymmetry of necessitation. Causes could be temporally prior to effects by definition. That is Paul Horwich's (1987: 132–45) line. However, I hope it is clear that Horwich's proposal would not be acceptable for epiphenomenalists. They need to explain why (C) is metaphysically necessary when causation is understood to involve asymmetric necessitation. So they must view causes as the kind of things that asymmetrically necessitate the common effects constitutive of the fork.

The problem with theories of this type is that they do not fit with our experience of temporal precedence. Mellor's theory fits our experience far better. When we experience one of our experiences as the cause of another of our experiences, we experience the first as temporally prior to the second. In experience, temporal priority and causal priority seem to go together. On reflection, causal priority and temporal priority might not fit so closely together as first thought, but the naturalness of the relationship is striking (see Mellor 1998: 114–15).

Here are two illustrations of the point. First, a theory of temporal precedence based on the fork asymmetry is forced to arrive at counterintuitive verdicts about certain simple worlds. For instance, if two neutrons revolve around each other in a universe and nothing else takes place, there would be no temporal precedence according to a theory based on the fork asymmetry (Tooley 1987: 226–8). By contrast, Mellor could allow that there is temporal precedence.

Second, since there are many de facto asymmetries in time, we need an explanation of why the fork asymmetry is special. Why should it outweigh the considerations in favour of an account of temporal precedence like Mellor's when many clearly do not? Mellor has provided a nice illustration of the worry by considering another de facto asymmetry, the movement of clock hands. Suppose that C-clocks are those in which the long hand moves from 1 to 2. Since almost all clocks have this feature, they are de facto irreversible processes. As far as the laws of nature are concerned, though, we have symmetry. It is consistent with the laws of nature that the reverse should happen. If we observed C's hands travel from 2 to 1 then we would suppose that a local reversal of time order had occurred. The situation would be different if we observed another type of clock, a C*-clock, designed so that its hands are made to move from 2 to 1. We would have reversal of the same type of events as those occurring in C but no time reversal.

What is the basis of the difference? In the case of C, we run the causal sequence backwards. This can be revealed by intervention. If we bend the long hand of a clock of type C at 1, it will still be bent at 2. In the time-reversed case, C's hand travels from 2 to 1, it will be bent at 2 and unbent at 1. By contrast, if a clock of type C*'s hand were bent at 1, it would not be bent after 1 up to 2 and beyond. Here the causal sequence is not running backwards. Instead, we just have a reversal of the movement of the hand because the clock is of a different type. This suggests that it is not so much the de facto asymmetry

which matters. Reverse the asymmetry and we get no time reversal. What matters is the reversal of the causal order just as we would expect if Mellor's approach were right (Mellor 1998: 120–1).

Of course, pointing out that the reversal of one kind of de facto asymmetry does not yield a reversal of time order does not imply that the reversal of any type of de facto asymmetry will be irrelevant. Nevertheless, the burden of proof is on the proponents of a particular de facto asymmetry to explain why the considerations Mellor adduces will not apply to theirs too. This is a challenge that it looks as if epiphenomenalists must take up. As things stand, they have no other means of providing a justification of their position.

In particular, epiphenomenalists will have to explain how, although a particular de facto asymmetry does not seem to be the obvious basis of time, in fact it is. The motivations which drive their position make this difficult. Epiphenomenalists emphasize the importance of taking our mental lives at face value. The difficulty of explaining how it could be constituted from the physical is taken to provide strong evidence that it is not. Fair enough, but they need to explain why the same scruples will not force us to conclude that temporal precedence is not constituted from the fork asymmetry and the like. As we have seen, temporal precedence is not just a relation in the world but a fact of our mental lives. If we are to take the phenomenology of these lives sufficiently seriously to reject physicalism and contemplate epiphenomenalism, then there had better be a good explanation of why we should not take our experience of the temporal precedence of our experiences sufficiently seriously to reject theories of temporal precedence based upon de facto irreversible processes such as the fork asymmetry.

Acknowledgement

Notes

1　When Mellor was committed to the laws of nature being metaphysically necessary (back in the days of 'In Defence of Dispositions') he might have had an answer to this concern (Mellor 1974). Now he is prepared to allow that the same properties may be instantiated even though the laws of physics are different (Mellor 1995: 172). As far as I can see, he has no reason to rule out the possibility that temporal priority may exist in a non-Einsteinian world.

2　I put forward this interpretation with some hesitation as I am not sure I understand Mellor here. His argument appears capable of establishing that there cannot be many facts located at a spacetime point. But this would be in conflict with the background of his theory of temporal precedence, although not the letter. Since the argument also occurs in the paperback edition of *The Facts of Causation*, where he refers to *Real Time II*, I assume that Mellor still endorses it.

3   Again, I am hesitant about this. Mellor published the paperback edition of *The Facts of Causation* in 1999, after *Real Time II*. Although he made some alterations, this is not one of them.

References

Broad, C. D. (1925) *The Mind and its Place in Nature*, London: Routledge and Kegan Paul.

Campbell, K. (1970) *Body and Mind*, Notre Dame, IN: University of Notre Dame Press.

Chalmers, D. (1996) *The Conscious Mind*, Oxford: Oxford University Press.

Crane, T. and Mellor, D. H. (1990) 'There is no question of physicalism', *Mind* 99: 185–206.

Feigl, H. (1958) 'The "Mental" and the "Physical"', *Minnesota Studies in the Philosophy of Science* 2: 370–497.

Horwich, P. (1987) *Asymmetries in Time*, Cambridge, MA: MIT Press.

Hume, D. (1978) [1888] *A Treatise of Human Nature*, L. A. Selby-Bigge (ed.) second edn, with text revised by P. H. Nidditch, Oxford: Oxford University Press.

Jackson, F. (1998) 'Epiphenomenal qualia', in *Mind, Method and Conditionals*, London: Routledge.

Lewis, D. (1986a) 'Counterfactual dependence and time's arrow', in *Philosophical Papers*, Vol. 2, Oxford: Oxford University Press.

—— (1986b) *Philosophical Papers*, Vol. 2, Oxford: Oxford University Press.

Mellor, D. H. (1974) 'In defence of dispositions', *The Philosophical Review* 83: 157–81.

—— (1981) *Real Time*, Cambridge, UK: Cambridge University Press.

—— (1991) 'McTaggart, fixity and coming true', in D. H. Mellor, *Matters of Metaphysics*, Cambridge, UK: Cambridge University Press.

—— (1995) *The Facts of Causation*, London: Routledge.

—— (1998) *Real Time II*, London: Routledge.

Noordhof, P. (1997) 'Making the change: the functionalist's way', *British Journal for the Philosophy of Science* 48: 233–50.

—— (1998) 'Critical notice: causation, probability, and chance, D. H. Mellor, *The Facts of Causation*', *Mind* 107: 855–77.

—— (1999a) 'Probabilistic causation, preemption and counterfactuals', *Mind* 108: 95–125.

—— (1999b) 'Causation by content?', *Mind and Language*, 14: 291–320.

Papineau, D. (1985) 'Causal asymmetry', *British Journal for the Philosophy of Science* 36: 273–89.

—— (1993) 'Can we reduce causal direction to probabilities?', in D. Hull, M. Forbes, K. Okruhlik (eds) *PSA 1992*, Vol. 2, East Lansing, MI: Philosophy of Science Association.

Reichenbach, H. (1956) *The Direction of Time*, Berkeley: The University of California Press.

Salmon, W. C. (1984) *Scientific Explanation and the Causal Structure of the World*, Princeton, NJ, Princeton University Press.

Smart, J. J. C. (1987) 'Sensations and brain processes', in J. J. C. Smart, *Essays: Metaphysical and Moral*, Oxford: Basil Blackwell.

Tooley, M. (1987) *Causation*, Oxford: Oxford University Press.

8 Is causation a genuine relation?

Peter Menzies

1 Introduction

Over a period of more than 30 years Hugh Mellor's writings have illuminated an enormous range of metaphysical issues to do with chance (Mellor 1971), dispositions, laws, properties (Mellor 1991) and time (Mellor 1981). His work has had a salutary influence in encouraging metaphysicians to think about these issues in clear-headed, realist ways.

His work on the metaphysics of causation (Mellor 1995), in particular, is distinguished by its rigour, cogency and originality. The main outlines of his theory of causation are well known. He has argued that causation relates facts primarily, with causation between events deriving from causation between facts; that causation comes in deterministic and probabilistic varieties; that each variety can be explained in terms of closest-world counterfactuals in which single-case chances play a crucial role; that the important connotations of causation are that causes precede, are contiguous with, are evidence for, explain and are means for bringing about effects; and that these connotations are consistent with, or imply, that causes increase the chances of their effects.

There is much in Mellor's theory of causation that I find congenial. Indeed, I hold many of the same views precisely because he has persuaded me of their truth. But we disagree on one issue that is central to the conceptual analysis and metaphysics of causation. The issue concerns whether causation is a genuine relation. Mellor believes that it is not, whereas I believe that it is. In this chapter I scrutinize his criticisms of the view that causation is a relation and, in passing, consider some related arguments for the same sceptical conclusion advanced by other philosophers. My conclusion that Mellor's scepticism on this matter is misplaced is not too surprising, but some of the arguments I rely on to reach this conclusion highlight a surprising and hitherto overlooked feature of the concept of causation.

2 Causation as an intrinsic relation

Before detailing Mellor's criticisms of the conception of causation as a real relation, it is worthwhile considering what reasons there might be that favour

this conception. One reason I have given is that our intuitive judgements about cases of pre-emption and overdetermination rely on the idea that causation is an intrinsic relation between a cause and its effect (Menzies 1996). If causation is an intrinsic relation, it must, a forteriori , be a relation. Allow me to rehearse briefly the reasons for thinking that it is an intrinsic relation.

Pre-emption and overdetermination examples are alike in that there are two or more processes leading to some effect. In pre-emption examples, only one of the processes goes to completion and brings about the effect, but in doing so it cuts off the other processes. In overdetermination examples, all of the processes go to completion, with no pre-emption of one process by another. Consider one familiar kind of pre-emption example.

> Case 1: Assassin A and assassin B, who are both deadly accurate marksmen, have been hired to kill a prominent political figure. They work independently of each other. But, as it happens, they come across their victim at the same time and place. Both assassins take careful aim, their fingers poised to pull their triggers. But assassin A fires first, his bullet hitting its mark. On seeing the victim collapse, assassin B refrains from pulling his trigger. However, if assassin A had not fired, assassin B would certainly have fired and hit his mark.

This kind of case appears to pose a difficulty for any theory such as Mellor's that employs a counterfactual increase-in-chance condition as a criterion for causation. Introducing a notion of counterfactual dependence at this point will make our discussion more precise.

> (1)   e *counterfactually depends* on c if and only if e's chance of occurring if c had not occurred would have been less than its actual chance given that c did occur.

As Mellor's theory analyses causation in terms of what I am calling counterfactual dependence, it encounters difficulties with pre-emption examples like Case 1. For it cannot explain our intuitive judgement that assassin A's action caused the victim's death, since A's action did not increase the chance of the victim's death. For if assassin A had not fired, the chance of the victim's dying would have been the same as its actual chance – that is, fairly close to 100 per cent – given the presence of the very reliable assassin B waiting in the wings.

Mellor has not discussed this kind of pre-emption example in his published work, as far as I know. I suspect that the reason for this is that he believes that his theory has the resources to be able to deal with it.[1] He argues that there is no unmediated action at a distance (Mellor 1995: 229–34). In other words, a cause and effect must be linked by a chain of intermediate contiguous causes and effects. In this case it seems that there is a ready solution to the pre-emption problem. There appears to be a chain of contiguous causes and

effects running from assassin A's actions to the victim's death, but no such chain running from assassin B's action. This seems to vindicate our intuitive judgements about the example.

However, this appearance of a ready solution dissolves on closer examination. It turns out that this kind of pre-emption example, which David Lewis (1986a) has described as *late pre-emption*, is more intractable than first appears.[2] The special problem posed by late pre-emption examples is that they make it hard to establish the existence of a chain of contiguous causes and effects running from the main pre-empting cause to the effect. Consider, for example, the chain of events running from assassin A's pulling the trigger and the victim's death. What is the last link in the chain of contiguous causes and effects? Could A's bullet travelling in mid-air towards the victim's body be the immediate cause of the victim's death? It would appear not, because this event did not increase the chance of the victim's death: even if there had been no bullet in mid-trajectory, the victim would have died anyway from a bullet fired a few seconds later by the back-up assassin B. The same kind of reasoning shows that no event contiguous with the effect satisfies the counterfactual dependence condition for being its immediate cause.

A common strategy for rescuing the counterfactual increase-in-chance condition from cases of late pre-emption is to insist on a very strict criterion of identity for the entities that serve as cause and effect. For example, if one insists that the victim's death could not occur at a different time from its actual time of occurrence, then one might be able to argue that A's bullet in mid-trajectory does in fact satisfy the counterfactual criterion: if A's bullet had not been in mid-trajectory, the victim would have died not at the time he did but a few seconds later, in which case he would have died a different death. Mellor is surprisingly reticent on the question of the criterion of identity for the facts that he takes to be linked by causation. He says in one place that Don's dying is 'his dying roughly then, there and as he does' (Mellor 1995: 14). Depending on how rough 'roughly then' is, he might appeal to this strategy of taking facts to be very fragile in order to rescue his account from the problem of late pre-emption. But there is another class of problems from which his theory cannot be rescued so easily.

This class of problems concerns overdetermination examples. For example, consider the following variant of Case 1.

> Case 2: Assassins A and B are deadly accurate marksmen, working independently of each other to kill a prominent political figure. They both come across their victim at the same time and place. Both fire bullets into their victim's heart at exactly the same moment.

Neither A's firing nor B's firing satisfies the increase-in-chance condition for being a cause of the victim's death. For if one of them had not fired, the victim would still have died, and indeed died at exactly the same time, from

the other's bullet. Once again the strategy of looking for a chain of contiguous causes and effects does not help because of the persistent problem of establishing the last link in the chain. One cannot say that either A's bullet or B's bullet in mid-trajectory is the immediate contiguous cause of the victim's death, because the absence of one or other would leave undiminished the chance of the victim's dying.

Examples of pre-emption and overdetermination such as these should, in my opinion, make us very sceptical about the prospects for an analysis of causation in terms of a counterfactual dependence condition. As an alternative to such an analysis, I proposed a functionalist theory of causation as a theoretical entity (Menzies 1996). Adopting a standard treatment of theoretical entities, I argued that the functional role of causation is given by certain crucial platitudes in the folk theory of causation. One crucial platitude is that causation is an intrinsic relation, which means, roughly, a relation determined by the intrinsic properties of its relata and of the process connecting them. In this respect, I think, the commonsense conception of causation simply conflicts with the Humean view of causation as an extrinsic relation depending on large-scale regularities.[3] Another crucial platitude of the folk theory is that the causal relation coincides for the most part with the counterfactual dependence condition, with the notable exceptions of pre-emption and overdetermination cases. Consequently, even if the counterfactual dependence condition cannot define causation, it can at least serve as a defeasible marker for the presence of the intrinsic relation that is causation. Combining these crucial platitudes, I offered the following functionalist analysis of causation:

(2)   c is *a cause* of a distinct event e if and only if the intrinsic relation that typically accompanies a counterfactual dependence between events holds between c and e.

This definition offers an a priori conceptual analysis of causation in terms of a certain counterfactually specified functional role. As with similar functionalist definitions, it can lead to a posteriori identification of the actual occupant of the functional role. Assuming that causation could be defined as an absolute relation in this way, I suggested that the intrinsic relation that occupies the functional role of causation might be the relation of exerting a force, or the relation of transfer of energy or momentum. Given some such identification, it is easy to see how the a priori analysis, combined with the a posteriori identification, leads to a uniform solution to the problems arising from pre-emption and overdetermination. For in each of the problem cases our judgement about which of the potential causes actually caused the effect tracks whether a complete process of a kind that could occupy the functional role defined above connects the potential cause with the effect.

The crucial feature of this analysis for my discussion here is the reference to causation as an intrinsic relation. In Menzies (1999) I explored several different ways in which the notion of an intrinsic relation might be explained, settling

on one explanation in terms of a robustly realist conception of universals, or perfectly natural properties and relations that carve nature at its joints. Assuming the existence of such properties and relations, I followed Lewis in defining an intrinsic relation as one instantiated by a pair of relata just in virtue of the perfectly natural properties and relations of that pair itself.[4] An intrinsic relation, so defined, supervenes on just the perfectly natural properties and relations of its relata. This definition certainly assumes a very robust conception of causation as a relation. It lays itself open, therefore, to Mellor's criticisms of this conception.

3   Mellor's critique of causation as a relation

Before detailing these criticisms, let me describe some of Mellor's background views on causation (Mellor 1995: 156–62). He argues that the canonical form of causal statements is given by 'E because C', where 'C' and 'E' state facts and 'because' is a sentential connective. Statements of this form certainly appear to state relations, in particular relations between facts. But he argues that this is so only on a broad sense of 'relation', according to which there is a relation corresponding to every relational predicate. It is not so on a narrower ontological sense, according to which relations are universals existing independently of thought and language. Moreover, it is only in a broad sense of 'facts', according to which facts correspond to true statements, that causal statements appear to relate facts. It is not so on a narrower ontological sense in which facts are the ontological grounds or ultimate truthmakers for statements. Mellor reserves the term 'facta' for the truthmakers of statements. In his discussion of whether causation is a real relation, Mellor is concerned with the question whether the facta that are the truthmakers for causal statements like 'E because C' consist in a genuine relation between facta.

Mellor advances two arguments against the view that the truthmaker for a causal statement is a relation between facta. One argument is that even if facta exist to act as relata – which, as we shall see, cannot always be taken for granted – a causal statement need not be made true by the existence of a relation between such facta (Mellor 1995: 162–5). He gives an example to illustrate this point. In a golf game, Sue pulls her drive to the left, making her ball bounce off a tree and, by a fluke, giving her a hole in one. Mellor's favoured description of the causal relations in this example is that Sue holed out in one *because* she drove her ball but *despite* the fact that she pulled her drive to the left. This accords with his counterfactual dependence criterion of causation. For Sue's driving the ball made it more likely that the ball would fall into the cup for a hole in one, but her pulling her drive to the left made it less likely. However, it is not plausible, he argues, to say the truthmaker for the positive causal statement consists in a relation between two facta. Even if we suppose that her holing out in one is a factum, it is not plausible to suppose that the cause of this, her driving the ball, is a factum as well. This is so because there is another fact that entails, but is not entailed by, this fact,

which must therefore be a factum, namely her pulling her drive to the left. But the relationship between these facta, Sue's pulling her drive to the left and her holing out in one, is not causal because the first does not increase the chance of the other.

For two reasons I find this argument to be the less compelling of Mellor's arguments. First, it depends on Mellor's very contentious claim that it was Sue's driving the ball, but not her pulling her drive to the left, that caused the ball to fall into the cup for a hole in one. As he notes, this example is a variant of some much discussed problem cases initially cited to show that a cause need not increase the chance of its effect (see, for example, Salmon 1984: 192–202). Those who have advanced these examples would insist – correctly in my view – that Sue holed out in one because she pulled her drive to the left. After all, the ball fell into the cup because it hit the tree and it hit the tree because Sue pulled her drive to the left. Of course, Mellor has independent reasons, stemming from his adherence to the counterfactual dependence criterion of causation, for thinking that his is the right description of the causal facts of the situation. But those of us who reject this criterion will find his description of the causal facts far from compulsory.

My second reason for finding the argument less persuasive is that it depends on a dubious principle about truthmaking facta. More precisely, it depends on the principle that if some fact P is entailed by, but does not entail, some other fact Q, then P cannot be a genuine factum. Is this principle at all plausible? In the particular example, Sue's driving is said to be entailed by, but not to entail, her pulling her drive to the left, which is then assumed to be a factum. But the very same principle that entails that Sue's driving is not a factum would surely imply that her pulling her drive to the left is not a factum either. For there is a fact that entails, but is not entailed by, this fact: that she pulls her drive to the left with a minute twist of her wrist. Indeed, the same style of argument would show that this cannot be a factum either because a still more precise description can be given of Sue's action. Indeed, since any positive event or action can be specified in more and more fine-grained ways, it would seem that Mellor's principle commits him to supposing that there are no facta at all, or ones that are maximally specific with respect to some particular sort of information. But without any account of the rules for determining such maximally specific sorts of information, the notion of a factum constrained by the above principle is useless in addressing the question of the truthmakers for ordinary causal statements. So, in future, I shall understand the notion of a factum in such a way that it need not conform to this principle.

Mellor's second argument against the view that causation is a relation between facta is more persuasive, I believe. David Lewis has called it the missing relatum objection (Lewis 1999). As Mellor formulates the objection, there are true causal statements involving negative occurrences such as:

(3)　Kim has no children because she took contraceptives.
(4)　Kim works full time because she has no children.

However, the absence that is said to be a cause in (4) and an effect in (3) is not a genuine factum, since there is nothing in the world to act as truthmaker for the statement 'Kim has no children'. If the truthmaker for a true causal statement were always a relation between facta, the truth of causal statements such as (3) and (4) would require the existence of a factum corresponding to Kim's lack of children. But since such a factum does not exist, the truthmaker for a causal statement cannot be a genuine relation between facta.

Lewis also endorses this objection against the view that causation is a relation, although he thinks that if it were a relation its relata would be events rather than facts (Lewis 2002). Moreover, he has an additional argument that is targeted directly at the more specific view that causation consists in an intrinsic relation (Lewis 1999). His counterexample to this view involves a case of so-called *double prevention*: a cause prevents something which, had it not been prevented, would have prevented the effect.[5]

> Case 3: A collision between billiard balls 1 and 2 prevents ball 1 from continuing on its way and hitting ball 3. The collision of 1 and 3, had it occurred, would have prevented the subsequent collision of balls 3 and 4. But since in fact the collision of 1 and 3 was prevented, the collision of 3 and 4 was unprevented. Accordingly, the collision of 1 and 2 causes the collision of 3 and 4. Indeed, there is a matching counterfactual dependence: if there had been no collision between 1 and 2, there would have been no chance of a collision between 3 and 4.
>
> (Lewis 1999: 13)

There are two problems posed by the example, according to Lewis. First, the counterfactual dependence is an extrinsic matter. Had there been some other obstruction that would have stopped ball 1 from hitting ball 3, the collision of 3 and 4 would not have depended upon the collision of 1 and 2. Second, there is no continuous chain of events running from cause to effect. Between the collision of balls 1 and 2 and the collision of balls 3 and 4, nothing much happens. What matters here, Lewis argues, is not what happens, but what does not happen.

It is interesting to note here that such examples of double prevention may also pose a problem for Mellor's theory. Recall that Mellor's theory does not require that an indirect cause be connected to its effect by counterfactual dependence, but merely that it be connected by a chain of contiguous causes and effects, each linked by a counterfactual dependence. It is reasonable to interpret this requirement in the light of his views that facta must be causation's relata if it has any. When interpreted in this way and when applied to Lewis's example of double prevention, the requirement necessitates that there be a causal chain of contiguous facta running from the collision of balls 1 and 2 to the collision of balls 3 and 4. However, since nothing happens in the spatial region between these collisions, there are no facta to form this causal chain.

In summary, then, Mellor's missing relatum objection and Lewis's objection from double prevention seem to raise genuine difficulties for the view that causation is an intrinsic relation. I am faced with a dilemma at this point. On the one hand, examples of pre-emption and overdetermination highlight the plausibility of this view. On the other hand, the fact that causes and effects can be absences, which are not real things, seems to lead into the kinds of difficulties Mellor and Lewis raise. How is this dilemma to be resolved?

4   Intrinsic relations reconsidered

We can begin to resolve this dilemma, I believe, if we refocus our attention on the notion of an intrinsic relation. So far I have relied on Lewis's explication of intrinsic relations in terms of perfectly natural properties and relations. On this understanding, an intrinsic relation is one that holds just in virtue of the perfectly natural properties and relations holding of its relata. There is, however, an independent reason to be dissatisfied with this explication. The notion of intrinsicality that it explicates has its most natural application in the actual world in fundamental physics. Lewis claims that the perfectly natural properties and relations are coextensive, in the actual world at least, with the fundamental physical properties and relations. However, we need to be able to explain the notion of intrinsicality as it applies to causal processes studied outside fundamental physics. For example, the commonsense causal claim that the terrorist attacks in the United States in September 2001 caused an immediate dramatic fall on Wall Street is made true, I claim, by an intrinsic process. We are justified in believing in the existence of such a process and seeing it as the truthmaker for this causal claim without having the faintest idea how it might be analysed in terms of the properties and relations of fundamental physics.

It seems best, then, to start from scratch to explicate the notion of intrinsicality in such a way that it applies smoothly to the causal processes studied in the higher-level sciences as well as those studied in physical science. We can make a start on such an explication by noticing an implicit relativity involved in our concept of intrinsicality, and indeed in the concept of causation that is to be explained in terms of it. This relativity reflects the fact that our causal thinking is steeped in abstraction. Within any spatio-temporal region there are many different levels of causation, and within each level many cross-cutting and intersecting causal processes. To determine the structure of these processes, we are forced to focus selectively on some aspects of what is going on and to background others. The causal schemas by which we interpret the world are irremediably permeated by abstractions that enable this selective focusing. One form of abstraction involves the identification within a given spatio-temporal region of *a system of a certain kind*.

A particular system of a certain kind consists in a set of constituent objects configured in specific ways. Clearly, the kinds of systems investigated by astronomers and cosmologists are different from the systems investigated by

biologists and economists: solar systems and galaxies involve different kinds of constituent objects from economies, markets, species and populations. However, a system is not just a set of objects, but a set of objects that have certain properties and relations. And not any old properties and relations are relevant to the identification of a system as being of a certain kind. For example, a set of astronomical bodies can be individuated as a kind of planetary system by way of each body's relation to other bodies in the system, but not by way of their relations to objects outside the system. In short, a system of a given kind is a set of constituent objects internally organized in a distinctive fashion. The properties and relations that configure the objects into a system must be intrinsic to that kind of system.

The concept of intrinsicality at issue here is not the concept of properties and relations intrinsic *tout court*, but those intrinsic to a kind of system. It will suffice for our purposes to explain the intuitive idea behind this concept, rather than to present a full analysis, which is beyond the scope of this chapter. Modifying an idea of Jaegwon Kim's (1982) concerning the simple concepts, I shall say the following:

> (5)   A property F is *extrinsic to a system of kind K* if and only if, necessarily, a member of a set of objects constituting a system of kind K has F only if some contingent object wholly distinct from the set exists.

For example, the extrinsic properties of an astronomical body that is part of a distant planetary system might include being observed by some human on earth.

The concept of a property intrinsic to a kind of system is defined in converse fashion:

> (6)   A property F is *intrinsic to a system of kind K* if and only if, possibly, a member of the set of objects constituting a system of kind K has F although no contingent object wholly distinct from the set exists.

For example, the intrinsic properties of a planetary system include the mass and shape of the individual astronomical bodies. But the intrinsic properties of the system need not all be intrinsic properties *simpliciter*. For example, the property of being gravitationally attracted to another member of the planetary system is an intrinsic property of the system, but it is not an intrinsic property *simpliciter*.[6] Notice that this definition does not prohibit negative, conjunctive or even disjunctive properties from being intrinsic properties of a system.

There is a vast multitude of kinds of systems, but very few are of real interest to us. For the most part, we are interested only in the kinds of systems that evolve in lawful ways. As examples of these kinds of systems, we need only consider the kinds of systems investigated in scientific theories. Typically speaking, a scientific theory provides an abstract description of a certain kind of system in terms of a select set of state variables, and explains the behaviour

of systems of the kind in question by showing how these variables change over time in conformity with certain laws. For example, classical mechanics employs the state variables of mass, position and momentum, and explains the motion of mechanical bodies, described in terms of these variables, by way of the Newtonian laws. Invariably, the state variables that a theory employs are intrinsic properties and relations of the systems under consideration. To summarize:

> (7) *A lawful kind of system* is a kind of system whose intrinsic properties and relations (state variables) evolve over time in conformity with a common set of laws.

In general, a lawful kind of system supervenes on a set of intrinsic properties and relations that conform to a common set of laws. In other words, any two particular systems with the same intrinsic properties and relations conforming to the same laws must both belong, or fail to belong, to a given lawful kind of system.

In terms of the concepts at hand, we are in a position to explain the notion of an intrinsic process that is going to play a central role in the modified functionalist analysis of causation. I suggest the following definition.

> (8) *An intrinsic process* holding in a lawful kind of system is a temporally ordered sequence of states that instantiate the intrinsic properties and relations that constitute that kind of system.

For example, the intrinsic process that I suppose is the truthmaker for the commonsense causal claim 'The terrorist attack in the United States caused a dramatic fall on Wall Street' might consist in some sequence of states such as the terrorist attack on United States facilities, vast economic losses to major companies, a loss of confidence among major investors, a delay in the reopening of Wall Street stock market and widespread panic among traders at the reopening. Here I assume that the kind of system that is implicitly being considered in this commonsense causal claim can be specified rather loosely as that of an open market economy. Whichever way the kind of system is precisely specified, it is clear that intrinsic processes of this kind are not identified in terms of the properties and relations of fundamental physics.

Finally, we are in a position to consider how the functionalist analysis of causation should be modified to accommodate this relativized understanding of an intrinsic process. First, we would expect that there should be a matching relativization in the causal concept. Elsewhere (Menzies 2002) I have argued that the causal concept must be understood as relativized to the contextual parameter of a lawful kind of system. We shall consider some evidence in support of this context relativity in the next section. Second, we would expect that the analysis of causation should encompass all causal statements, whether they concern positive or negative occurrences. In order to be as neutral as possible

over the contentious issue of the nature of the causal relata, I shall simply talk of them as property instances.[7] Whichever way such property instances are to be understood, they are to include instances of negative as well as positive properties. With these two preliminary remarks, let me state the modified functionalist analysis:

> (9)   If c is an instance of property F and e an instance of property G in a lawful system of kind K, then c and e are *causally related* if and only if (a) there is a kind of intrinsic process that typically holds in systems of kind K when a G-instance is counterfactually dependent on an F-instance; and (b) a process of this kind holds in the particular system of kind K that includes c and e.

This analysis is meant to apply uniformly to all causal statements, whether they concern positive or negative occurrences.

This analysis works as well as the old one when it comes to explaining our intuitions about pre-emption and overdetermination. For example, suppose that we see Case 1 as exemplifying the kind of system that consists of a sole assassin shooting with a rifle at an unprotected person. There are clearly counterfactual dependences between shootings and deaths in this kind of system; and furthermore there is an obvious kind of intrinsic process that typically accompanies these dependences. This kind of process involves an assassin pulling the trigger of his rifle, a bullet being released by the rifle, the bullet travelling through the air and hitting the body of the person, followed by the person's death. In Cases 1 and 2 we look for a process of this kind to discriminate the actual from the potential causes.

But it is important to recognize that the analysis applies just as readily to the examples in which the cause and effect are absences, that is instances of negative properties. Take Mellor's example in which Kim's use of the contraceptive pill causes her not to have children. If we suppose that the relevant kind of system involved in this causal statement is that of a human female's body functioning according to the laws of human anatomy, then we will find counterfactual dependences holding in systems of this kind between the use of contraceptives and the absence of children. Moreover, we will find that there is a kind of intrinsic process that typically accompanies such counterfactual dependences, a process consisting of ingestion of oestrogen, disruption of ovulation, absence of fertilization and absence of fetus formation. (Note that all these properties count as intrinsic properties of this kind of system.) Moreover, if this kind of intrinsic process obtains in Kim's case, then there is a truthmaker for the claim that Kim does not have children because she uses the contraceptive pill.

Similarly, the analysis applies straightforwardly to the example of double prevention that Lewis discusses. Let us suppose that the kind of system involved in the relevant causal claim is one consisting of four billiard balls with momenta of the same magnitude and direction as those in the original

example. Once more, we can expect to find a kind of intrinsic process that typically accompanies a counterfactual dependence between collisions in this kind of system. The process will consist of balls 1 and 2 colliding, with each moving in a different direction from its initial direction, and then balls 3 and 4 colliding without interference from the other balls. Even though there may be a spatial gap between the collisions, this process is nonetheless temporally continuous. Since such a process obtains in the actual situation under consideration, there is a truthmaker for the causal claim that the collision of balls 1 and 2 caused the collision of balls 3 and 4.

Accordingly, I suggest that causal situations involving absences or double prevention present no difficulty for the modified account. This account allows me to hold on to the view that causal statements are made true by intrinsic processes, while maintaining that there is but one causal concept whose analysis applies to all causal statements, regardless of whether they relate positive or negative occurrences.

Of course, I can maintain both views because I now construe the notion of an intrinsic process in a broad and flexible way. It might be thought that the notion of an intrinsic process has been made so broad and flexible as to be theoretically useless. But this is not so, I would argue. The notion of an intrinsic process still plays an essential role in explaining our intuitions about causation in cases of pre-emption. Consider the following modification of Case 3, Lewis's example of double prevention:

> Case 4:   The set-up is the same as in case 3, but there is a billiard ball 5 that is on a collision course with ball 1. If ball 2 had not first collided with ball 1, then ball 5 would have a bit later on, so that one or other collision would have prevented ball 1 from colliding with ball 3. So ball 5 is back-up preventer of the collision between balls 1 and 3, which would have prevented the collision of balls 3 and 4.

This example of so-called *pre-emptive prevention* is an interesting test case for the modified functionalist analysis.[8] The analysis should be able to discriminate the actual from the potential preventer of the collision between balls 1 and 3. Notice that a pure counterfactual analysis cannot do this. For the absence of a collision between balls 1 and 3 does not depend counterfactually either on the motion of ball 2 or on the motion of ball 5: if one of these events had occurred without the other, the collision between balls 1 and 3 would still have been prevented. But this example is readily handled by the modified functionalist analysis. If we consider the situation under consideration as an instance of the lawful kind of system in which one billiard ball collision prevents a later one, we can see that the kind of intrinsic process that underlies this prevention obtains in this particular case. An essential part of such an intrinsic process is a collision of two balls that disrupts one of them from its collision course with a third ball. Clearly, we can see that the motion of ball 1, but not the motion

of ball 5, initiates a process of this kind; and so we are able to discriminate the actual from the potential preventer in this example.

5   The relativity to a kind of system

It might be thought that a defect of the present account of causation is that it makes the concept of causation context sensitive by making it relative to a lawful kind of system. However, I think, contrary to this line of thought, that this apparent weakness is one of the great strengths of the analysis. Our concept of causation is marked by a certain degree of indeterminacy and vagueness: we display ambivalence in our causal judgements about certain kinds of situation. By understanding the causal concept as involving a contextual parameter that can be set in various ways in different contexts, one can explain this indeterminacy. Let me illustrate this with just one type of example that is germane to our discussion here.

Cases of pre-emptive prevention have been much discussed of late as interesting test cases for theories of causation. It appears that an indeterminacy affects our judgement about them. Consider the following example:

> Case 5: You reach out and catch a passing cricket ball. The next thing along in the ball's direction of motion was a solid brick wall. Beyond that was a window. Did your action prevent the ball hitting the window? (Did it cause the ball to *not* hit the window?)
>
> (McDermott 1995: 525)

People express conflicting intuitions about this example. When it is pointed out that the presence of the brick wall means that the window was never in any danger of being broken, people are inclined to say that your catch did not prevent the ball hitting the window. On the other hand, when it is pointed out that something must have prevented the ball hitting the window, they agree that it must have been your catch that did the preventing.

All of this makes sense in terms of the theory of causation presented above. There are different ways of modelling the causal structure of the situation depending on which kind of system one sees it as instantiating. Suppose one thinks of the relevant kind of system as one that includes you, the ball, the window and the brick wall with their given spatio-temporal arrangements. There is no counterfactual dependence between your catch and the ball's not hitting the window in this kind of system and so a forteriori no intrinsic process accompanying such a dependence. On the other hand, suppose one thinks of the situation as instantiating the kind of system that abstracts away from the presence of the brick wall – a kind of system that includes you, the ball and the window but excludes the brick wall as an object extrinsic to the system. Then there are counterfactual dependences between your catch and the ball's not hitting the window, and indeed these dependences pick out an

intrinsic process of a certain kind. Moreover, a process of this kind holds in the particular situation under consideration, so supporting the judgement that your catch prevented the ball from hitting the window. In this way, the indeterminacy in our causal judgements can be traced to the multiple ways in which the contextual parameter of a kind of system can be fixed.

This account predicts that our readiness to accept the causal judgement that your catch prevented the ball's hitting the window goes hand-in-hand with our readiness to see the situation in terms of a kind of system that abstracts away from the presence of the brick wall. In McDermott's example, our readiness to do this wavers somewhat. But now consider a modification of the example introduced by John Collins:

> Case 6: You reach out and catch a passing cricket ball. The next thing along in the ball's direction of motion was my hand. (I leapt up to catch the ball, but because of your faster reaction you caught the ball just in front of the point at which my hand was raised.) Beyond our outstretched hands is a window. Did your action prevent the ball hitting the window?
>
> (Collins 2001: 223)

Collins detects some indeterminacy in our causal judgement about whether your catch prevented the window from being broken. But he claims, correctly I think, that we are more inclined to accept it in this example than in McDermott's original example. He explains this in terms of how far-fetched it is to entertain the absence of the back-up preventer. It is easy to entertain the absence of my hand ready to take the catch: one simply imagines that I get my timing wrong so that when I leap I do so not at the right moment to be ready to take the catch. It is more far-fetched, on the other hand, to suppose that the brick wall is absent or that the ball would miraculously pass straight through it (Collins 2001: 227–9).

I think that Collins's explanation is on the right track to the extent that it links our willingness to accept the judgement that your catch prevented the ball's hitting the window with our willingness to abstract away from the presence of the back-up preventer. I would go further and explain this linkage in terms of the way we model the causal structure of a given situation in terms of kinds of systems that abstract away from the presence of factors that are viewed as extrinsic to the system. Again, I think there is something to Collins's explanation of our varying degrees of willingness to do this in terms of how far-fetched it is to imagine the absence of the back-up preventer. But I would prefer to see the matter of how far-fetched it is to imagine such things as rooted in fairly objective issues about the features of the situation itself: How permanent a feature of the set-up is the back-up preventer? Is it something that is an external intrusion in an otherwise isolated system? Would the system that abstracts away from its presence fall under wider, more robust laws than the system that retains its presence? Such considerations can

yield fairly objective reasons for modelling a situation in terms of one kind of system rather than another.

However, it is crucial in such discussions to keep in mind the implicit relativity of causal judgements to a lawful kind of system. Overlooking this context relativity makes one more liable to fall into conceptual traps. As an illustration of this, consider an argument of Collins's to the effect that examples of pre-emptive prevention falsify any theory that takes causation to consist in an intrinsic process. He observes that your catch prevents the window from breaking when it pre-empts my catch from preventing the window from breaking, but not when it pre-empts the brick wall from doing so. Yet the only difference between these cases, he says, has to do with features extrinsic to the simple process involving your catch. 'The process that includes the ball's flight, your catch and the window's not breaking is causal in the case where my hand was poised behind yours to take the catch, but it is *not* causal in the case where a brick wall is there instead of me.' (Collins 2001: 226).

A sufficient counter to this argument starts from the observation that the concept of an intrinsic process, like that of the causal concept, must be seen to be relative to a lawful kind of system. Indeed, as we have seen, intrinsic processes are often widespread features of entire systems, rather than localized parts of the system. Now notice that the two causal judgements that Collins's argument turns on involve quite different kinds of systems. The judgement about the modified example that your catch prevented the ball hitting the window involves the kind of system that *excludes* the back-up preventer of my outstretched hand. The opposite judgement about the original example involves a kind of system that *includes* the back-up preventer of the wall. In order to compare the intrinsic processes that could act as truthmakers for these judgements, we have to consider sequences of states involving all the intrinsic features of these systems. In the first system the intrinsic process will consist of a sequence of states holding true of you, the ball and the window, whereas in the second system the intrinsic process will consist of a sequence of states holding true of you, the ball, the brick wall and the window. The presence of the brick wall in one system but not the other makes a big difference about what counts as the intrinsic features of the set-ups. It is false to say, therefore, that the two set-ups agree in intrinsic processes and differ only in matters extrinsic to these processes, namely the presence of the back-up preventer. The difference between the set-ups with respect to the presence of the brick wall makes for a difference in intrinsic processes, a difference that ultimately underlies our readiness to accept a causal judgement about one set-up but not the other.

Notes

1 See Noordhof (1998) for a good discussion of this point.
2 In *early* pre-emption examples, *the main process* that goes through to completion and brings about the effect cuts short all alternative processes before the effect has occurred, whereas in *late* pre-emption examples the main process

goes through to completion, but it is *the effect* itself that cuts short all the other alternative processes after it has occurred.

3 For further discussion of the conflict between the commonsense conception of causation as an intrinsic relation and the standard Humean position see Menzies (1998).

4 Lewis's definition of an intrinsic relation appears in Lewis (1983a; 1986b). He actually defines two kinds of intrinsic relations: relations intrinsic to their relata and relations intrinsic to their pairs. The relevant kind of intrinsic relations I consider in connection with causation correspond to relations intrinsic to their pairs.

5 Similar cases of double prevention are discussed in McDermott (1995) and Hall (2002). Hall explicitly draws out the implications of such cases for the supposed intrinsic character of causation.

6 A problem infects these definitions parallel to the problem Lewis (1983b) pointed out for Kim's definition of the simple concepts. Modifying some concepts introduced by Lewis, let us say that a system is *accompanied* if and only if it coexists with some contingent object wholly distinct from it, and *lonely* if and only if it does not so coexist. The definitions I have presented amount to saying that that the extrinsic properties of a system are those implied by the accompaniment of the system and the intrinsic properties of a system are those compatible with its loneliness. The problem is that loneliness of a system is intuitively an extrinsic property of the system (since it can differ between duplicates of the system), but it counts as an intrinsic property by the definition (since it is compatible with itself). One possible remedy to this problem may be to adapt for our purposes the refinement of Kim's original idea to be found in Langton and Lewis (1998). This refinement is supposed to circumvent the defect Lewis detected in Kim's original idea.

7 For a more detailed discussion of the nature of the causal relata see Menzies (1989), in which I argue that fact-like entities, which I call situations, are the primary relata of causation.

8 Examples of pre-emptive prevention have been discussed in McDermott (1995), Lewis (1999) and Collins (2001).

References

Collins, J. (2001) 'Preemptive prevention', *Journal of Philosophy* 97: 223–34.

Hall, N. (2002) 'Two concepts of causation', in J. Collins, N. Hall and L. Paul (eds) *Causation and Counterfactuals*, Cambridge, MA: MIT Press.

Kim, J. (1982), 'Psychophysical supervenience', *Philosophical Studies* 41: 51–70.

Lewis, D. (1983a) 'New work for a theory of universals', *Australasian Journal of Philosophy* 61: 343–77.

—— (1983b) 'Extrinsic properties', *Philosophical Studies* 44: 197–200.

—— (1986a) *Philosophical Papers*, Vol. II, Oxford: Oxford University Press.

—— (1986b) *On the Plurality of Worlds*, Oxford: Basil Blackwell.

—— (1999) 'Causation as influence', University of Melbourne Preprint 1/99. Reprinted in shortened version in *Journal of Philosophy* (2001) 97: 182–97.

—— (2002) 'Void and object', in J. Collins, N. Hall and L. Paul (eds) *Causation and Counterfactuals*, Cambridge, MA: MIT Press.

Langton, R. and Lewis, D. (1998) 'Defining "intrinsic"', *Philosophy and Phenomenological Research* 58: 33–45.

McDermott, M. (1995) 'Redundant causation', *British Journal for the Philosophy of Science* 46: 523–44.

Mellor, D. H. (1971) *The Matter of Chance*, Cambridge, UK: Cambridge University Press.

—— (1981) *Real Time*, Cambridge, UK: Cambridge University Press.

—— (1991) *Matters of Metaphysics*, Cambridge, UK: Cambridge University Press.

—— (1995) *The Facts of Causation*, London: Routledge.

Menzies, P. (1989) 'A unified account of causal relata' *Australasian Journal of Philosophy* 67: 59–83.

—— (1996) 'Probabilistic causation and the pre-emption problem', *Mind* 105: 85–117.

—— (1998) 'How justified are the Humean doubts about intrinsic causal links', *Communication and Cognition* 31: 339–64.

—— (1999) 'Intrinsic versus extrinsic conceptions of causation', in H. Sankey (ed.) *Causation and Laws of Nature*, Dordrecht: Kluwer.

—— (2002) 'Difference-making in context', in J. Collins, N. Hall and L. Paul (eds) *Causation and Counterfactuals*, Cambridge, MA: MIT Press.

Noordhof, P. (1998) 'Critical review of *The Facts of Causation*', *Mind* 107: 428–39.

Salmon, W. (1984) *Scientific Explanation and the Causal Structure of the World*, Princeton: Princeton University Press.

9 Dispositions and conditionals

Isaac Levi

1 Introduction

Are dispositions equivalent to conditionals? Does '*x* is fragile' mean the same as 'if *x* were dropped it would break'?

Hugh Mellor and I agree that disposition predicates are truly or falsely predicable of objects or systems. Whether a particular glass is fragile or not is a question of fact (once various ambiguities are set to one side). On the other hand, I follow Ramsey in thinking that conditionals like 'if *x* were dropped, it would break' are neither true nor false. One cannot be in suspense coherently whether it is true or false. One cannot assign coherently a (subjective) degree of probability to its being true.

Conditionals express judgements of conditional possibility and impossibility. Like expressions of judgements of serious possibility and impossibility and like expressions of judgements of subjective probability, they are *supported* by the inquirer's state of full belief and (in the case of probability judgement) the inquirer's confirmational commitment but are not *entailed* by it.

If I am right about conditionals, satisfaction conditions for disposition predicates cannot be supplied by invoking conditionals.

I shall attempt here to offer a brief elaboration of the position I favour without hoping to convince Hugh Mellor of its correctness. I have found out on more than one occasion that Mellor is immune to the approach to dispositions, abilities, chances and the like I favour, as well as to my ideas concerning conditionals. Immune or not, I have always found bouncing my ideas off the walls of his impregnable conceptual fortress rewarding. This confrontation is but a poor but sincere token of my friendship, respect and admiration for him and his work.

2 Dispositions as placeholders

Does tossing the coin near the surface of the earth cause it to land heads-or-tails? We may plausibly claim that the coin invariably lands heads up or tails up on a toss *given that it has the surefire disposition to do so*. The universal regularity we can invoke here is that any coin having the surefire disposition

that is tossed near the surface of the earth invariably lands heads up or lands tails up (lands heads-or-tails).

Some coins do not have that disposition. They have the ability to land on their edge when tossed near the surface of the earth. Coins with heads on both sides do have the disposition to land heads-or-tails on a toss. Unlike 'normal' coins, they lack the ability to land tails on a toss. They have the surefire disposition to land heads on a toss. Yet deviant coins of both kinds may sometimes land heads or tails every time in the actual history of tossings that they are made to endure.

There is no true universal regularity correlating tossings of coins near the surface of the earth and landings heads-or-tails. Coins possessing the surefire disposition to land heads-or-tails on a toss near the surface of earth invariably land heads-or-tails when tossed. Coins with thick edges lack the disposition. Whether or not they always land heads-or-tails, their doing so is not explanatory. In such cases, causality cannot be explicated by appealing to a constant conjunction between tossings and landings heads-or-tails. We need to appeal to the presence of the disposition to land heads-or-tails as well.

Appealing to dispositional properties to account for causality is sometimes alleged to be an abandonment of the covering law model of explanation. This, I think, is Elster's (1999) view. We cannot claim that everything breaks when dropped. Some things do and some things do not. An individual who has beliefs he wishes were false may indulge in wishful thinking. But he may alter his values and preferences instead. Or he may resign himself to the grim truth. Elster despairs of devising a predictive theory for anticipating how individuals will respond to bad news. But he thinks we can explain retrospectively why an individual does respond by appealing to what he calls 'mechanisms'. Elster's mechanisms appear to me to resemble what are generally called 'dispositions'. With the aid of dispositions we can formulate regularities linking landing heads-or-tails with tossings in a covering law manner. Tossing coins may not invariably lead to landing heads-or-tails. But tossing coins with the appropriate disposition does. Appealing to mechanisms or dispositions provides a 'measure of explanatory power'; but Elster seems to think that further integration into systematic theory is not promising – at least not in some areas of the social sciences. In the social sciences at any rate, inquirers need to settle for such second-rate explanation.

It seems to me that Elster has things backwards. Disposition predicates are used as placeholders in stopgap covering laws to provide explanation sketches that inquirers seek to fill out within the framework of more systematic theory. In claiming that they are placeholders I am not suggesting that they fail to be predicates meaningfully applied to objects or systems. Nor am I suggesting that the stopgap covering laws fail to be covering laws. The stopgap laws in which dispositional placeholders appear are not fully satisfactory for the purposes of explanation. Inquiry is required in order to integrate the placeholders into explanatorily adequate theories. But there would be no point to introducing dispositional concepts were it not for their role in enabling

stopgap explanations pending the conduct of inquiry seeking to convert the dispositional predicates into explanatorily adequate theoretical terms in some explanatorily satisfactory theory.

Suppose that despair of being able to integrate the stopgap covering laws in explanatorily satisfactory theories sets in and inquirers settle for the explanation sketches as the best that can be done. In that setting, explanation by disposition becomes suspect in the way that explaining the responses of those who imbibe opium by appealing to its dormitive virtue is.

The explicit disposition predicate 'D(R/S)' (to be read as 'is disposed to respond in manner R on a trial of kind S') may be taken as a primitive term characterized by the following postulate:

(1) $(x)(t)\{D(R/S)xt \supset [Sxt \supset Rxt])\}.$

Postulate (1) is a covering law that can be used to explain why some object responded in manner R. We cannot claim that everything R's when S'd as a matter of natural law. But we can invoke the claim that everything with the disposition D(R/S) R's when S'd. Far from leading to the abandonment of covering-law explanation, the use of disposition predicates enhances it.

Condition (1) is one component of what Carnap used to call a bilateral reduction sentence. It fails to secure necessary and sufficient satisfaction conditions for the disposition predicate 'D(R/S)' in terms of test behaviour characterized by means of trials of kind S and responses of kind R. However, inquirers are not prevented from finding out that attributions of dispositions are true of systems by an appeal to test behaviour just because necessary and sufficient satisfaction conditions exclusively in terms of test behaviour are lacking. Appeal to test behaviour may need to be supplemented by additional information. Or the inquirer may resort to drawing ampliative inferences. The background information obtained in this fashion may be fully believed and, hence, judged to be certain by X. Indeed, (1) itself is part of the inquirer's full beliefs and is no more or less certain than other items. The distinctive character of (1) is not found in its certainty. Other full beliefs may be vulnerable to being given up to a greater degree than (1). In most contexts, (1) is maximally well entrenched.[1]

There is a good reason for this. Postulate (1) is introduced, as already indicated, in order to supply a covering law for the purpose of explaining why the system in question responded in manner R. (1) is not established experimentally. But it is not a product of definition. The primitive disposition term is introduced in tandem with (1) as a stopgap measure. The inquirer thinks or at least hopes that future research will find an acceptable way to integrate the primitive and its postulate into an acceptable comprehensive theory. Once explanatory integration has proven successful, the erstwhile disposition term is no longer a placeholder in stopgap explanation. The theory may then be revised in a manner that leads to abandoning (1) without cost. *But as long as the disposition term continues to function as a placeholder in stopgap explanation, postulate (1)*

needs to be retained. Without the postulate the placeholder can no longer serve its placeholding function.

Thus, viewing dispositions as placeholders in stopgap explanations supports Elster's contention that dispositions or mechanisms are introduced for the purpose of supplying somewhat disreputable explanations. Elster, however, thinks that progress towards removing placeholder status is unlikely. We should resign ourselves to the best we can get. The placeholder view resists this attitude of despair. Explanation by disposition should not be offered with resignation but with recognition of the need for future inquiry in order to integrate the placeholding disposition predicate into an explanatorily satisfactory theory and with the hope that future inquiry will succeed. To use disposition predicates for purposes of explanation without any expectation that future inquiry will improve on the explanations yielded and without the intention of promoting such inquiry is to make a mystery out of dispositionality and to defend the use of vacuous explanation as the best that can be had.

There are other kinds of placeholders besides dispositions. Let us grant then that (1) is postulated to be a law. Its postulation is defended by noting that some things respond in manner R when S'd and others do not. Things disposed to respond in manner R upon being S'd respond in manner R when S'd. It is possible for the other things to fail to R when S'd.

To say that it is possible for x to R when S'd is to attribute the ability to R upon being subjected to a trial of kind S. The ability predicate 'A(R/S)x' is the dual of the surefire disposition predicate 'D(R/S)'. A(R/S)x if and only if $\sim$D($\sim$R/S)x. 'A(R/S)' has sufficient satisfaction conditions in terms of test behaviour given by (2).

(2) $(x)(t)\{Sxt \ \& \ Rxt \supset A(R/S)xt\}$.

Postulate (2) fails to specify a necessary condition for the presence of the ability to respond in manner R on a trial of kind S (the possibility *for* object or system x to R on being S'd). Nonetheless, information that x has the ability and that a trial of kind S has been implemented on x at a given time warrants a judgement of the serious or doxastic possibility *that* an outcome of kind R of test of kind S will occur unless the inquirer has relevant information ruling out such possibility.

Carnap (1950) and many others who deployed his notion of bilateral reduction sentences to characterize disposition predicates thought that the following condition holds:

(3) $(x)(t)[Sxt \ \& \ Rxt \supset D(R/S)xt]$.

The sufficient condition for the presence of the ability in (2) became a sufficient condition for the presence of the disposition in (3). This is surely a mistake. Obtaining heads on a single toss of a coin is not sufficient for its having the surefire disposition to land heads on a toss. Hugh Mellor (1974) explicitly noted this important point.

It is true that data about the behaviour of magnets rotating in copper coils might warrant coming to the conclusion that under appropriate conditions such experiments will invariably induce an electric current. But this conclusion is not entailed by the data specified alone. It is either entailed by some theory of electromagnetism or is obtained by some series of well-designed repetitions of the experiment. But might not such a theory at least on some occasions imply the truth (and support the lawlikeness) of a generalization of the form (3)? What is the difference between the status of (1) and the status of (3)?

The placeholder account of dispositions proposed in Levi and Morgenbesser (1964) cannot distinguish between postulates in which a theoretical term is embedded with respect to analyticity. The difference between (1) and (3) is not meaning theoretic, whatever that is. The placeholder account characterizes disposition predicates in terms of their function in deliberation and inquiry. As already noted, postulate (1) ought not to be given up as long as the disposition predicate serves its placeholder function. As a consequence, (1) carries informational value that inquirers should be reluctant to give up. (1) is, in the terminology of AGM-type (after Alchourrón, Gärdenfors and Makinson) theories of belief revision, well entrenched (see Gärdenfors 1988). Additional structure may be added to the characterization of the disposition predicate, including, perhaps, generalizations like (3). But as long as the disposition predicate is problem raising, postulate (1) must be better entrenched than (3).

Postulate (1) is insufficient to secure conformity with two requirements Mellor (1974: 118) imposes on 'real' properties: that they display themselves in more ways than one and that two objects which differ with respect to one property must differ with respect to another.

Mellor insists that if dispositions are real properties they satisfy these two conditions. Neither condition specified by Mellor need be met, however, by problem-raising disposition predicates that have not as yet been integrated adequately into a theoretical framework. Certainly, as Mellor (2000) emphasizes, some, and perhaps most, disposition predicates in natural language are associated with a variety of 'conditions of application'. According to Mellor, such specifications of conditions of application are given by reduction sentences that are variants of Carnapian reduction sentences such as (1). They differ from Carnap's reduction sentences because they are subjunctive conditionals. Since I deny that such conditionals can carry truth values whereas statements attributing dispositions to things do, clearly I cannot agree with Mellor's approach to supplying conditions of application to disposition predicates.

Nonetheless, we often use the same predicate (such as 'is fragile' or 'is magnetic') in diverse Carnapian reduction sentences relating that disposition predicate to many different kinds of outcomes or 'displays' or spelling out different kinds of conditions under which a given display may be realized. To the extent that this is so, we may expect Mellor's two conditions to be satisfied. Even so, satisfying these requirements would not be sufficient for successful integration relative to typical research programmes. Disposition

predicates satisfying Mellor's requirements would fail to be integrated in a sense satisfying such research programmes. Perhaps all well-integrated, problem-solving disposition predicates characterize real properties in Mellor's sense. But many problem-raising disposition predicates appear to be real in Mellor's sense as well.

Unlike Mellor, I have neither understanding of nor interest in an 'ontological' distinction between predicates characterizing real properties and predicates that do not. I do recognize methodological distinctions between predicates that play diverse roles in inquiry. For me, the contrast between problem-raising disposition predicates and non-problematic predicates is just such a distinction. Both types of predicates are true or false of things and, in this sense, characterize 'real properties'. But in an another sense, only non-problematic predicates characterize real properties. This distinction cannot be satisfactory to Mellor; for it relativizes real properties to research programmes and the state of knowledge at a given time. Even so, Mellor and I agree that disposition predicates are true or false of things and, in that sense, real. That is enough agreement for my purposes.

Mellor complains that Morgenbesser and I denied that changes in dispositions could be causes or effects. We did indeed deny that changes described using *problem-raising* placeholder disposition predicates could be causes or effects. Perhaps, we should have been slightly more careful. Causal explanations appealing to changes described using placeholders or explaining changes described using placeholders are stopgap explanations. For what it is worth, one can say that changes in dispositions ascribed in such explanations characterize stopgap causes and effects.

As far as explanations invoking problem-solving disposition predicates where the predicates have been well integrated into an explanatorily satisfactory theory are concerned, Levi and Morgenbesser acknowledged that dispositions could be causes or effects. At no point in our discussion did Levi and Morgenbesser suggest that dispositions are not quite real because they are mere potentialities or possibilities. Mere potentialities are abilities. Such abilities are duals of dispositions and as such they are not dispositions. Of course, to concede that changes described using non-problematic disposition predicates can be invoked in satisfactory explanations and, in this sense, can be causes and effects is to exploit a methodological distinction and not an incursion into stormy ontological waters.

Whether dispositions are real properties or not, the thesis of Levi and Morgenbesser was designed to address the schizophrenia that has plagued discussions of dispositions, abilities, potentialities and the like. On the one hand, they are perfectly acceptable theoretical terms and, on the other hand, they make reference to properties that no self-respecting admirer of modern science ought to countenance.

Levi and Morgenbesser drew a distinction between three types of disposition attribution:

(1) problem-raising attributions, where the predicate serves as a placeholder for the purposes of stopgap explanation;
(2) problem-solving attributions, where the predicate has already been integrated into an explanatorily adequate theory or, it is thought, it could be so integrated with minimal and fairly routine inquiry;
(3) mystery-raising attributions, where inquirers have not integrated the disposition predicates into an explanatorily adequate theory and judge the explanations they offer as perfectly satisfactory.

Those who use disposition predicates in a mystery-raising manner are fairly accused of using disposition terms in an inappropriately vacuous manner. Changes in such dispositions can be neither causes nor effects. That is to say, causal explanations invoking such changes are empty. Yet, the emptiness goes unrecognized. This is a typical feature of appeals to occult powers.

Appeal to changes in problem-raising dispositions may also be empty at certain stages of inquiry. However, changes described using problem-raising dispositions are intended for use in stopgap causal explanation. The inquirer who uses such disposition predicates is committed to recognizing the value of inquiry aimed at integrating such predicates into an adequate theory. Such an inquirer need not be absolutely certain that such inquiry will succeed. What is required of the inquirer is that he or she honours the legitimacy and value of pursuing such inquiry.

To the extent to which disposition terms become integrated into an explanatorily adequate theory they cease being placeholders and are to be treated like other theoretical terms. Levi and Morgenbesser (1964) called such terms problem-solving disposition predicates. They could just as well have been called non-dispositional theoretical terms.

Thus, the defects in disposition and ability predicates in scientific inquiry, when such predicates are considered defective, are defects in their use for explanatory purposes. Such defects need not, however, deter us from treating such predicates as being true or false of systems, set-ups or objects. In this sense, we can be as 'realist' as can be about dispositions and abilities.

The defects in the use of placeholders for explanatory purposes do not forbid the use of such placeholders in stopgap explanations. We need not be deterred from using them for purposes of explanation *as long as the stopgap character of the explanation is recognized and taken seriously as relevant to the direction of further inquiry*.

Disposition predicates are used in explanation in an objectionably vacuous manner when there is no recognition of the need for further inquiry aimed at integrating such predicates in an explanatorily more adequate theory.

As an example of such objectionable use consider cases in which the predicate 'is rational' is taken to be a disposition to obey the principles of rational belief, desire and choice for the purpose of explaining human behaviour. Those who think of principles of rationality as explanatory have good reason to favour such a view, for otherwise they lack 'covering laws' to invoke

in developing such explanations. Authors, like Davidson, who adopt this view sometimes insist that there are no psychophysical principles allowing for the reduction of psychology to physics (see Davidson 1980: Ch. 14). This suggests that integrating the predicate 'is rational' in a more comprehensive natural-ized theory is hopeless. Fulfilling the promise to cash out the promissory note is abandoned. We may have to rest content with the vacuous covering laws as Davidson is resigned to do. In my opinion it would be preferable to abandon the view that principles of rationality explain behaviour.

One may well ask what constitutes adequate integration into theory. I have no fixed answer. Of course, the theory should be part of the established body of full beliefs or, perhaps, of the shared beliefs of the community to which the inquirer belongs. It should also fulfil the demands of a research programme to which the inquirer or the community subscribes to some satisfactory degree. In my judgement, the relativity of adequate integration to a research programme precludes the possibility of a standardized characterization of adequate inte-gration. The adoption of a research programme is not the endorsement of the truth of some metaphysical scheme. It is rather the adoption of a value commit-ment that is open to revision in the ongoing activity of inquiry. When inquirers disagree concerning the conditions demanded of explanatory adequacy, the disagreement is often a disagreement about the aims that the given inquiry ought to be promoting, fuelled, perhaps, by information already available about the extent to which realizing a given programme is feasible.

It may, perhaps, be possible to identify some minimal and very weak conditions on the adequacy of explanatory research programmes. Mellor's two conditions on the 'reality' of properties may find a place among such conditions. And, perhaps some day, something may be said about how inquir-ers committed to competing programmes may engage in joint inquiry to iron out differences in their research programmes. But the status of disposition, ability and sample space predications as placeholders relative to research programmes may be supported without identifying a fixed set of conditions of adequacy for explanatory research programmes.

In the case of coin tossing, adopting the characterization of tosses as heads inducing and tails inducing may be helpful as long as we think of such descrip-tions as replaceable in terms of descriptions in classical mechanics. Even if we manage somehow to effect the replacement, we may not be in a position to apply the product to a given case of coin tossing without already knowing the result (because the boundary conditions will not be known without knowing the result).

Perhaps inquirers abandon the replaceability of conceptions like 'heads-inducing toss' by notions of classical mechanics. Perhaps inquirers do not even believe that there is a correct and adequately specifiable system of initial and boundary conditions. The outcome of tossing may be radically indeterministic. In that case, one might still invoke Elster's mechanisms of type A or of type B; but there can be no hope of integrating the placeholders into an adequate theory. In that case, the appeal to mechanisms is vacuous.

When an agent faces unpleasant news, he or she may, as Elster suggests, change desires or change beliefs. The prospects for identifying biologically based conditions that will replace the mechanisms Elster introduces when explaining these different reactions are not great. Perhaps, we might obtain statistical explanation instead by considering the responses of the agent as a matter of chance. But then we still need the sample space property according to which the agent responds to unpleasant news by accepting it, changing desires or changing belief. Here a surefire disposition is attributed to the agent. A stopgap covering-law explanation is offered about why one of these responses occurs. But unless the disposition can be integrated into an explanatorily more adequate theory, the stopgap explanation is a dead end and remains vacuous.

According to the placeholder theory of dispositions, problem-raising dispositions must be characterized by postulates like (1). If the postulate is removed, the placeholding function of the disposition term is abandoned. This is acceptable when the disposition term has been integrated into an explanatorily adequate theory and becomes indistinguishable from any other theoretical term.

If one understands matters in this way, there can be no 'finkish', problem-raising dispositions (Lewis 1999: 133). If the disposition to R upon being S'd is never manifested by object or system x because whenever a trial of kind S is instituted x loses the disposition, the postulate (1) is no longer judged a true lawlike generalization. So the disposition predicate loses its placeholder function. Of course, if the disposition predicate is problem solving, finkishness is no longer precluded. But problem-solving disposition predicates no longer serve a placeholder function. In this respect, they are no longer dispositional except by origin. They are much more like what are known as 'theoretical terms'.

3   Conditionals

Peirce thought that predicating dispositional properties of things is equivalent to asserting so-called 'counterfactual' or 'subjunctive' conditionals. To say that the coin has a surefire disposition to land heads or tails up on a toss is to say that, if it were tossed, it would land heads up or tails up. Of course, supplying truth conditions for disposition statements in terms of conditionals presupposes that conditionals carry truth values. We need truth conditional semantics for conditionals if conditionals are equivalent to disposition and ability statements. And the equivalence seems desirable, for in that way necessary and sufficient truth conditions for disposition and ability statements in terms of (conditional) judgements of test behaviour become available. This line of reasoning drives many erstwhile empiricists into the madness of possible worlds semantics in general and the introduction of some 'closest worlds' semantics for conditionals in particular.

I shall argue that equating disposition statements with conditionals should

be considered unacceptable. My aim is to undermine at least one motive for the madness.

If-sentences are sometimes understood to predicate true of things primitive theoretical predicates characterized by postulates of the forms (1) and (2). As such, these sentences are rephrasals of disposition statements. Dudman (1985) has pointed out that this is especially so when the 'if'-sentences are of the types illustrated by 'If it drops, it breaks', 'if it dropped, it broke'. Dudman calls such if-sentences 'generalizations'. If disposition statements carry truth values, so do such generalizations.

Sentences like 'if it had been dropped, it would have broken', 'if it were dropped, it would break' or 'if it drops, it will break' are quite distinct grammatically from generalizations. Perhaps, in spite of this, they can be used in effective communication with the same understanding as generalizations can. Dudman thinks that as a matter of fact English speakers normally do not do so. I think he is right; but I am not in the business of prohibiting anyone from using subjunctive conditionals or future indicatives as attributions of dispositions no matter how abnormal such usage might be. However, I am under the impression that such conditionals generally express modal judgements of serious possibility and impossibility on a supposition. Conditional sentences of the form 'If x were S'd, x might (would) R' express the modal judgement that x might R (that x would R) *on the supposition made for the sake of the argument that x is S'd*. Here is how I think such judgements ought to be construed.

Let us represent agent X's state of full belief by a deductively closed theory K in some regimented language L. Three cases can be distinguished.

(1)   *The open case*: X is in suspense about whether system a is subject to trial of kind S or not. That is to say, X is committed to recognizing there being a fact of the matter about whether a is subject to a trial of kind S. X's mind concerning this issue is not made up.

(2)   *The belief-contravening case*: X is certain that a trial of kind a has not been conducted on x.

(3)   *The belief-conforming case*: X is certain that a trial of kind S has been conducted on a.

To suppose that a is S'd in the open case is to expand K by adding a sentence expressing that a is subject to a trial of kind S. If the sentence 'a does not R' is inconsistent with the expansion, the modal judgement that a would or must R is made conditional on the supposition. If the sentence 'a does not R' is consistent with the expansion, the modal judgement that a might not R is made conditional on the supposition.

In the belief-contravening case, K is first contracted by removing the claim that a trial of kind S has not been conducted on x. This is to be done so that the loss of informational value is kept at a minimum. The contraction is then treated like the open case. Add the supposition and then make the modal judgement.

In the belief-conforming case, K is first contracted by removing the claim that a trial of kind S has been conducted on *x*. Again loss of informational value should be kept to a minimum. As before, treat the contraction like the open case. Add the supposition and make the modal judgement. There is an alternative account according to which supposition in the belief-conforming case makes modal judgements relative to K itself. The former approach I call *Ramsey Revision* (Levi 1996) and the latter *AGM Revision* after the justly celebrated account of revision in Alchourrón *et al.* (1985).

To repeat, I do not intend to legislate linguistic usage. There are other construals of suppositional reasoning and the associated conditionals on offer. In particular, there is the view that provides closest worlds semantics for conditionals along the lines of Stalnaker (1968) or more impressively of Lewis (1973). Closest worlds approaches are incompatible with Ramsey's approach to open conditionals that I have followed here unless distance between worlds is gerrymandered in a fashion that makes that relation doxastic and subjective.

I agree with Peirce and Mellor that attributions of dispositions and abilities are true or false of the objects of which they are attributed. My concern is to question reasoning from realism about dispositions and abilities to the conclusion that conditional sentences expressing modal judgements on suppositions carry truth values.

If such reasoning were cogent, the epistemic account I have sketched of modal judgement conditional on a supposition based on Ramsey revision would be undermined. Replacing Ramsey revision by AGM revision cannot help. Nor does the use of an epistemized version of closest worlds analysis (invoking imaging transformations of belief-states). Such conditional modal judgement lacks truth value whereas the conditionals allegedly entailed by disposition and ability statements would have truth values. Realism about dispositions would then seem to support the cogency of attempts to supply truth conditions for conditionals as closest worlds semantics does.

My contention is this. Dispositions *in the sense in which it is desirable to allow for a realistic construal of dispositions in scientific inquiry* do *not* entail 'would' conditionals carrying closest worlds truth conditions. Abilities (duals of dispositions) do not entail closest worlds 'might' conditionals.[2] Whatever other applications closest worlds semantics may have, closest worlds semantics does not contribute to the understanding of dispositions.

4   The monotonicity of dispositionality and the non-monotonicity of closest worlds conditionals

Let inquirer X fully believe that object or system *a* has the disposition to R when S'd. X fully believes that D(R/S)*a*. X also fully believes the appropriate instance of schema (1) and, hence, also of the following:

(1T) $(x)[D(R/S)x \supset (Sx \ \& \ Tx \supset Rx)]$.

(1T) together with D(R/S)*a* does not of course entail D(R/S & T)*a*. Even so, since X fully believes D(R/S)*a*, X cannot coherently fully believe that A(∼R/S & T)*a* (which is equivalent to ∼D(R/S & T)*a*). So either X fully believes that D(R/S & T)*a* if X lives up to X's commitments or X suspends judgement on the matter. But suspending judgement in this case does not allow for consistently expanding X's belief-state by concluding that *a* lacks the disposition. To prevent this unattractive result, we require another postulate:

(4)   (*x*)(D(R/S)*x* ⊃ D(R/S & T)*x*.

In spite of the opportunities for equivocation available, the placeholder account of dispositions and abilities makes one thing clear. Disposition terms are introduced as primitives with the Carnapian reduction sentence (1) as postulate in order to provide stopgap universal generalizations for purposes of explanation. This function would be undermined if disposition terms were not 'surefire'. And this feature requires that the disposition not be undermined as trial conditions are strengthened. In this sense, dispositionality is monotonic.

Let us suppose, for the sake of the argument, that D(R/S)*a* entails that if *a* were subject to a test of kind S it would respond in manner R. I am conceding, for the sake of the argument, that such conditionals are truth value bearing. But since D(R/S)*a* entails D(R/S & T)*a* via postulate (4), we must conclude that, if *a* were subject to a test that is both S and T, it would respond in manner R.

However, assuming that conditionals carry truth values (as closest worlds conditionals do), they should be 'variably strict implications', as Lewis (1973) rightly notes. That is to say, they are non-monotonic. Adding more qualifications to the prodosis of the conditional may undermine the conditional. 'If *x* were S & T, *x* might fail to R' might be true even though 'if *x* were S, *x* would R' is true.

Thus, 'If this match were struck, it would light' could be true according to Lewis, whereas 'If this match were struck in the absence of oxygen, it would light' is false.

By way of contrast, either it is false that this match has the surefire disposition to light upon being struck or it is true that this match has the surefire disposition to light upon being struck in the absence of oxygen. I conjecture that most of us think that the match lacks both dispositions. Yet we think that the match has the surefire disposition to light upon being struck under conditions C. We are not ready, however, to spell out minimal conditions C (although, perhaps, we can specify some of them). In spite of our ignorance, we may coherently be convinced that the conditions C are satisfied on some occasion and that the match is disposed to light when struck under conditions C. In lieu of 'conditions C' we might use *ceteris paribus*.

We are then committed, so I submit, to endorsing the view that the match is disposed to light when struck under conditions C by Bill Clinton or under

conditions C when D is true regardless of what D asserts. This is the monotonicity constraint on surefire dispositions.

If K implies that Bill Clinton did not strike the match, supposing that he did might lead to giving up the claim that the match has the surefire disposition to light upon being struck. In that setting, the inquirer might judge that if Bill Clinton struck the match under conditions C it would not light. Yet, K implies that the match has the surefire disposition to light when struck and via (4) to light when struck by Bill Clinton. So even if the inquiring agent is convinced that the object has the disposition to light upon being struck under conditions C by Bill Clinton, the closest-world conditional 'if the match were struck under conditions C by Bill Clinton, it would light' would be judged false.

Lewis himself has acknowledged that attributions of dispositions may fail to entail the corresponding would-conditionals in cases where dispositions go finkish (Lewis 1999: 133). In the example given above, the disposition is not finkish. Bill Clinton's striking the match under the conditions C does not cause the match to lose its surefire disposition to light upon being struck by Clinton under conditions C. The inquirer is certain that Bill Clinton was not the one who struck the match. So when the inquirer supposes that Clinton performed the act, the inquirer might be led to abandon the claim that the match does have the surefire disposition in question without thinking that the match would have lost a disposition it had beforehand. There is no change in disposition implied.

In any case, finkishness can arise only when disposition predicates have been well integrated into theory and can be treated as non-dispositional theoretical terms. When disposition predicates are serving their placeholder functions, finkishness, as noted before, cannot arise. But even in cases where there is no finkishness, disposition predicates cannot entail would-conditionals as specified according to closest worlds semantics.

5   The non-monotonicity of ability and the centring condition

Suppose that coin x has the ability to land heads on a toss. But it lacks the ability to land heads on a toss by Morgenbesser. That is to say, it has a surefire disposition to *fail* to land heads on a toss by Morgenbesser.

Let it be true that coin x is tossed by Morgenbesser and lands tails. The fact that this happens does not undermine the fact that at the time of this occurrence the coin had the ability to land heads on a toss.

Shall we say that the ability of the coin x entails the judgement that, if the coin were tossed, it might have landed heads? That is to say (according to closest worlds analysts), in at least one closest world to the actual world that is a coin-tossing world, the coin lands heads. Given that the actual world is a coin-tossing world where the coin is tossed by Morgenbesser and the coin lands tails, the former condition cannot hold. According to Lewis's 'centring condition', when the closest world to the actual world in which the coin is

tossed by Morgenbesser is the actual world, that world is the uniquely closest such world.

Abandoning the centring condition does not alter the situation. No matter what requirement on the nearness of possible worlds to the actual world is imposed, either the set of closest worlds includes at least one case where the coin lands heads or no such cases. According to the first alternative, the 'might land heads' conditional is true and the 'would land tails' conditional is false. According to the second alternative, the 'might' conditional is false and the 'would' conditional is true. But *both* the conditionals should be true if the corresponding disposition and ability statements are true and Morgenbesser tosses the coin. Lewis's theory (and other closest worlds and selection function theories) must do one of two things:

(1)   Declare the following to be an inconsistent triad: (a) that the coin is able to land heads on a toss, (b) that the coin is constrained to land heads on a toss by Morgenbesser and (c) that the coin is tossed by Morgenbesser.
(2)   Recognize the consistency of the triad but deny that that disposition statements are equivalent to conditionals.

Insisting that the ability and disposition attributions together with the claim that Morgenbesser tosses the coin form an inconsistent triad is untenable. That is to say, it is untenable if the attribution of the ability to land heads on a toss to the coin is to be neutral with respect to whether the process of coin tossing can be correctly redescribed according to a deterministic model.

James Bernoulli (1713) insisted in *Ars Conjectandi* that there can be neither objective possibility nor objective probability if there is objective necessity, i.e. determinism. One of the concerns of those introducing notions of objective or statistical probability in the nineteenth century was to characterize objective probability so that its use in characterizing macroprocesses could be neutral with respect to whether these macroprocesses might be described microscopically in a neutral fashion.

To relativize probability attributions to kinds of trials was critical to such views. And this called for relativizing attributions of dispositions and abilities to kinds of trials as well. We may coherently acknowledge that coin *a* has the ability even if we also think that situating the coin in a certain mechanical state and subject to appropriate boundary conditions constrains it to land tails and believe that the coin is so situated upon being tossed by Morgenbesser. If we think that the best contemporary physical theory is not deterministic, we can also allow for the ability. Attributing the ability is neutral with respect to underlying determinism.

Relativizing ability and disposition attributions to kinds of trials allows for this kind of neutrality, however, only if triads of the kind illustrated above are not incoherent. Equating disposition and ability statements with corresponding closest worlds conditionals requires that such triads be inconsistent. So much the worse for the equation.

6   Dispositions support but do not entail conditionals

Since dispositions neither entail nor are entailed by truth value-bearing closest-world conditionals, appeal to the truth value-bearing status of disposition statements cannot be sufficient to argue for the truth value-bearing status of conditional modal judgements.

This is fortunate. Conditional modal judgements (modal judgements conditional on suppositions) of possibility, like unconditional judgements of possibility, cannot carry truth values for reasons quite different from what has been considered thus far.

If conditional modal judgements could carry truth values, it would be coherent to suspend judgement concerning their truth or falsity, to judge them probable to varying degrees, to desire that they be true in varying degrees and the like. There are, in my opinion, powerful arguments for calling such coherence into question. Some of these arguments are surveyed in Arló Costa and Levi (1996) and Levi (1996) and it is fortunate that the kind of realism about dispositions required by scientific practice does not call for a different verdict.

Indeed, it is still open to us to admit that *full belief* that *a* is disposed to R upon being S'd *supports* but does not *entail* the judgement that it is not a serious possibility that *a* fails to R on the supposition that *a* is S'd. The supported judgement of serious possibility is neither true nor false.

Consider then the judgement that *a* might fail to R on the supposition that *a* is S & T'd. This judgement is consistent with the full belief that *a* has the surefire disposition to R upon being S'd provided that *a* is S'd and T'd licenses giving up the background assumption that *a* has the surefire disposition in question. Such a licence is not available if the disposition statement entails the closest-worlds conditional.

On the other hand, consider the full belief that it is possible for the coin to land heads on a toss, the claim that it is not possible for the coin to land heads on a toss by Morgenbesser and the further claim that the coin is tossed by Morgenbesser. This system of beliefs supports the following bits of suppositional reasoning:

(1)   On the supposition that the coin is tossed, it may land heads.
(2)   On the supposition that the coin is tossed by Morgenbesser, it will land tails.

In case (1), supposing that the coin is tossed (where it is already believed that it is tossed by Morgenbesser) calls for 'contracting' by giving up the claim that the coin is tossed. This will plausibly call also for giving up that either it is not tossed or tossed by Morgenbesser. Hence, restoring the supposition that the coin is tossed will not return the claim that it is tossed by Morgenbesser. On this basis, the conditional (1) will be supported.

On the other hand, in case (2), supposing that the coin is tossed by Mor-

genbesser will require supposing not only that the coin is tossed but that it is tossed by Morgenbesser.

These results obtain if one uses the account of supposition based on Ramsey revision. Supposition based on AGM revision cannot provide this result. Nor can supposition based on imaging or closest worlds revision.

The fact that supposition based on Ramsey revision can adequately avoid the two difficulties confronting the equation of disposition statements with truth value-bearing conditionals suggests, so I think, that there should be no pressure to invoke truth value-bearing conditionals for the sake of understanding dispositions. I suspect that it might also relieve the pressure on overblown efforts to understand causal attributions in terms of conditionals.

7   Are dispositions real?

Throughout this discussion, I have maintained that there is indeed a fact of the matter of whether attributions of dispositions to things are true or false. We can coherently be unsure whether a glass is fragile or not and can even judge the hypothesis of fragility to be probable to some degree. In these respects, dispositions are, so I claim, as real as can be.

On the other hand, whether the additional demands imposed on the 'reality' of properties are those favoured by Mellor or are the requirements of some research programme for explanation, the placeholder view suggests that dispositions are real only when their placeholding mission has been accomplished and they are no longer problem raising. The reality of dispositions is a work in progress. That, I believe, is the crux of my disagreement with Mellor.

Whatever the status of dispositions, belief that they are true or false of things supports conditionals. But such support does not presuppose that conditionals carry truth values. Thus, Isaac Newton's use of inductions on the acceleration fields induced by central bodies to draw conclusions about the system of the world may presuppose some modest realism concerning dispositions and abilities but should not offer authority to semantics supplying truth conditions for conditionals as illustrated by closest worlds analysis.

An elaboration of the 'logic' of conditionals based on Ramsey revision is beyond the scope of this chapter, as is an account of iterated conditionals.[3] My aim here has been to point out that one can coherently endorse a form of minimal realism for dispositions without indulging in the excesses of possible worlds semantics for conditionals.

No doubt the realism about conditionals I endorse is not robust enough to satisfy Mellor's requirements. I hope, however, that these remarks testify to the extent to which I have sought to accommodate his realist insights within the framework of my own irremediably methodological and pragmatist approach. More important yet, I hope they testify to the respect I have for him as a friend and for his philosophical contributions.

Notes

1 For a discussion of the distinction between certainty and incorrigibility, see Levi (1980: Ch 1; 1991; 1996).
2 Assuming a closest worlds semantics for 'would' and 'might' conditionals, the thesis that dispositions entail the corresponding closest worlds 'would' conditionals and abilities entail the corresponding 'might' conditionals holds if and only if disposition statements are equivalent to the corresponding 'would' conditionals.
3 See Levi (1988; 1996: Chs 3 and 4; 1998).

References

Arló Costa, H. and Levi, I. (1996) 'Two notions of epistemic validity', *Synthese* 109: 217–62.

Alchourrón, C., Gärdenfors, P. and Makinson, D. (1985) 'On the theory of logic change: partial meet functions for contraction and revision', *Journal of Symbolic Logic* 50: 510–30.

Bernoulli, J. (1713) *Ars Conjectandi*, Bing Sung (trans.), Department of Statistics Technical Report 3, Cambridge, MA: Harvard University.

Carnap, R. (1950) *Testability and Meaning*, reprint by the Graduate Philosophy Club, Yale University, with corrections and additional Bibliography of the paper published in *Philosophy of Science* Vol. 3 (1936) and Vol. 4 (1937).

Davidson, E. (1980) *Essays on Actions and Events*, Oxford: Oxford University Press.

Dudman, V. H. (1985) 'Towards a theory of predication in English', *Australasian Journal of Linguistics* 5: 143–93.

Elster, J. (1999) *Alchemies of the Mind*, Cambridge, UK: Cambridge University Press.

Gärdenfors, P. (1988) *Knowledge in Flux*, Cambridge, MA: MIT Press.

Levi, I. (1980) *The Enterprise of Knowledge*, Cambridge, MA: MIT Press.

—— (1988) 'Iteration of conditionals and the Ramsey Test', *Synthese* 76: 49–81.

—— (1991) *The Fixation of Belief*, Cambridge, UK: Cambridge University Press.

—— (1996) *For the Sake of the Argument*, Cambridge, UK: Cambridge University Press.

—— (1998) 'Contraction and informational value', http://columbia.edu/~levi.

Levi, I. and Morgenbesser, S. (1964) 'Belief and disposition', *American Philosophical Quarterly* 1: 221–32.

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

—— (1999) *Papers in Metaphysics and Epistemology*, Cambridge, UK: Cambridge University Press.

Mellor, D. H. (1974) 'In defense of dispositions', *Philosophical Review* 83: 157–81.

—— (2000) 'The semantics and ontology of dispositions', *Mind* 109: 757–80.

Stalnaker, R. C. (1968) 'A theory of conditionals', in *Studies in Logical Theory*, Oxford: Basil Blackwell.

10 Structural properties

Alexander Bird

1 Introduction

Dispositional essentialists claim that dispositional properties are essentially dispositional: a property would not be the property it is unless it carried with it certain dispositional powers. Categoricalists about dispositional properties deny this, asserting that the same properties might have had different dispositional powers had the contingent laws of nature been otherwise.

As I have described it, that debate concerns properties that can be characterized as dispositional. We could expand that debate to include another one. How many different metaphysical kinds of property are there? Just one, or two or more? The monist thinks that there is just one kind of property. The categoricalist described above is likely to be a monist, asserting that all properties are categorical in nature. On this view, all properties are alike in essence; they confer, of themselves alone, no potentialities, no causal powers. A (categorical) property can confer such powers, but only because there is a law relating that property to some other property. Armstrong is a categorical monist (Armstrong *et al.* 1996: 15–18; Armstrong 1997: 69, 80–3). Another kind of monist thinks that the distinction between different kinds of property is misconceived, and that dispositionality and categoricity are different aspects of one kind of property. Martin and Mumford have expressed this sort of view (Armstrong *et al.* 1996: 71–5; Mumford 1998: 64–7). A dualist may think that the distinction is well conceived and that some properties are categorical (i.e. are just as the categorical monist thinks all properties are), whereas some others are essentially dispositional. One could, perhaps, be a more liberal pluralist, thinking that substance and kind properties (being gold, being a tiger) and mathematical properties (being odd, being well founded) are yet different kinds of property, being neither dispositional nor categorical. Dualists and other pluralists may be egalitarian – none of the different kinds of property has any special priority relative to the others. Or they may be hierarchical, holding that one kind of property (the categorical, for example) explains or is the basis for the other kind(s).

In this chapter I wish to examine the prospects for *dispositional monism*. This view is monistic in that it holds that there is only one kind of property,

or, more circumspectly, that there is only one kind of property in the metaphysics of science. (The properties I am discussing in this essay are Lewis's sparse properties; the dispositional monist need not account for non-sparse abundant properties.) But, in a mirror image of categorical monism, dispositional monism asserts that all properties are essentially dispositional. None is categorical.

This view faces severe challenges on more than one front. For example, dispositional essentialism is committed to the metaphysical necessity of the laws of nature. If some property D is *essentially* the disposition to manifest M whenever stimulated by S, then the conditional $(Dx \wedge Sx) \rightarrow Mx$ (possibly with an added *ceteris paribus* clause) is necessarily true. Our intuitions are that the laws of nature are contingent; our intuitions thus favour categoricalism about properties. The correct response to this challenge is simply to deny the dialectical force of intuition in this case. Our intuitions concerning necessity are notoriously unreliable, as Kripke has shown. Furthermore, it can be proved that even the categoricalist must accept that some apparently contingent higher level laws are in fact necessary.[1]

2   Structural properties

In the following I shall examine another, perhaps stiffer, challenge which is presented by properties that seem not to be dispositional at all and are held up as paradigms of categorical properties. These are structural, typically geometrical, properties. Take the science of crystallography. The explanation of the properties of a crystal will refer to its structure, which is a matter of the geometrical relations of the ions or molecules that constitute the crystal.[2] Since spatial relations are structural in the current sense, all sciences will depend on structural properties. A categoricalist might think that an object that consists of a set of masses in a particular spatial configuration has just been described in purely categorical terms. Whether the dispositionalist can account for mass is a question to be pursued elsewhere.[3] The present, greater challenge is to account for the spatial relations in dispositional terms.

Being triangular, for example, seems to bring with it no powers in the way that, say, being elastic or being negatively charged does. This is why those who think that *some* properties are essentially dispositional might be inclined to be dualists, permitting structural properties to be categorical. If one is inclined to be a dispositionalist across the board, how might one defend the claim that structural properties are, despite appearances, dispositional also? As in the objection concerning the alleged contingency of laws, the dispositional monist must argue that appearances are deceptive. It is not the case that structural properties do not confer powers necessarily.

'Conferring a power' has traditionally been cashed out in terms of entailing a counterfactual or subjunctive conditional. If that is appropriate, then the categoricalist challenge is committed to the following necessary condition on being a dispositional property:

(A)   P is a dispositional property only if for some S and M and for all *x*:

Px entails if Sx were the case, then Mx would be the case.

The categoricalist about structural properties argues that structural property ascriptions entail no such conditionals:

(S)   If P is a structural property then there are no S and M such that for all *x*:

Px entails if Sx were the case, then Mx would be the case.

(A) and (S) together entail that structural properties are not dispositional properties.

For example, if '*x* is triangular' entails no non-trivial subjunctive conditional then, given (A), triangularity is not dispositional. If on the contrary there is a sound argument that '*x* is triangular' does entail some such conditional, i.e. that (S) is false, then the categoricalist will have failed to show that triangularity is non-dispositional, and so triangularity cannot be employed as a counterexample to the claim that all properties are dispositional.

Even so, such an argument showing that (S) is false would not have shown that triangularity *is* dispositional. For that we would need the reverse of (A) to be true:

(B)   P is a dispositional property if for some S and M, and for all *x*:

Px entails if Sx were the case, then Mx would be the case.

The conjunction of (A) and (B) yields a biconditional that is the so-called conditional analysis of dispositions:

(CA) P is a dispositional property if and only if for some S and M, and for all x:

Px entails if Sx were the case, then Mx would be the case.

3   A contest

Hugh Mellor argues that the ascription of triangularity does entail a subjunctive conditional, and hence that (S) is false for triangularity (and so (A) cannot be employed to show that triangularity is non-dispositional) (Mellor 1974). If his claim can be made to stick, then Mellor's argument may be used by the dispositional monist against the attack based on structural properties. In rooting for Mellor the dispositional monist will decry his oer her opponent in the ensuing debate, Elizabeth Prior, who argues that Mellor's

alleged entailment does not hold (Prior 1982). (It should be pointed out that Mellor's aim is not to defend dispositional monism; rather it is to undermine the prejudice against dispositions that says they are not real or that if they are real that is only because they are identical to or supported by a basis composed entirely of categorical properties. As we shall see, Mellor's aim does not entail dispositional monism; but it is congenial to it. Prior is herself a dualist but a hierarchical one – categorical properties form the causal basis for dispositional properties.)

In what follows I shall see where this debate leads. Ultimately, I shall argue, it shows how a dispositional monist can indeed mount a satisfactory defence and can account for structural properties – although not in quite the way initially suggested by Mellor. This is an outcome with which the dispositional monist can be happy, since structural properties are prima facie a counterexample to their position. However, I will not be arguing here that structural properties *must* be accounted for as dispositions. In that sense the outcome will be a draw, in that both the dispositionalist and the categoricalist have, as far as the debate surrounding their relation to conditionals is concerned, satisfactory accounts of structural properties.

4 Dispositions and conditionals

However, before we look at that debate, we need to note that success for Mellor is, as it stands, not after all even a necessary condition for the truth of dispositional monism. Nor, for that matter, is it a sufficient condition for a successful dispositional account of structural properties. Success would be both a necessary and a sufficient condition on showing that structural properties are dispositional if the conditional analysis were true. But the conditional analysis is false – in both directions of the biconditional in (CA).

Given (A), success for Mellor is necessary for the truth of dispositional monism in this sense. There must be some conditional entailed by '*x* is triangular'; if there is not, triangularity is a counterexample. (Of course, Mellor's particular conditional might not be the right one – we shall return to this.) But if (A) is false, then the non-existence of such a conditional will not show that triangularity is non-dispositional. And (A) is indeed false, as is shown by the possibility of finks and antidotes (Martin 1994; Lewis 1997; Bird 1998).

Finkish dispositions are those which cease to exist upon the instantiation of the disposition's characteristic stimulus. Since the disposition ceases to exist, the manifestation is not brought about. So at some time when there is no stimulus event, the disposition exists. But the counterfactual 'were the stimulus to occur, the manifestation would follow' is not true. Lewis (1997) gives the following example. A sorcerer wants to protect a favourite but very fragile vase from breaking. His method of protection is to cast a spell that almost instantaneously changes the structure and material of the vase in such a way that it is no longer fragile, whenever (but only when) it is struck, dropped, etc. So the vase will not break when struck even though it is very fragile. Finks

operate by changing the disposition (or its intrinsic causal basis). But a disposition may depend for its characteristic functioning not only on the causal base that is intrinsic to its possessor but also upon properties of the environment; it may depend upon properties of the possessor of the disposition that are not part of the disposition's causal basis (such as properties acquired after the possessor has received the stimulus of the disposition – an example will make this clear shortly). An *antidote* works so as to interfere with the role of these other properties in the operation of the disposition (Bird 1998). An antidote to a poison may work either by changing the patient's physiology so that the poison cannot do the damage it normally does or by repairing the damage done before it can result in illness or death. In such cases the antidote to the poison is an antidote in my sense, since it changes the environmental conditions required for the poison to do harm. An antidote to a poison might also work by reacting with the poison before it can affect the patient. In this case, the poison's disposition to cause illness or death is a finkish one, and the antidote is not strictly an antidote in my sense. When a disposition receives its normal stimulus but in the presence of an antidote, the normal manifestation will fail to occur. Hence we have a disposition without the corresponding conditional. Changing Lewis's example, the sorcerer might alternatively decide to protect his vase by instructing a demon to repair any cracks that appear in the vase at lightning speed. So although striking the vase leads to cracks appearing in the vase as normal, these are repaired before they can join up, so preventing the vase from breaking. The normal functioning of fragility in causing breaking requires the cracks, which are properties of the vase, to remain; the antidote in this case works by changing properties of the possessor (rather than of the environment) that are brought about by the stimulus.

So it looks as if finks and antidotes make life more comfortable for the dispositional monist. By showing (A) to be false they seem to undermine the possibility of counterexamples before they get off the ground. On the other hand, showing that '*x* is triangular' does entail a conditional is not sufficient [*pace* (B)] to prove the dispositional monist right, for two reasons. The first reason mirrors the problem of finks and antidotes. Parallel arguments show the falsity of (B).[4] Finkishness can operate in reverse, so that an event S causes a disposition to come into existence and to yield its manifestation M. So, just before that event, the conditional 'if S were to occur, then M would occur' is true, but at that moment there is no disposition. Similarly, environmental conditions can conspire to make a conditional true without there being any disposition in the offing; this is a *mimic*.[5] A trivial case of this concerns any two actual facts p and q. According to Lewis, 'if p were true, then q would be true' is true. But we do not think that any two actual facts are conjoined dispositionally. Mimics and reverse finkishness are counterexamples to the equivalence of dispositional statements and counterfactual or subjunctive conditionals, because they show that the conditional does not entail the existence of the disposition.

The second reason why success for Mellor does not entail the dispositionalist view is rather different. The categoricalist *can* endorse the claim that *some* statements asserting the instantiation of a non-dispositional property do entail a conditional. The categoricalist acknowledges that there are dispositional property terms, such as 'elastic', 'irascible' and so forth. The meanings of these terms, says the categoricalist, may well be conveyed by subjunctive conditionals. Hence there might well be some conditional C such that '*x* is elastic' entails C. But that will not show that the property we call 'elasticity' is essentially dispositional. The categoricalist view of dispositions is that elasticity is the name given to a certain categorical property complex in virtue of the fact that, in this world, with this world's laws, that property causes its possessor to stretch, temporarily, rather than break or deform permanently, when subjected to a moderate force. That is consistent with its being the case that the same property complex would not have that effect in a world with different laws. Putting things another way, the categorical monist can be happy with the thought that there are two kinds of predicate, categorical and dispositional, and that the difference between them turns on whether there is an analytic relationship between the predicate and a subjunctive conditional.

5   Rules of the contest

Should the conclusion of the previous section be that the truth of subjunctive conditionals is a red herring as regards dispositionality? No, I think not – but we should be careful. There is, as Martin (1994) has said, clearly some sort of connection between dispositions and conditionals, even if it is not one of straightforward entailment (Armstrong *et al*. 1996: 178). So we can still follow the debate, only we must umpire the debate by forbidding moves that exploit the differences between conditionals and dispositions discussed above. One side in the contest seeks to show that a property is dispositional by showing that it possesses an intimate link (that falls short of outright entailment) to a characteristic conditional; the other side will deny such a link. The contest is governed by two rules:

> Rule 1: Any link established between a property and a conditional must be a metaphysical rather than a merely analytic one.

> Rule 2: The existence of a link between a property and a conditional may not be refuted by appeal to finks or antidotes (or established by appeal to finks or mimics).[6]

With these rules in place and with careful umpiring to see that they are respected, we may still have an informative debate centred on the existence or otherwise of a relation between properties and conditionals.

6   Let the games commence …

The challenge to the dispositional monist is the claim that geometrical shape entails nothing as regards counterfactual or subjunctive conditionals. There is no C, it is asserted, such that C is a genuine, non-trivial, modal conditional and '*x* is triangular' entails C. Mellor states that there is just such a C. His candidate is 'if someone were to count *x*'s corners correctly, then the result would be 3', which, he says, is entailed by '*x* is triangular'. Hence triangularity is at least no proven counterexample to dispositional monism – and (S) is false for triangularity. And to the extent that (B) can be relied upon, triangularity is shown to be dispositional.

The subsequent debate hinged on the interpretation of 'correctly'. Prior held that Mellor's claim acquired prima facie plausibility only because of the use of this word. For without it we would see that the entailment does not hold – people frequently count things and get the wrong answer. More significantly, we are entitled to consider another possible world in which the laws of nature are different so that its inhabitants make systematic errors in counting. (Prior suggests perceptual errors, but one could imagine deeper neurophysiological interference also.) The inclusion of 'correctly' is significant because it seems to rule out such cases. But, says Prior, it does so only because we take the claim that a task was carried out 'correctly' as meaning that it was performed successfully, that it got the right result. Since it is analytic that triangles have three corners, it is also analytic that someone who counts the corners of a triangle correctly gets the answer 3. And so the entailment does not seem to reflect the metaphysics of the property of being a triangle. Rather it depends only on analytic relations and so Mellor's argument falls foul of Rule 1.

Prior (1982) notes that Mellor states in a footnote that 'correctly' is intended to refer not to the result of counting but rather to the manner of counting. But she thinks that, if this is so, then the entailment fails, since if it is only the manner of counting that is invoked, then counting in the unusual world with systematic error may be carried out in the correct manner without getting the correct result.

Prior has a second argument that invokes a different unusual world, in which the laws of nature are such that, when one starts to count the corners or a triangular object, the object is caused to change the number of corners it has. Hence, if one counts as well as one can one will get an answer other than 3.

What defence has Mellor against these two objections? He does not address the second. But he does not need to. The umpire rules out this objection as a foul – it is a clear contravention of Rule 2, since in the world considered, triangularity is finkish, in that the stimulus, counting, causes an object to lose its triangularity. As regards the first objection, here the accusation is that it is Mellor who has broken the rules.

Mellor responds that he can spell out precisely what counting correctly is without referring to the correctness of the result: it is to count the items in question once each (and once only), which is to put them in a '1–1 correspond-

ence with an initial segment of the sequence of positive integers 1, 2, 3 … The highest number in the segment is the result of the counting' (Mellor 1982: 97). Does this reply block Prior's appeal to Rule 1?

Let us compare 'x is even', which entails 'if x were to be divided by 2, then the result would be an integer'. On one understanding, where dividing is understood as an abstract mathematical operation, this is clearly true. Does this make 'being even' a dispositional property? If so, it would be difficult to deny that being triangular or any other property is dispositional. If Mellor's claim is understood analogously, with 'counting correctly' taken to be an abstract mathematical operation, it might well be regarded as analytically trivial, and so outlawed by Rule 1. It is analytic that the set of corners of any triangle has three members. It is analytic that, when any three-membered set is put into 1–1 correspondence with an initial segment of the positive integers, the highest number in the segment is 3. So Mellor's entailment is analytic. But is it *merely* analytic?

We need a test for the application of Rule 1, a test that distinguishes a merely analytic entailment from one that reflects the metaphysics of the entities involved. The test is this: if the entailment is not merely analytic it should continue to hold when we employ any rigid designator to pick out the entity in question. So 'S is the inventor of bifocals *entails* S invented bifocals' is a merely analytic entailment, since 'S is Benjamin Franklin' does not entail 'S invented bifocals'. While even if one thought that being H_2O is part of the definition of water, 'x is water *entails* x is H_2O' would not be merely analytic, since, for example, 'x is that substance which, in the actual world, is the main component of living things on earth *entails* x is H_2O' is also true (but not analytic).

By this test Mellor's entailment will not come out as merely analytical, since for any rigid designator 'D' that picks out the property of triangularity, 'x is D' entails 'if someone were to count x's corners correctly, then the result would be 3' (where 'counting correctly' is still understood abstractly). Yet we should note that the efficacy of the test depends on the difference in modal properties between definite descriptions and rigid designators. But there is no such difference between mathematical definite descriptions and corresponding rigid designators. So the test does not seem to applicable here, and it is not clear that Mellor's entailment does not infringe Rule 1.

However, a different reason for dismissing Mellor's claim, on this understanding, is that it is in conflict with the thought that the stimulus of a disposition is a cause of the manifestation – dropping the fragile vase *caused* it to break, pulling the elastic *caused* it to stretch, and so forth. Although this is contentious in the eyes of some, we could add a third rule. Rule 3 would state that there must be a causal or nomic connection between the antecedent of the conditional and its consequent. Mellor's claim would outlawed by Rule 3 on the mathematical interpretation.

On the other hand, we might understand the dividing as an intellectual, psychological operation, not as an abstract mathematical one. This allows the stimulus (i.e. dividing) to cause the manifestation (getting an integer as the

answer). If we regard the process of counting the corners of the triangle in this way, then Mellor's claim looks to be a substantial one. However, we might then ask, can we be sure that his entailment holds under this interpretation? Someone batting for Prior's team could argue as follows that it does not. For now there is a gap between the fact of the corners of the triangle having been correlated with the set of numbers $\{1, 2, 3\}$ and the fact of the subject's being in the mental state of getting the answer 3. In normal cases this gap is traversed without difficulty. But in unusual cases it need not be. Where environmental conditions or the laws of neurophysiology are different, the counting may have been carried out correctly, the appropriate correlations having been made, yet the answer achieved is a number other than 3. For example, we may imagine a 'killer triangle' whose particular size and angles interact with a subject's neurophysiology to kill them or to cause mental aberration. More directly, we could take the case of a triangle painted killer yellow.

Hence the conditional is not entailed by the ascription of triangularity. However, this does not prove that Prior is right. We already know that in general disposition ascriptions do *not* entail the corresponding conditional, because of finks and antidotes. We saw that Prior's case of a world where triangles ceased to be triangles when counting began is an invocation of a fink. The cases considered in the previous paragraph do not invoke finks (the triangles remain the same), but they do involve antidotes, since they interfere with the normal operation of the stimulus. Hence Prior's moves break Rule 2 again.

As we shall see, the debate is by no means concluded. Nonetheless, after the first innings it looks as if Mellor has the upper hand, just, and that triangularity is no less related to its conditional than dispositions in general are related to theirs.

7   Will the real disposition please stand up?

Even so, there is still all to play for. The problem next bowled at Mellor is the thought that, although the entailment discussed seems to indicate a disposition, it is not, on reflection, clear that the disposition lies with the triangle rather than the counter. Consider the following:

> *x* is a normal observer *entails* if *x* were to count the corners of a triangle correctly then *x* would get the answer 3.

Modulo finks and antidotes, this seems to be true. Given the link between dispositions and conditionals upon which we have been trading, this suggests that being a normal observer is dispositional, which is plausible enough. Note, however, that this entailment is equivalent to Mellor's entailment, if finks are excluded. So it look as if we have two dispositions for the price of one. Which is the real disposition?

That said, it is not clear that we have to choose between the two dispositions. Indeed, it might be perfectly correct to accept both. Martin (Armstrong *et al.*

1996: 135–6) has pointed out that dispositions frequently come in pairs of reciprocal disposition partners. The negatively charged electron is disposed to attract the positively charged proton; the proton is disposed to be attracted to the electron (and also to attract the electron towards it). In fact, our discussion suggests that dispositions might always come in reciprocal pairs. For the following are, in general, equivalent (again in the absence of finks):

> X *entails* were it the case that Y, then Z would be the case; and

> Y *entails* were it the case that X, then Z would be the case.[7]

So it seems too hasty, simply because there is dispositionality in the subject (the counter), to exclude triangularity from genuine dispositionality. However, the resulting position remains unsatisfactory from the points of view of both the categoricalist and the dispositional monist. On the one hand, the dispositional reciprocity between the triangle and the observer that is suggested by Mellor's account makes triangularity look like a secondary property, akin to a colour. But there is a clear disanalogy between structural properties like triangularity and secondary properties like colour, in that the latter have an explanatory role only in a limited portion of science, primarily the behavioural sciences. That is as it should be, since the manifestations of colours and all other secondary properties are the mental states of sentient observers. Yet structural properties play a role at the most general and basic level in science. And their doing so is independent of any power to produce effects in human observers. This does not show but does suggest that the reciprocity between triangle and observer is one sided, that the dispositionality comes primarily or even completely from the observer and not from the triangle.

On the other hand, the same line of reasoning will suggest to the dispositional monist that Mellor has not shown which disposition triangularity is. The existence of a property may be related to all sorts of conditionals. But not all of them reflect the nature of that property. In this case the conditional in question seems to make being triangular a secondary property, a property whose nature is to be a disposition to cause a certain effect in a human observer. One can deny that triangularity is a secondary property without asserting that it is a categorical (primary) property. Perhaps it should be understood as a genuinely tertiary property, one that is a disposition which is manifested not in human subjects especially but in some other, broader, class of entities, a class specifiable at a more general level in science.

In so far as we are still employing subjunctive conditionals as a sign of dispositionality, we should look for a conditional that reflects the nature of the (alleged) disposition, and a sign of this will be that the stimulus and manifestation reflect the role of the property in scientific explanation. In effect, both sides should accept this as Rule 4. Triangles may exist in pretty well any possible world that has a physical component. It would be odd, if triangularity is a dispositional property, that it should be one whose dispositional nature,

if it can be specified, is specifiable only in terms of entities (things that can count) that exist at a very limited range of possible worlds. Rule 4 says that if triangularity is to be shown to be genuinely dispositional, we should look for a conditional characterization that has appropriate generality.

8   Properties and geometries

I believe that there are conditionals for structural properties that come much closer to obeying Rule 4 than Mellor's. For example, the following:

> if two entities travelling at constant speed were simultaneously emitted from A, one along the line AC and the other along the line AB, where it is reflected along the line BC, the former will reach C first

is a conditional that, at first sight, seems to be entailed, barring finks and antidotes, by the proposition that ABC is an triangle.

The problem with this proposal is that the entailment suggested does not hold after all. The conditional is true in worlds where the geometry of space is Euclidean, but may not hold in worlds where the geometry is, for example, Riemannian. But that suggests a position for the dispositionalist. What the latter will hold true is (again barring finks and antidotes):

> ABC is an Euclidean triangle

> *entails*

> if two entities travelling at constant speed were simultaneously emitted from A, one along the line AC and the other along the line AB, where it is reflected along the line BC, the former will reach C first.

Different subjunctive conditionals will be made true by Riemannian triangles, and triangles in Lobatchevsky–Bolyai geometry, and other kinds of geometry. So we have lots of different kinds of triangle property, each of which is dispositional. Space in Riemannian geometry has uniform positive curvature, whereas in Lobatchevsky–Bolyai geometry it has uniform negative curvature; strictly, we might expect a different property of triangularity for each degree of curvature.

So what of triangularity in general? What sort of property is that? Is it dispositional? The dispositionalist will deny that it is a dispositional property. It may nonetheless be a property in some acceptable but more general sense of 'property'. The dispositionalist is not required to account for everything we call a property. Rather, the dispositionalist is required to account only for *sparse* properties. Abundant properties, which correspond (more or less) to predicates, form a much wider class that will include non-dispositional properties. So the dispositionalist's position will be that whereas 'being a Euclidean triangle',

'being a Riemannian triangle with curvature *r*' and so forth may denote sparse properties, 'being a triangle' denotes only an abundant property. 'Being a triangle' is a generalization of the specific, sparse triangularity properties. Since the different triangularity properties do not have any dispositional powers in common, 'being a triangle' is not a dispositional property and no characteristic subjunctive conditionals are entailed by the fact of possessing it. This is no challenge to the dispositional monist, since there is no reason to take the general property of triangularity to be a sparse property.

We might bypass the question of whether any triangularity property, general or specific, is really a sparse property by asking about the dispositionality of spacetime itself. Clearly, the possibility of an explanation that invokes *x*'s triangularity supervenes on the spatio-temporal arrangements of *x*'s parts. This does not show that the supervening property is not a genuine sparse property. But if we were content that the subvening properties are all dispositional, we need not exercise ourselves so greatly over the status of the supervening ones. The set-up that is often invoked as exemplifying categorical but not dispositional properties is a set of masses arrayed in spacetime. The lesson of general relativity is just that we may see the components of this set-up as dispositional. Each spacetime point is characterized by its dynamic properties, i.e. its disposition to affect the kinetic properties of an object at that point, captured in the gravitational field tensor at that point. The mass of each object is its disposition to change the curvature of spacetime, that is to change the dynamic properties of each spacetime point.[8] Hence all the relevant explanatory properties in this set-up may be characterized dispositionally.

Before concluding that the dispositionalist has succeeded in defending a dispositional view of structural properties, we should ensure that the entailments being appealed to do obey all the rules laid down. Rule 4 requires an appropriate level of generality. We moved our attention from Mellor's entailments to these ones precisely to achieve that. Rule 2 required that we do not appeal to finks or antidotes or mimics in refuting or establishing an entailment, and clearly we have not done that. Rule 3 required that there be a causal connection between the antecedent of the conditional and its consequent. In this case that would require the transmission of the two entities along different paths to be a cause of their arriving at different times. Although not indisputable, this does seem a defensible view. It is true that the claimed causal connection may well be a metaphysically necessary one, but the dispositionalist has no problems on that score, as I mentioned at the outset. Rule 1 looks more contentious, since the entailment is analytic. Once again the issue is whether the entailment is *merely* analytic. We will need to apply the test of rigid designation. Let ABC be a triangle in Euclidean space. We might rigidly designate the property of ABC that we are interested in using 'D' ('D' might be 'ABC's basic geometrical shape'). Then the question is, does 'XYZ is D' entail the appropriate conditional? That all depends upon which property is indeed designated by 'D'. Remember that it is sparse properties that we are interested in. So if 'D' designates the property of being a Euclidean triangle,

then the entailment holds; but, if 'D' designates the property of being a triangle in general, then the entailment does not hold. Our view on this depends on which we think the sparse property really is, and as explained that depends on whether we are dispositionalists or categoricalists. So there is no untendentious application of Rule 1 to penalize this entailment. The dispositionalist has a consistent position that is in conformity with all the rules.

9   Dispositionalism versus categoricalism

The above sketches the account that the dispositionalist should adopt when faced with the challenge of geometrical properties. Geometrical property terms as we typically use them do not always denote sparse properties. That does not, however, prevent us from using them in explanations. This is because, whenever it is true that some object or collection of objects possesses the geometrical property in question, it will also be the case that the object or collection possesses some sparse geometrical property or complex of sparse geometrical properties. The sparse geometrical properties will belong to specific geometries of spacetime. Correspondingly, their instantiation will entail that the appropriate geometry does govern the local structure of spacetime and will entail appropriate subjunctive conditionals. What goes for geometrical properties goes for structural properties more generally.

Although this is a coherent position for the dispositionalist to adopt, one which thus defuses the challenge presented by structural properties, it does not show that dispositionalism *is* the correct story about structural properties. The conclusion of the previous section shows that the categoricalist has an equally coherent story to tell. According to the categoricalist, the generic structural properties are the real sparse properties; the specific structural properties are properties compounded of a generic property plus a specification of the nature of the space in which the particular possessing the property exists. So, for example, 'x is a Euclidean triangle' is equivalent to 'x is a triangle and x exists in Euclidean space'.

In turn, this suggests that neither the dispositionalist nor the categoricalist is likely to be able to win the debate between them by pointing to particular properties or classes of property that are alleged to be explained by one side but not the other. For every story the dispositionalist can tell, the categoricalist can tell a story and vice versa. Let the dispositionalist allege that D is a property whose possession by x entails (modulo finks and antidotes) the truth of the conditional: 'were it the case that Sx then Mx'. Then it will be a law that, *ceteris paribus*, $\forall x(Dx \wedge Sx) \rightarrow Mx$. The dispositionalist will claim that D is a sparse property, that the entailment is metaphysical, and that the law is necessary. The categoricalist can respond by saying that there is no *sparse* property D. Instead the sparse property is some categorical B where 'x is D' entails 'x is B' and 'x is B' itself entails no conditional. The law is the contingent $\forall x(Bx \wedge Sx) \rightarrow Mx$. The proposition '$x$ is D' is analytically equivalent to 'for some categorical property B, x is B and x exists in a world where the law

$\forall x(Bx \wedge Sx) \rightarrow Mx$ holds'. The entailment is thus merely analytic. Conversely, if the categoricalist points to some property B that entails no laws and conditionals, then the dispositionalist can respond that whenever B is actually instantiated by some x then x also instantiates some sparse and truly dispositional property D that does entail laws and conditionals, and it is D that does the explanatory work that the categoricalist ascribes to B.

10 Conclusion

I have presented the debate between Mellor and Prior as a contest between the attempt to prove, subject to certain conditions (Rules 1–4), that the instantiation of a structural property entails the truth of a subjunctive conditional and the attempt to prove (subject to the same conditions) that it does not. The last section shows that at the end of play we must declare a draw in this particular game. However, as I indicated in the introduction, the Mellor–Prior debate is relevant to a larger competition between dispositional monists and their opponents. In that competition, the existence of structural properties was prima facie a stiff challenge to the dispositional monist. Mellor's side, in so far as it was representing dispositional monism, was playing away from home. A victory, proving that triangularity must be understood as essentially dispositional, would have been a very good performance indeed. But a draw away from home is highly respectable. Resisting the attack that structural properties must be understood categorically is a very useful result indeed for the dispositional monist. We have seen that what looked at first to be a reason to reject dispositional monism turned out to be no compelling reason at all. There is a perfectly coherent story to be told about structural properties as dispositional. Dispositional monism has resisted relegation and will live on to play another day.[9]

Notes

1 For details, see Bird (2001).
2 Note that in talking of 'structural' properties I am *not* intending to talk of properties as may be conceived of by structuralists of various kinds. A structuralist may maintain that all there is to some set of entities is the set of more or less formal relations between them. On such a view the essence of a property might just be its relations with other properties. This might indeed make properties dispositional, and certainly dispositional monism might be regarded as a structuralist account of properties, in that sense. But I am not begging the question in this chapter by thinking of 'structural' properties in this sense; rather, they are the properties of objects that exist in virtue of their spatial relations or in virtue of the spatial relations of their parts.
3 I return to this again briefly below.
4 Mellor (1974: 179–80) himself also rejects (B), for slightly different reasons.
5 Mimics are raised by Johnston (1992).
6 One could reformulate the conditional analysis (CA) so as to exclude finks and antidotes, and so remove the need for Rule 2. This is in effect what Mellor (2000) proposes. It is contentious whether the reformulation still constitutes

an analysis. Either way it is more convenient for the following discussion, but equivalent to Mellor's proposal, to keep the simple conditional analysis and to exclude finks, antidotes and mimics via Rule 2.

7   These entailments are not equivalent *simpliciter*.

8   We can see Charlie Martin's reciprocal dispositionality, mentioned above, at work here.

9   I am grateful to Huw Price for helpful comments on a draft of this chapter.

References

Armstrong, D. M. (1997) *A World of States of Affairs*, Cambridge, UK: Cambridge University Press.

Armstrong, D. M., Martin, C. B., Place, U. T. and Crane T. (eds) (1996) *Dispositions: A Debate*, London: Routledge.

Bird, A. J. (1998) 'Dispositions and antidotes', *Philosophical Quarterly* 48: 227–34.

—— (2001) 'Necessarily, salt dissolves in water', *Analysis* 61: 267–74.

Johnston, M. (1992) 'How to speak of the colors', *Philosophical Studies* 68: 221–63.

Lewis, D. (1997) 'Finkish dispositions', *Philosophical Quarterly* 47: 143–58.

Martin, C. B. (1994) 'Dispositions and conditionals', *Philosophical Quarterly* 44: 1–8.

Mumford, S. (1998) *Dispositions*, Oxford: Oxford University Press.

Mellor, D. H. (1974) 'In defense of dispositions', *Philosophical Review* 83: 157–81.

—— (1982) 'Counting corners correctly', *Analysis* 42: 96–7.

—— (2000) 'The semantics and ontology of dispositions', *Mind* 109: 757–80.

Prior, E. (1982) 'The dispositional/categorical distinction', *Analysis* 42: 93–6.

11 Laws, explanations and the reduction of possibilities

Arnold Koslow

1 Introduction

There is a good case to be made for the idea that explanations delimit or 'narrow down' a certain range of possibilities, if the concepts of possibility and the narrowing down or reducing of possibilities are understood in a way that differs from the standard candidates for them that can be found in the literature. So the task, as I see it, is to make the case for these new types of possibilities,[1] and to describe the special way that sets of possibilities get narrowed down by laws and explanations. These possibilities (let's call them *natural possibilities*) might easily be dismissed as no possibilities at all, but merely a case of speaking with the vulgar. Nevertheless, there is, I think, good reason to take these examples as seriously *modal*. Indeed, they represent a kind of modality that opens the way to a new account of the way in which scientific explanations and laws are related to possibilities.

2 Natural possibilities: cases

Let us begin with some familiar cases, where things are usually and naturally described as possibilities:

(1) A die is thrown and there are, as we say, six possibilities. They refer to what happens when the die is subject to a certain experiment, such as tossing, and an outcome (a die with six uppermost) is usually taken to be one possibility among others.

(2) In sample spaces generally, the members of the space are usually described as possibilities. In some cases they are the possible outcomes of an experiment, but this need not generally be the case. Usually the members of the sample space are said to represent all the possibilities.

(3) Declarative sentences are described as being either true or false, and this is described very naturally (in the case of standard logic) to be the only two possibilities.

(4) Another example, like the preceding one, only more complex, can be found in some versions of possible worlds semantics. It is assumed that a certain collection of worlds is such that it contains all the worlds, and that

they are mutually 'incompatible' in some sense. Sometimes this so-called incompatibility is supplemented with a maximality condition to ensure that none of these worlds is a part of any other. When the notion of the truth of a statement at a world and a few assumptions are added which state how the truth of a statement at a world depends upon the truths of its parts in various worlds, we then have the beginnings of a semantic theory of possible worlds. The terminology of possible worlds, of calling each of these worlds a *possibility*, seems entirely natural.

(5)   For most physical theories there is an associated notion of the states of that theory. The collection of all these, the state space of the theory, is commonly described as setting forth the (physical) possibilities for those systems under study by the theory.

(6)   Suppose that in a state space of some physical theory, two points A and B are distinguished. Usually one says that there is an actual path or curve along which the system passes from state A to state B. All the other curves connecting A and B are described as *possible* routes or paths or orbits from A to B. So, rather than the theoretical states, sometimes it may be the curves or paths which are the natural possibilities.

(7)   Another example arises with mathematical proofs or arguments that proceed by cases. For example, either n is a prime number, in which case 'A' is true, or n is composite, in which case 'B' is true. The natural thing is to say that the proof involves two, three or more possible cases.

There are a good many more examples, covering epistemic, ethical, mechanical, physical and mathematical possibilities. The point we wish to emphasize is that, no matter how similar or different these examples may be to each other, they are all examples of what we shall call natural epistemic, ethical, mathematical and physical possibilities.

Why should we think of the options, situations, cases, etc., of these examples as *possibilities*? Why should these colloquial references to possibilities be regarded as anything that is seriously modal? The insistence that all these cases are genuine modal possibilities seems to be overstated. The modal way of speaking is natural, but to take all these cases as possibilities would result in a diluted notion of possibility: they are almost everywhere. Indeed, even the classical *sentential* calculus would be an essentially modal subject, since it would involve the study of things that are either true or false, and those are, as almost any book on logic will tell you, the only *possibilities* in classical logic.

Nevertheless, it would be a mistake to think that there was nothing seriously modal involved in these natural possibilities. In the next few sections we shall explain the modal character of these examples, and settle the question of what kind of modal it is. Moreover, once the modal issues are settled, we can begin to make a case for the serious modality of the states of a physical theory, and we can also give some sense to the idea that the scientific laws of that theory rule out or exclude certain possibilities (theoretical states). Towards an exploration of this situation, we must first say something more about what is modal about all these examples.

We shall describe a mini-theory of natural possibilities, which includes the previous examples as special cases. According to the theory, natural possibilities may be abstract (numbers, numerical equations, truth values) or concrete (one particular act, say eating a banana, or another, slipping on it). They may be object-like, property-like or neither. Nor does it matter whether they have structure (paths from one place to another, vectors, ordered *n*-tuples of physical magnitudes) or appear to be structureless (a real number). What does matter for this general notion of possibility are the following conditions, which suffice for their having modal character.

3   Natural possibilities: a mini-theory

We shall describe a non-empty set N as consisting of *natural possibilities*, if and only if

(1)   N has at least two members;
(2)   any two members of N do not overlap, intersect or have anything in common, and are 'mutually incompatible' in some sense of that term; and
(3)   being an N is in a sense the widest, most inclusive possibility under consideration.

Given the enormous variety of the kinds of things that can be possibilities, it is evident that the second and third conditions are vague and highly metaphorical for it is not clear that there is a uniform sense of 'overlap', 'intersection' or 'incompatibility' which covers all of them. Moreover, since we want to take these possibilities as genuinely modal, we have to say something about how they are systematically related by logical operators like negation, conjunction, disjunction, conditionals, quantification, and so forth. The difficulty of doing this may again seem insuperable, when so many of the examples that we have described do not even have truth values.

Although these problems may appear insurmountable, in fact there is a simple theory of these natural possibilities that provides a clear and uniform form for the conditions (1)–(3), and which will also allow for these possibilities to be negations, conjunctions, conditionals and quantifications of other possibilities. In short, it allows for there to be a logic of natural possibilities.

The basic idea is that the modal character of the members of a set N of alternatives (a set of natural possibilities) becomes evident once N is situated as a member of the implication structure consisting of the power set, $\wp$ (N), of N, together with an implication relation on the power set, i.e. an implication relation on the subsets of N.

In particular, let N be a set with at least two members, and let $\wp$ (N) be its power set (its members are all the subsets of N). In what follows, we let the power set of any set A be designated by A*. There are many different implication relations that can be defined on the power set of N. The one which we

use to develop our account of natural possibility uses the implication relation '⇒', on the set N*, which is defined as follows:

> Any members of N* together imply any member B of N* if and only if their intersection is a subset of B.

We shall call the ordered pair, $I_N = <N^*, ⇒>$, the *(implicational) structure of the N-possibilities*, and we shall call the unitary sets $\{x\}$ for each x in N the *natural possibilities of the structure* I_N. These unitary sets are of course in the power set of N, and so they belong to this structure. And the set-theoretical union of all the natural possibilities of the structure is the set N itself.

By locating the set N within its power set, we now have a clear and uniform reading of the conditions (1)–(3), on natural possibilities. Thus, for any members P and Q of the structure $< \wp (N), ⇒>$, we can define the *negation* of P, ¬P, to be the set N–P, which is also a member of $\wp (N)$, the *disjunction* P ∨ Q as the set-theoretical union of the two: P ∪ Q, the *conjunction* P ∧ Q as the set-theoretical intersection of P ∩ Q. Similar remarks hold for all the other logical operators on the elements of $\wp (N)$.

Although we have stated the conditions for negation and the other logical operators as definitions, these conditions can actually be *proved* to hold once one uses definitions of the logical operators that hold for all implication structures and apply them to the special case of the implicational structure of N-possibilities, as described above.

There are some easy benefits. Conditions (2) and (3) hold for all the natural possibilities of the structure N*. That is:

> (2)   Any two natural possibilities of the structure $<N^*, ⇒>$ based on N are incompatible.

This is so, because any two natural possibilities of the structure are two singletons $\{x\}$ and $\{x'\}$, for distinct members x and x' of N. Consequently, their intersection is the empty set. This shows that the two singletons together imply every member of the structure, and so they are incompatible. The third condition is also provable:

> (3)   The disjunction of all the natural possibilities of N* is a thesis (that is, it is implied by every member of N*).

The reason is that the disjunction of all the natural possibilities of N* is the set-theoretical union of all the sets $\{x\}$ for all x in N. That union is N, and since every member of N* is a subset of N, it implies N. In effect, then, we have shown how to embed any set N (with at least two members) into an implication structure on a set N* with the subset relation as an implication relation on it. The ordered pair $<N^*, ⇒>$ is a special case of what we have

elsewhere called an implication structure.[2] The use of implication structures has an added virtue. It allows for a simple description of modal operators, and so it facilitates the investigation of the question of what sort of modal character the members of N might have.

4   Natural possibilities: their modality

The basic idea is to think of any modal operator φ on an implication structure as a function that maps the structure to itself in such a way that two conditions are satisfied:

(N_1) For any members $A_1, \ldots, A_n$, and B, of the structure

if $A_1, \ldots, A_n \Rightarrow B$

then

$\varphi(A_1), \ldots, \varphi(A_n) \Rightarrow \varphi(B)$.

(N_2) There are some A and B in the structure such that

$\varphi(A \vee B) \Rightarrow \varphi(A) \vee \varphi(B)$

fails.

The first condition says simply that a modal operator preserves implication. If some elements of the structure are related by implication, then their corresponding values under φ are also related by implication. The second condition requires that there are at least two elements of the structure such that φ of some disjunction does not imply the disjunction of the φs of the disjuncts.

Most of the familiar modal operators satisfy these two conditions. For example, a necessity operator that maps any statement A to the statement 'It is necessary that A' is easily seen to satisfy our two conditions.

If we consider an operator which is the dual of a modal operator that satisfies (N_1) and (N_2), for example 'It is possible that A', then it can be proved that it will be a modal operator if and only if the dual of the conditions of (N_1) and (N_2) hold. That is:

(P_1) For any $A_1, \ldots, A_n$ in the structure

$\varphi(A_1 \vee \ldots \vee A_n) \Rightarrow \varphi(A_1) \vee \ldots \vee \varphi(A_n)$.

(P_2) There are A and B in the structure, such that

$\varphi(A) \wedge \varphi(B) \Rightarrow \varphi(A \wedge B)$

fails to hold.

Although these conditions are not familiar, they do cover most if not all the usual cases of modal operators in the literature, and they are useful as a benchmark in any inquiry into the modal character of an operator.[3] We turn now to an investigation of the modal character of sets like N, using their associated implication structures I_N. First we define an operator, $\Diamond$, on N* that represents the natural possibilities of the structure. And, second, we shall justify the use of the diamond notation for it, by showing that it is a modal operator satisfying (P_1) and (P_2).

Let $\Diamond$ be an operator which maps the members of N* to itself, such that

$$\Diamond(x) = \left\{ \begin{array}{l} \varnothing, \text{ if } x \text{ is } not \text{ a singleton (a set having exactly one member of N)} \\ \\ N, \text{ otherwise.} \end{array} \right.$$

It follows that

(1) $\Diamond(x) = N$ if and only if x is a natural possibility of I_N.
(2) For any x and y in N*, $\Diamond(x \cup y) \Rightarrow \Diamond(x) \cup \Diamond(y)$.
(3) There are x and y in N*, such that $\Diamond(x), \Diamond(y) \Rightarrow \Diamond(x \cap y)$ fails.

By (2) and (3) the operator $\Diamond$ is a possibility modal operator, and by (1) it is clear that it represents the natural possibilities in the same way that the members of some set, say X, can be represented by a characteristic function which has only two values: one for all those members of X and the other for all those not in X. Similarly, the function $\Diamond$ is a characteristic function with one value, N, for all the natural possibilities of N*, and another value, $\varnothing$, for all those members of N* that are not natural possibilities (i.e. not singletons of N*). More concisely, if P is the set of the natural possibilities, then any x is in P if and only if $\Diamond(x) = N$. The *natural possibilities* (the singletons of members of N) are one thing and of course the *possibility modal* (the operator $\Diamond$) is another. The proofs of (1)–(3), and of the results to come, are straightforward, and we defer them to another occasion.

If we consider how the modal operator $\Diamond$ behaves on just the natural possibilities of N*, then we have some results that begin to tell us something about what kind of a modal operator it is:

(4) If x is a natural possibility of the implication structure I_N (that is, a singleton whose only member is a member of N), then $x \Rightarrow \Diamond(x)$. There is a close converse:
(5) If x is non-empty and $x \Rightarrow \Diamond(x)$, then x is a natural possibility.
(6) If $\Diamond(x) \Rightarrow x$, then x is not a singleton (and so not a natural possibility).

Since $\Diamond$ does not 'collapse' on the natural possibilities of N*, it does not collapse on N*. It is a genuine modal.

In order to see what kind of a modal operator it is, we need to say something about $\Box$, the necessity modal on the structure $I_N = \langle N^*, \Rightarrow \rangle$. We shall refer to this as *natural necessity*, which we shall define this way: for any member x, of N^*, let

$$\Box(x) = \left\{ \begin{array}{l} N, \text{ if } x = N \\[2mm] \varnothing, \text{ otherwise.} \end{array} \right.$$

Then it is straightforward to show that

(7) '$\Box$' is a modal operator on N^* with respect to the implication already defined for the members of N^*.

(8) '$\Box$' is a necessitation modal (where a modal operator φ is a necessitation modal if and only if $\varphi(x)$ is a thesis whenever x is a thesis, and by a thesis of a structure we mean any member of it which is implied by all the members of that structure).

(9) '$\Box$' is a T-modal. That is, for all x in N^*, $\Box(x) \Rightarrow x$.

(10) '$\Box$' is a K4-modal: for all x in N^*, $\Box(x) \Rightarrow \Box\Box(x)$.

(11) '$\Box$' is an S5* modal: for all x in N^*, $\Diamond(x) \Rightarrow \Box\Diamond(x)$, and the dual $\Diamond\Box(x) \Rightarrow \Box(x)$, also holds for all x in N^*.

(12) $\Box(x) \Rightarrow \Diamond(x)$ does *not* hold for every x in N^* (N is a counterexample).

However, (12) does hold for all the natural possibilities, x, of N^* (that is, all the singletons of N^*) (for then $\Diamond(x) = N$).

In the failure of the box to imply the diamond, this modal operator parts company with many familiar modals, but it shares this property with the Gödel–Löb modal. Moreover, there is no collapse of necessity and possibility on the singletons. In fact, for each singleton x of N^*, $\Diamond(x)$ fails to imply $\Box(x)$, since the former is always N and the latter is the empty set.

There is one unusual feature of natural possibility and natural necessity. In most modal theories (where negation is classical), one or the other of the box and diamond is taken to be primitive, and the other is defined using the not-not formula. However it is easy to see that

(13) For all x in N^*, $\Box(x) \Rightarrow \neg\Diamond(\neg x)$ (the converse fails).

(14) $\neg\Diamond(\neg x) \Rightarrow \Box(x)$ does not hold for all x in N^* (it fails for $x = \varnothing$)

So the full equivalence of box and non-diamond-not does not hold; only one half does. And it is also easy to see that the full equivalence of diamond and not-necessarily-not does not hold; only one half does. That is

(15) For all x in N^*, $\Diamond(x) \Rightarrow \neg\Box(\neg x)$ (the converse fails).

The kind of modality that we have described is very like one familiar strong

modal, and strikingly unlike it. It is very like the modal studied in C. I. Lewis's system S5 in that it is a normal modal satisfying the necessitation condition as well as the conditions that are characteristic axioms for the modal systems T, K4 and S5 [(9), (10) and (11)]. The difference lies with the relation between necessity and possibility. Our necessity implies not-possibly-not (but not conversely), and our possibility implies not-necessarily-not (but not conversely), and this reflects the fact that the necessity modal is a T-modal (9), but the possibility modal is not [by (4) and (5)].

5   Laws and the reduction of possibilities

Collections of states of theories, states of a particle (or field) or system of particles (or fields), collections of paths or orbits in space, spacetime or members of a phase space, as well as collections of outcomes of some experimental device, are serious modal possibilities. Let us consider two claims about laws and explanations that concern how they narrow down or reduce possibilities of this kind. The claims are

(LP) laws narrow down possibilities; and
(EP) explanations narrow down possibilities.

Both principles involve the notion of 'narrowing down', which has to be understood in a special way which we shall explain presently. First, there is an ambiguity that has to be resolved, whatever 'narrowing down' may mean. There are at least two readings of (LP):

(LP1) if anything, say A, is a law, then A narrows down possibilities; and
(LP2) if anything, say A, is a law, then the statement 'It is a law that A' narrows down possibilities.

The difference between the two can be described with the help of some notation. Let '£' stand for the prefix 'It is a law that …', then (LP1) and (LP2) can be written as

(1)   for any A, if £(A), then A narrows down possibilities; and
(2)   for any A, £(A) narrows down possibilities.

The difference is this: according to (1), it is the law itself that narrows down possibilities, but according to (2), it is the condition expressed by 'It is a law that A' that narrows down possibilities. In the presence of certain reasonable assumptions, (1) implies (2), and it is this (stronger) version of nomic narrowing that we will defend.

There is also an ambiguity of scope that should be noted, whichever version, (1) or (2), we advocate. Consider (1). It could require that there is a set of possibilities such that, if any A is a law, then A narrows down those possibilities.

That is the wide-scope version. According to the narrow-scope version, if A is a law, then there is a set of possibilities such that A narrows them down. The narrow-scope version allows that the set of possibilities could differ from law to law, whereas wide scope has it that there is this super-set of possibilities, which gets reduced by every law. The narrow version yields a more refined and more accurate account of the way things happen scientifically.[4]

We turn now to an explanation of the special way in which laws exclude possibilities. It involves several assumptions whose full support we must postpone for another occasion. The first is (1) to each scientific law £, there is associated a certain non-empty set of possibilities $\wp_£$ associated with it. The idea is that there is at least one such set of possibilities; there may however be several. We do not insist on the uniqueness or even the maximality of each of these sets. The second is (2) that some of the members in *some* set of possibilities $\wp_£$ are 'ruled out' or excluded by the law £. In this sense, 'the' set of possibilities is narrowed down.

Before we provide our version of how laws rule out possibilities, it is worthwhile considering one familiar way of explaining this feature of laws, if only to discard it as useless. A notion of *nomic* possibility is sometimes defined this way: A is *nomically* possible if and only if it is compatible with the set of all laws. And A is *nomically* necessary if and only if it is a consequence of the set of all laws. From which it follows immediately that every law is *nomically* necessary. This is not a deep thought, just an immediate consequence of definitions. It does ensure however that for any A

(N) $\Box (A) \Leftrightarrow \neg \Diamond (\neg A)$.

If we think of one statement excluding another if and only if it is incompatible with it, then (N) guarantees that the necessity of any A, and of any law L in particular, excludes the statement $\Diamond(\neg A)$. In particular, the necessity of any law L excludes that it is *nomically* possible that L is false. However, this does not tell us much about *nomic* possibility; there are many modal systems that satisfy condition (N), from which it follows that the necessity of any A excludes the possibility of not-A. Moreover, on this proposal, it is the necessity of L($\Box$(L)) that excludes some possibilities, whereas the stronger and, we believe, the correct claim is that it is the law itself which excludes various possibilities.

Two further assumptions are needed for our more structured account of how possibilities are excluded. The third is (3) that each law involves certain physical quantities or magnitudes. We shall not say much at present about these physical magnitudes. For present purposes it is enough that some physical quantities (e.g. mass, length, velocity, density, kinetic energy, temperature charge, etc.) are functions that map physical entities or structures of physical entities to elements of some mathematical structure (e.g. a real number, a vector, matrix, tensor, etc.) This is not the whole story. Not all physical quantities are mappings from physical entities, or structures of them, to mathematical structures. There is an important group of physical magnitudes, perhaps the

most important ones, that are *functionals*. They figure prominently in laws of classical mechanics, electromagnetism, thermodynamics and a whole range of contemporary field theories, relativistic and quantum theoretical. If we think of some of the physical quantities as mappings from physical entities or systems of physical entities to mathematical structures, then those physical quantities that are *functionals* are mappings from functions to mathematical structures. That is, the *functionals* are mappings from functions to mathematical structures, rather than mappings from physical entities to those mathematical structures. The point is that certain functionals are physical magnitudes. The final assumption is (4) that to each scientific law not only is there some set of possibilities $\wp$ associated with that law, but there is also a special kind of property, Σ_Φ (which depends on a functional magnitude Φ), which holds or fails to hold of the possibilities in $\wp$. We shall refer to the property Σ_Φ as a *functional property*. So here is the idea of how it is that laws exclude or narrow down possibilities:

> (LRP) (1) For each law L there is some property Σ_Φ expressed with the aid of a functional Φ which holds or fails to hold for the members of a certain set of possibilities, and (2) L implies that some possibilities fail to have that property.

According to (2), if L is a scientific law, then for some possibility α of a set of possibilities $\wp$, $L \Rightarrow \neg\Sigma_\Phi(\alpha)$.[5] It may also happen that some law excludes all the possibilities that are associated with it, but that is not generally so for all laws. Furthermore, a law L may actually guarantee that some of the possibilities have the functional property Σ_Φ, that is, for some possibility β, $L \Rightarrow \Sigma_\Phi(\beta)$.

On our proposal, it is not the possibilities themselves that are implied (or not) by the law L; it is the *statements* $\Sigma_\Phi(\alpha)$ (that the possibility α has the functional property Σ_Φ) that are implied (or not) by L.

There is little space to show in any detail how typical laws reduce possibilities. Any law which can be expressed as a special case of Hamilton's Principle of Least Action easily falls under the concept expressed by (LRP). The reason is that such applications proceed by specifying a particular Lagrangean (T–U) for some physical system, where T is the kinetic energy and U the potential energy of that system. A set of curves f, g, … is specified between two points in a state space. Think of these curves as the possibilities, and there is a functional called the *action* for the particular Lagrangean L, which is given by

$$\Psi\big[f\big] = \int_{t_0}^{t_1} L\big(q, dq/dt, t\big)\, dt$$

(where the integral is taken along the curve f). With the help of the functional Ψ, we can define a functional property Σ_Ψ of the possibilities (curves f, g, …) as follows:

Σ_ψ holds of the curve f if and only if f is a curve for which Ψ, the action, is an *extremal* – that is, $\Psi(f)$ is either a maximum or a minimum.

In all these 'Hamiltonian' cases, possibilities are ruled out or excluded if and only if it is implied that they fail to satisfy the functional property Σ_ψ. That is, they fail to be extremal paths.

Here is a simple example of this kind of case. For a free particle moving in Euclidean three space, the potential, U, is 0, so that the Lagrangean for the free particle is $L = T = mr^2/2$. If one uses generalized coordinates, $L = m/2(q_1{}^2 + q_2{}^2 + q_3{}^2)$. The Lagrange–Euler equations then yield that the generalized momentum $p = \partial L/\partial q_i$ is constant $[dp/dt = d(\partial L/\partial q_i) = 0]$, since the Lagrangean is not a function of the generalized coordinates q_i. This simple result can be expressed by saying that straight lines are the extremals of the action of free particles.[6] That is, the law of inertia in this formulation rules out any possible path that is not a straight line.

A law need not be formulated using Hamilton's Principle of Least Action in order to show how it excludes possibilities. Here is a sketch of how a traditional presentation of Newtonian Gravitation Theory can also do the job.[7] Assume among other things that the force acting upon a body is a central force with the potential $V = -\alpha/r$ with α positive, so that the force is attractive, central and proportional to the inverse square of the distance. Using polar coordinates, the equation of the orbit is given by

$$r(\theta) = \lambda(1 + \varepsilon)/[1 + \varepsilon\cos(\theta - \theta_0)]$$

where the constant λ is defined to be $|L|^2/[m\alpha(1 + \varepsilon)]$, where L is the total angular momentum of the planetary body, and α in this, the gravitational case, is Gm_1m_2. This equation is the focal equation of a conic section, with eccentricity ε. (Recall that any conic section can be described with the aid of a line, the directrix, a fixed point F (a focus) and the ratio of the distance r between the body and the focus F, to the distance between F and the directrix.) The following table is a standard result for the types of conic that satisfy the equation of the orbit:

$0 < \varepsilon < 1$	$\lambda > 0$	ellipse
$\varepsilon = 0$	$\lambda > 0$	circle
$\varepsilon = 1$	$\lambda > 0$	parabola
$\varepsilon > 1$	$\lambda > 0$ or $\lambda < 0$	hyperbola

Consequently, assuming that the gravitational force is attractive, central and inverse square, it will follow that there are four possibilities left open. The idea conveyed by (LRP), that scientific laws exclude certain possibilities, is

illustrated in the gravitational case this way: the possibilities or curves for planetary orbits consist at the very least of differentiable trajectories, and the Newtonian Law of Gravitation narrows down those possibilities to just the conics. The reduction or narrowing down can be cast in the same terms that were used for all the Hamilton Least Action cases, only in this case a simpler functional property is available: let $r(\theta)$ represent the path of a body in polar coordinates, and the functional property $\Sigma_\Phi(\ )$ holds of any curve $r(\theta)$ if and only if there are some $\lambda, \varepsilon, \theta$ and θ_0, such that $r(\theta) = \lambda(1 + \varepsilon)/[1 + \varepsilon\cos(\theta - \theta_0)]$. Clearly, the requirement that the gravitational force is central, attractive and inverse square will rule out all non-conic curves $X(\theta)$, because these requirements on the force function imply that $\neg\Sigma_\Phi(X(\theta))$.

The gravitational case shows in a clear way that, although laws reduce the possibilities, they need not reduce them to all but one. In the gravitational case, several possibilities (the various conics) may be left, and in the case of some laws, the so-called impossibility laws, all of the possibilities may be ruled out. An *explanation* of the elliptical orbit of Mars narrows down the possible orbits from the conics to the ellipse. This further reduction beyond that furnished by the law of gravitation is obtained through further information in the explanation – say the values of λ and ε, as in our discussion, or in the specification of definite values of other parameters that would pick out the ellipses from the other types of conics.[8]

6   Explanations and the reduction of possibilities

Do explanations rule out possibilities? From the *facticity* condition on explanation,[9] it follows that those explanations that are either explanations of laws or involve laws as part of the explanation of something else will certainly exclude some possibilities. The reason is that by *facticity*, in either of the two cases, laws are implied, which in turn implies that some possibilities are excluded. Thus, for explanations of this kind, clearly (EP) holds. However, there are some accounts of explanation that explicitly eschew the use of laws and some which are silent about the need for them. We have proposed that scientific explanations narrow down possibilities. This is an important feature, but not the only feature that they have. It is remarkable, however, given the various specific models of explanation, and given the various adequacy conditions for explanation that have been proposed, that almost none of them guarantees that explanations reduce possibilities. One of the few exceptions is a constraint on explanations that Hugh Mellor has advocated. It yields the result that explanations do indeed narrow down possibilities, and the reduction is carried out in a way which is very like that seen in the scientific cases.

Let E(A; B) stand for the existential statement that there is some explanation of B in which A is all or at least part of the explanans. The constraint on explanation Mellor has advocated[10] is that

(M)   E(A; B) implies that $C_A(B) - C_{\neg A}(B) > 0$.[11]

That is, if there is an explanation according to which A explains why B, then it follows that the chances of B, given A, are greater than the chance of B, given ¬A (i.e. the chance of B in the absence of A).

There are two questions of interest. First, does (M) show that explanations narrow down possibilities? We wish to show that it does, if we think of possibilities along the lines that we have suggested. Second, is the narrowing down of possibilities similar to the way that scientific laws and explanations narrow down possibilities? Here again, the answer is positive. If you add to these two positive points the observation that Mellor's theory of chance regards it as a measure of something objective (degrees of possibility?) so that it becomes like an objective physical magnitude, then the result is a feature of explanations that it shares with those explanations which are at the heart of scientific practice – not the awful examples of birds and their colour, but the awesome ones like planets and their orbits.

In Mellor's account, any B is a possibility if and only if it has a non-zero chance. From which, with some weak assumptions, it follows that those things are necessities that have a chance of one. There are two ways of showing that (EP) holds, given Mellor's (M). One uses the three cases (positive, zero and negative) that can hold for $C_A(B) - C_{\neg A}(B)$ to generate a set of possibilities. For lack of space, we set this aside. The other generates a different set of possibilities using two conditions on the chance of B: that $Ch(B) = 0$, or that $Ch(B) \neq 0$. This second set of possibilities is closer to the version that Mellor advocates, although it is not identical to it. We hasten to add that either version follows the pattern of the scientific examples.

Suppose that there is an explanation of B of which A is a part, or all, of what explains B. That is, E(A; B). Chance is a function that maps to the reals in the closed interval [0,1]. So for any B in that domain, we associate two *constant* functions, f_B and g_B, defined as follows: For all E in the domain of the chance function:

$$f_B(E) = 0, \text{ iff } Ch(B) = 0, \text{ and}$$

$$g_B(E) = k, \text{ iff } Ch(B) = k, \text{ where } k \neq 0.$$

We can now introduce a functional Φ, which maps f to 0 and g to whatever constant non-zero value, k, that g has (and complete, in the obvious way, the definition of Φ for those functions other than f and g in the relevant space of functions). Now that we have the possibilities f and g and the functional Φ, we can see this example as a special case of the scientific ones, once we define a (*functional*) property. Let be Σ_Φ be defined so that for every function h that maps to the closed unit interval

$$\Sigma_\Phi(h) \text{ if and only if } \Phi(h) = 0.$$

On the Mellor account, $E(A; B)$ implies $C_A(B) - C_{\neg A}(B) > 0$. Since explanations are factive, $E(A, B)$ implies A (and B). Now, according to Mellor's account of chance, $C_A(B) - C_{\neg A}(B) > 0$ together with A imply that $Ch(B) \neq 0$. Consequently, $E(A; B)$ implies that $Ch(B) \neq 0$.

What needs to be shown is how it is that (M) reduces possibilities. Suppose that $E(A; B)$ and $\sum_\Phi(f_B)$. $E(A; B)$ implies, as we have seen, that $Ch(B) \neq 0$. On the other hand, $\sum_\Phi(f_B)$ implies that $\Phi(f_B) = 0$, which implies that $f_B = 0$. And that in turn implies that $Ch(B) = 0$. The assumption leads to a contradiction so that $E(A; B) \Rightarrow \neg\sum_\Phi(f_B)$. This shows, on our account, that if there is any explanation of B, then some possibility associated with B is ruled out – namely f_B. Of course $E(A; B)$ implies $\sum_\Phi(g_B)$, but that is not an issue. Some possibilities rather than others are ruled out. As a consequence, there is a certain contrastive feature to explanations and to laws as a byproduct of their reduction of possibilities.

Notes

1 I am indebted to Steve Leeds's (2001) seminal paper for his insights into physical possibilities. I now think of them, thanks to him, as a special case of the natural possibilities in this chapter, but I doubt that this is his view. The case for reducing them and the natural possibilities to a kind of metaphysical possibility now looks more implausible than ever.
2 See Koslow (1992).
3 For an extended discussion of these conditions for modal operators, see Koslow (1992).
4 We shall have to defer the defence of this assertion for another place.
5 Here we intend by the double-shafted arrow that some standard notion of implication be used. It is not to be confused with the special implication relation on the set N*, which, by an abuse of notation, is also indicated by a double-shafted arrow.
6 Compare with Arnold (1978: 61).
7 For present purposes, any of a number of contemporary accounts will serve. For example, Corben and Stehle (1957), Dubyago (1961), Berger and Olsson (1973) or Arnold (1978).
8 One has to be careful about individuation. By the reduction of the possibilities to one (the ellipse) we obviously mean a type of curve. If the possibilities included ellipses of various sizes, or of the same size but different orientations in space, then our gravitational explanation would not narrow the possibilities to one.
9 By the facticity condition on explanation we mean the condition that $E(A; B)$ implies A as well as B, where '$E(A; B)$' means that there is an explanation of B in which A is a part or perhaps all of the explanans.
10 Mellor (1995: 73).
11 There may be a difference between Mellor's proposal and a Mellor-style (M): Mellor may believe that it is every particular explanation that implies the difference between the two chances, whereas in our version the difference of the chances follows from the existential claim that there is an explanation of B that involves A. On our general version, no one is forced to settle whether explanations are arguments, rules, statements or anything else. That is, no specific model of explanation is needed to state the condition.

References

Arnold, V. I. (1978) *Mathematical Methods of Classical Mechanics*, New York: Springer Verlag.

Berger, V. and Olsson, M. (1973) *Classical Mechanics, A Modern Perspective*, New York: McGraw Hill.

Corben, H. C. and Stehle, P. (1957) *Classical Mechanics*, New York: John Wiley & Sons, Inc.

Dubyago, A. D. (1961) *The Determination of Orbits*, New York: Macmillan.

Koslow, A. A. (1992) *Structuralist Theory of Logic*, New York: Cambridge University Press.

Leeds, S. (2001) 'Possibility: physical and metaphysical', in C. Gillett and B. Loewer (eds) *Physicalism and its Discontents*, Cambridge, UK: Cambridge University Press.

Mellor, D. H. (1995) *The Facts of Causation*, London: Routledge.

12 What is wrong with the relational theory of change?

Gonzalo Rodriguez-Pereyra

1

Things, or objects, change their properties: a banana is green one day and some days later it is yellow; a kettle is hot at one time and some time later it is cold; a person is bent at the times when he or she is sitting and straight at the times when he or she is standing. How can a banana be both green and yellow all over? By being green and yellow at different times, of course, since for something to change it must have incompatible properties at different times.[1] But how is change possible? Given that certain properties cannot be had at the same time, why is it possible to have them at different times? Why and how does a difference in time make possible what is otherwise impossible? Why is it not a contradiction that a banana is green and yellow, i.e. not green, all over at different times? This is the problem of change, and several solutions have been proposed.

Some philosophers, such as David Armstrong (1980) and David Lewis (1986: 202–4), think that a difference in time makes possible what is otherwise impossible because a difference in time is also a difference in parts. No doubt it is possible for a thing to be green and for another to be yellow, and this, according to these philosophers, is what happens in the case of the banana: it is one thing that is green, a certain temporal part of the banana, and another one that is yellow, another temporal part of the banana. Others, among them presentists such as Mark Hinchliff (1996: 123–6), think that a difference in time makes possible what is otherwise impossible because a difference in time is also a difference in tense. No doubt it is possible for a thing to be green and not to be other colours and this, according to these philosophers, is what happens in the case of the banana: the banana *is* just green and any other colour is a colour that the banana *has had* or *will have*.

Yet other philosophers propose yet another solution. Hugh Mellor, in his *Real Time* (1981: 111–14), thought that a difference in time makes possible what is otherwise impossible because a difference in time is always a difference in relata. No doubt it is possible for a thing to bear a relation to a thing and an incompatible relation to a different thing, and this, Mellor thought, is what happens in the case of the banana: it is one thing with respect to which the

banana is green, time t, and another thing with respect to which the banana is yellow, later time t'.

Mellor in 1981 held what I call the *relational theory of change*. In its canonical version, the theory holds that changeable properties are really relations between things and times. It thus explains the change of the banana by saying that it bears the relation *green-at* to a time t and the relation *yellow-at* to a later time t'. Thus, according to this theory, although it is impossible to be green and yellow all over at the same time, it is possible to be green and yellow at different times, because this involves different relata. The relational theory, also held by Peter van Inwagen (1990), has recently been abandoned by Mellor in his *Real Time II*, in which he argues against it. In this chapter I shall try to show why the relational theory fails to account for change, and I shall also criticize the arguments of several philosophers, including Mellor, against the theory.

My aim in this chapter is to present a new argument against the relational theory of change. Since the relational theory has already been rejected by many philosophers, before presenting my own argument against it I shall show why these other arguments against it are not effective. Thus, in Section 2, I shall say something more about the problem of change and the relational theory; in Sections 3–6 I shall criticize the arguments of several contemporary philosophers, including Mellor, against the relational theory; and in Section 7 I shall give my new argument that the relational theory fails.

2

The problem of change, which the relational theory tries to solve, is sometimes called the 'problem of temporary intrinsics'. The problematic entities are supposed to be properties; they are temporary because they are not had by their subjects at every time, i.e. they are changeable properties; and they are intrinsic because if a thing has such a property this is supposed to be independent of any and every relation the thing in question bears to anything.

But taking the problem of change to be the problem of temporary intrinsics is making relational change either inexistent or unproblematic.[2] Relational change, however, is as existent and as problematic as intrinsic change. That there is relational change is proved just by giving examples, i.e. a is hotter than b at t and a is colder than b at t'. This example, however, does not prove relational change to be something over and above intrinsic change; for here relational change is clearly supervenient upon intrinsic change of a and b, namely change in a a's and b's temperatures. But there are relations, such as some spatial ones, which supervene upon no intrinsic properties of the relata. Thus, that a and b are 2 miles apart at t and they are 1 mile apart at t' is genuine, irreducible *relational* change on a relational theory of space. And on a substantival theory of space, change of distance is explained in terms of change with respect to the region occupied, so that when a occupies region x at

t and it occupies region y at t' there is genuine, irreducible relational change. So, in general, spatial change is genuine relational change.

Whatever one's theory of space, change of distance is as problematic as intrinsic change. How can a and b be both 1 mile apart and 2 miles apart? By being 1 mile apart at a time and being 2 miles apart at a different time, of course. But how is this change possible? Change consists indeed in having incompatible properties or relations at different times, and so the problem of change is to explain change, both in properties and in relations. It is particularly important not to neglect relational change, since according to the relational theory all change is relational change. In conclusion, the relational theory is, or should be, a solution to the problem of change in general, not just to the problem of intrinsic change.

In the case of allegedly intrinsic properties like *being green*, *being hot* or *being bent*, the relational theory says that these are really relations to times. Thus, for a to be green at t is for it to bear the relation *green-at* to t. In the case of relations such as *being 2 miles apart* the relational theory must claim that this is only apparently a two-place relation. Really, the theory claims, it is a three-place relation that holds between the things that are 2 miles apart and the times at which they are 2 miles apart. In general, the theory has it that apparently n-adic relations are really $n + 1$-adic relations, with an extra place for a time. This version of the relational theory, which I shall call the *canonical version*, is the one which Mellor held in *Real Time*.

There are two other versions of the relational theory.[3] The second version of the relational theory says that, although *being green*, *being bent* and the like are indeed properties and not relations, they are not intrinsic properties but *relational* ones. Relational properties are those which are held by a thing in virtue of that thing standing in some relation to some other thing or things. For example, the property of *being admired* is a relational property of Socrates, which he has in virtue of the relational fact that Plato admires Socrates. In this version, then, the relational theory claims that an allegedly intrinsic property like *being green* is really a relational property held by green things in virtue of a certain relation (the *green-at* relation) holding between them and the times at which they are green. This version of the theory says that an apparently two-place relation such as *being 2 miles apart* really is a two-place relation, but one which holds between any two things x and y at any time t in virtue of x, y and t standing in a corresponding three-place relation (a three-place relation of *being 2 miles apart at*). In general, according to this version, an apparently n-adic relation is really n-adic, but it holds between its relata in virtue of an $n + 1$-adic relation between those relata and the time at which they are so related. I shall call this version of the relational theory, which as far as I know has passed unnoticed, the *relational property version*.

Finally, there is a third version of the relational theory which, as Lewis (1999: 188, fn. 1) would say, puts the relationality not in the properties themselves but in the having of them. I shall call this version of the relational theory the *instantiation version*. The instantiation version comes in three different variants.

According to the first variant, which Lewis (1999: 188, fn. 1) calls *adverbial*, for *a* to have an intrinsic property **F** at *t* is really for a three-place relation of *instantiation* to hold between *a*, **F** and *t*. In the case of relations, this variant must hold that for *a* to stand in an *n*-adic relation **R** to $x_1, \ldots, x_{n-1}$ at *t* is really for an *n* + 2-adic relation of instantiation to hold between *a*, $x_1, \ldots, x_{n-1}$, **R** and *t*.[4] The two other variants of the instantiation version are introduced and defended by van Inwagen (1990: 247). One of these variants has it that for *a* to have an intrinsic property **F** at *t* is really for *a* to bear the relation of *instantiation* to the time-indexed property **F-at-*t***. In the case of relations this variant must hold that for *a* to stand in an *n*-adic relation **R** to $x_1, \ldots, x_{n-1}$ at *t* is really for the relation of instantiation to hold between *a*, $x_1, \ldots, x_{n-1}$ and *the time-indexed relation* **R-at-*t***. The second variant distinguished by van Inwagen has it that for *a* to have an intrinsic property **F** at *t* is really for *a* to bear the time-indexed relation of *instantiating-at-t* to **F**. In the case of relations this variant must hold that for *a* to stand in an *n*-adic relation **R** to $x_1, \ldots, x_{n-1}$ at *t* is really for the time-indexed relation of *instantiating-at-t* to hold between *a*, $x_1, \ldots, x_{n-1}$ and **R**.[5]

In the next four sections I shall show why various arguments against the relational theory fail. In the last section I shall produce a new argument which shows that the relational theory, in all of its versions, fails to account for change, including relational change.

3

The simplest argument against the relational theory is Lewis's (1986: 204; 1999: 188), who just takes the position to be untenable because it denies that there are any temporary intrinsics. He says that shapes are properties, not relations, and that we know that this is so. Hinchliff (1996: 121–2) and Merricks (1994: 168) adhere to Lewis's view, and reject the relational theory because it conflicts with our intuition that shapes, for instance, are not relations.

Notice that this affects only the canonical version of the relational theory, for according to both the relational property and the instantiation versions colours, temperatures and shapes still count as *properties*, not relations. Lewis (1999: 188, fn. 1) is aware that his complaint does not touch the first variant of the instantiation version, but he insists that it still amounts to a denial that things have temporary intrinsics. The same must be true of the relational property version and the other two variants of the instantiation version for Lewis's rejection of the relational theory to be solid.

But Lewis's argument is hardly an argument at all. Indeed, do we *know* that allegedly intrinsic properties are not really relations to times? What we have is, at most, an intuition, in the sense of a pre-theoretical and uncritical belief, that they are not relations to times. But then I echo Forbes's remark that he does not see 'how we could be confident that shape is not a relation to a time if we are unsure whether proximity is two-place or three-place' (Forbes 1987: 140, fn. 3).

Furthermore, the counterintuitiveness of the relational theory need not be a defect since strict adherence to our intuitions would make the progress of knowledge impossible. My point is simply that being counterintuitive is not enough to reject a theory, especially in the case in question, since all theories about change are counterintuitive to some degree or other. To put it in an *ad hominem* way: our intuition that shapes, colours and temperatures are not relations is neither stronger nor more credible than our intuition that things like people, bananas and kettles do not have temporal parts, but that such things have temporal parts is what Lewis (1986: 204) believes. Thus, Lewis has produced no reasons to reject the relational theory – at best he has shown that it is counterintuitive.[6]

4

Another argument against the relational view is advanced by Johnston (1987: 113, 128). Exact duplicates are things sharing all their intrinsic properties, and duplicates existing at different times are as much duplicates as duplicates existing at the same time. But then, Johnston thinks, having a changeable intrinsic property cannot really be bearing a relation to a time – otherwise 'duplicates existing at different times would have different intrinsics' (Johnston 1987: 113), which contradicts the original characterization of exact duplicates.

This is anything but conclusive. For the relational theorist might just accept that, under Johnston's definition, no exact duplicates could exist at different times. This, of course, does not mean that things bearing exactly the same relations to different times do not look exactly the same.

A further complaint of Johnston's (1987: 113) is that the relational theory requires things to change their properties continuously, even if they suffer no apparent qualitative change. A simpler way to put this point is to say that the relational theory makes things change continuously. But this is just confusion. For change is having incompatible properties or relations at different times. Consider the canonical version of the relational theory: according to it a banana might well bear the *green-at* relation to two consecutive times *t* and *t'*. If so, the banana bears the *same* relation to different consecutive times and so it has not changed since, of course, the *green-at* relation is not incompatible with itself. There is nothing in the relational theory that requires things to bear incompatible relations to consecutive times.

5

Another argument against the relational theory is Hawley's (1998: 213), which implicitly assumes that a certain distinction between internal and external relations is exhaustive. Thus, she argues first that changeable properties cannot be *internal* relations and then that taking them to be *external* relations makes them mysterious entities, and so, she thinks, the temporal parts theory should be preferred over the relational theory.

The distinction between internal and external relations can be drawn in several different, but related, ways. For Hawley, internal relations are those that supervene upon the intrinsic nature of the relata. By this she means, I take it, that if **R** is an internal relation which *a* bears to *b*, then necessarily every two things *x* and *y* with the intrinsic natures of *a* and *b* respectively are such that *x* bears **R** to *y*. Provided temperatures are intrinsic properties, the relation of *being hotter than* is an internal relation in this sense.

Could changeable properties be internal relations that things bear to times? Hawley (1998: 214) thinks not, for two different reasons. Basically, her argument is that if changeable properties are taken to be internal relations to times then one is committed not only to absolute time but also to the strange theory that times have intrinsic properties. On the other hand, if changeable properties of things are internal relations to times, then things have very few intrinsic properties and it is difficult to see how the great number of a thing's changeable properties can be accounted for in terms of those few non-changeable intrinsic properties.

At this point, I think, someone could invoke a different notion of internal relations, according to which entities having no intrinsic properties can enter into internal relations provided these supervene upon the intrinsic properties of the other *relatum*. But this will not help the relational theory. For although this will be compatible with times having no intrinsic properties, we shall still have the problem of how to account for the many changeable properties of things in terms of a few non-changeable intrinsic properties. Perhaps then we could resort to a notion of internal relations according to which they are those which supervene upon the identity of the terms? This will be of no help, for it has the awkward consequence that for all changeable properties **F**, if a thing has **F** at *t* then it is essential for that thing to have **F** at *t*.

Thus, I agree with Hawley that changeable properties cannot be internal relations to times. Can they be external relations? Hawley (1998: 215), again, thinks not. For her external relations are those determined by or supervenient upon the intrinsic properties of the fusion of the relata (Hawley 1998: 215). As Hawley (1998: 215) says '[i]f the distance between an object's parts is one of its intrinsic properties, then spatial separation is an external relation'.[7]

Thus, supposing that there is a thing which is the fusion of the banana and the time *t* at which it is green, what intrinsic property of that fusion could determine the external relation of being *green-at* which the banana bears to *t*? In other words, what external relations hold between the banana and *t*? One might think that the answer to these questions is just the spatio-temporal separation of the banana and *t*. But this, Hawley says, will not do, for such a separation is a temporary or changeable property of the banana–*t* fusion, 'since the banana gets closer to *t*, then further away, as time passes' (Hawley 1998: 215). Hawley concludes that the intrinsic properties of things–time fusions that determine the changeable properties of things must be special, permanent, non-spatio-temporal and non-causal properties of the said fusions. What these properties are nobody knows. They are mysterious properties.

But then, since changeable properties cannot be internal relations and taking them to be external relations makes them mysterious entities, Hawley (1998: 215–16) concludes, the relational theory should be rejected.

Hawley's argument, if it worked, would devastate the relational theory, for although the relational property and the instantiation versions of it do not make properties relations they make those properties, or the having of those properties, depend on relations which seem neither internal nor external in Hawley's sense.

But Hawley's argument does not work. The problem with Hawley's argument is that her distinction between internal and external relations is not exhaustive. Internal relations are those that are determined by the intrinsic properties of the relata, whereas external ones are those determined by the intrinsic properties of the fusions of the relata. This leaves room for relations which are determined by no intrinsic properties of anything. Couldn't changeable properties be of this kind? Unless Hawley can give an argument to support a negative answer to this question she has not undermined the relational theory.[8]

6

In his *Real Time II* (1998) Mellor argues against the relational theory adopted in his *Real Time* (1981). Why does he now reject the relational theory? Take the example of the banana again. He says that if the banana is green at t then the banana and t are co-located in time. But, Mellor says, relations do not entail, in general, that their relata share temporal location. Thus, Mellor concludes, changeable properties are not relations between things and times (Mellor 1998: 93–4).

This, however, is not strictly true. Indeed the instantiation relations posited by the instantiation versions do seem to entail temporal location among their relata, and so Mellor's argument does not work against all versions of the relational theory. But let us see whether Mellor's argument can rule out the other versions of the relational theory. Is it then true that relations (other than the instantiation relations of the instantiation versions of the relational theory) do not entail that their relata share temporal location? Mellor (1998: 94) is aware that there are some exceptions and he cites simultaneity. This, as he suggests, is a rather trivial example. So it is important to note that the phenomenon is a more extended one and that there are many other relations which cannot hold unless their relata share temporal location: *being in contact with*, *living in*, *working with*, *being married to* are some examples. So why could not all changeable properties be like these?

Mellor says that what makes the fact that the banana is green at t entail that the banana is located at t is that *being green* is a non-relational or intrinsic property of the banana which requires the banana to be located whenever and wherever the banana is green (Mellor 1998: 94). That is, for Mellor, changeable properties are not relations, because they are intrinsic properties.

Why does Mellor think changeable properties are intrinsic properties? Because he requires that real changes of properties have effects, 'and for them to be changes in the things to which we ascribe those properties, that is where their first effects must be' (Mellor 1998: 88). This causal test of change provides two related tests for changeable properties which Mellor thinks rule out relational properties as changeable properties. The first is a causal test for properties according to which real changeable properties are those whose changes have their first effects on or near the things we ascribe them to. So, for instance, since the first effects of a change in Lenin's fame are on or near those whose thinking of him makes him famous, Lenin's fame is not a property of his. Mellor then extends his case for saying that *being famous*, *being taller than Jeff* and *being an only child* are not real properties to *relational properties in general* not being real properties (Mellor 1998: 88).

But is he justified in this generalization? Consider the relational property of *being in contact with b*. If, at *t*, *a*, which is cold, is in contact with *b*, which is hot, then their being separated at *t′* is a change which has effects on the thing to which we ascribe the property of *being in contact with b*, namely *a*, since it, or part of it, will suffer a change of temperature as a consequence of its separation from *b*.

Consider Mellor's second test for changeable properties: a thing's properties should be detectable just by inspecting that thing (Mellor 1998: 88). This also prevents *being famous*, *being taller than Jeff* and *being an only child* being properties of the things they are ascribed to. But again this test does not rule out all relational properties: *having been murdered*, for instance, is a relational property detectable by inspecting the person it is ascribed to. Otherwise, forensic experts could not determine whether they are in the presence of a murder, a suicide or an accidental death unless they have seen how the death occurred.

In conclusion, Mellor's general claim that relational properties are not real properties of the things they are ascribed to is unjustified. For all Mellor has shown, there are some relational properties which are real changeable properties. Thus, Mellor has not shown that changeable properties must be intrinsic properties, and so he has not shown that they are not relational properties or relations to times. And, of course, *being green* and *being yellow* pass both Mellor's tests for changeable properties. If the banana changes from green to yellow then the first effects of this are in the banana itself and, of course, the colour of the banana is detectable by inspecting the banana. So even if *being green* and *being yellow* are properties, they may be relational ones, in which case Mellor has not shown that it is not the case that the banana is green at *t* and yellow at *t′* by bearing the *green-at* relation to *t* and the *yellow-at* relation to *t′*.

7

If the previously examined arguments against the relational theory fail, why do I still maintain that this theory is false? The reason is simple: change is having *incompatible* properties or relations at different times, but in the relational

theory's picture of change this incompatibility disappears. Thus, change is not what the relational theory says it is. In other words, if the relational theory is true then there is no change, for then nothing has incompatible properties or relations at different times. This eliminativist feature makes the relational theory untenable, since its purpose was precisely to account for change, not to deny it.

Why does the relational theory fail to account for the incompatibility required by change? The reason is that the relational theory makes all change relational change, and for there to be relational change a thing must bear incompatible relations to the *same* entity at different times, but the relational theory fails to provide such a single entity, since on that theory incompatible relations like *green-at* and *yellow-at* are borne to *different* entities, namely *different* times.

Consider again what the canonical version of the relational theory says about the banana. The banana bears the *green-at* relation to t and the *yellow-at* relation to t'. Since *green-at* and *yellow-at* are relations, their incompatibility means that nothing can bear both of them to the *same* entity (at the same time). *Liking* and *disliking* are incompatible relations because nothing can bear those relations to the same entity at the same time, although of course there is no incompatibility in liking Tom and disliking Tim. Indeed, that Mike likes Tom at t and dislikes Tim at t' constitutes no change for Mike. For Mike might both like Tom and dislike Tim at the same time t. Mike would change if, for example, after liking Tom at t he came to dislike Tom at t'. For liking Tom is incompatible with disliking him.

Thus, for the banana to change it should pass from bearing the *green-at* relation to t, to bearing the *yellow-at* relation to t. But that of course never happens. And of course this is not what we get in the canonical version of the relational theory; instead, according to this version, the banana bears incompatible relations to *different* times, the *green-at* relation to t and the *yellow-at* relation to t'. But this is no more change for the banana than for someone to like Tom at t and dislike Tim at t'. After all, someone can like Tom and dislike Tim at the same time. Thus, bearing the *green-at* relation to t and the *yellow-at* relation to t' is no change since the banana bears those relations to different times. Indeed, since those relations are borne to different times, they can be and are borne at the same times: at both t and t', for instance, the banana bears the *green-at* relation to t and the *yellow-at* relation to t'. Thus, the canonical version of the relational theory does not account for change of allegedly intrinsic properties.

For similar reasons the canonical version cannot account for relational change either. For it has it that the change of a and b from being 2 miles apart at t to their being 1 mile apart at t' consists in a three-place relation of *being 2 miles apart* holding between a, b and t at t and a three-place relation of *being 1 mile apart* holding between a, b and t' at t'. But this is no change, since these relations can hold at the same time; indeed, there is no more incompatibility here than in a and b being south of c and a and b being north of d.

It has been suggested to me that the relational theorist could reinterpret the notion of relational change so as to allow for change when incompatible relations are borne to different entities, provided these are *times*. Thus, on this account, bearing the *green-at* relation to t and the *yellow-at* relation to t' would count as a change. But this is clearly an ad hoc manoeuvre whose true effect is to remove all credibility from the relational theory. For why should times be such special relata that, unlike other relata, they need not remain fixed for there to be relational change? If all change is relational change, as the relational theory has it, how can a thing undergo relational change by bearing an incompatible relation to a *different* entity, even if this is a time? There appear to be no convincing answers to these questions. Perhaps the way to make more plausible the ad hoc manoeuvre here discussed would be to argue that times are really not relata of *green-at*, *yellow-at* and the like, and that these are really not relations. But then nothing remains of the relational theory.

Thus, the canonical version fails to account for change, both intrinsic and relational. But perhaps the other versions of the relational theory are immune to this objection? The account of change given by the relational property version is exactly the same as that of the canonical version, since they differ only in that for the canonical version something like *being green* is a relation to a time while for the relational property version it is a relational property had in virtue of a relation to a time. Thus, the relational property version also says that for the banana to pass from being green at t to being yellow at t' is for the banana to bear the *being green-at* relation to t and to bear the *being yellow-at* relation to t'. Similarly, it says that for a and b to pass from being 2 miles apart at t to being 1 mile apart at t' is for them to stand in the three-place relation of *being 2 miles apart* with t and in the three-place relation of *being 1 mile apart* with t'. But this, as we saw, is no change and so the relational property version also fails to give a correct account of change, both intrinsic and relational.

Let me now consider the instantiation version in its three variants. According to the first variant, for the banana to pass from being green at t to being yellow at t' is for the three-place *instantiation* relation to hold between the banana, the property of *being green* and t at t and the three-place *instantiation* relation to hold between the banana, the property of *being yellow* and t' at t'. But there is no change here, since these relations can both hold at the same time; what cannot hold at the same time are the three-place *instantiation* relation between the banana, *being green* and t, and the three-place *instantiation* relation between the banana, *being yellow* and t. *Mutatis mutandis* in the case of relations. Thus, the first variant of the instantiation relation fails in its account of change, both intrinsic and relational.

Similarly for the other two variants of the instantiation version. Bearing the relation of *instantiation* to the time-indexed property *green-at-t* is not incompatible with bearing the relation of *instantiation* to the property *yellow-at-t'*. The incompatibility is, of course, between instantiating *green-at-t* and instantiating *yellow-at-t*. Similarly, bearing the *instantiating-at-t* relation to the property of *being green* is not incompatible with bearing the *instantiating-at-t'* relation to the

property of *being yellow*. The incompatibility is, again, between instantiating-at-*t* the property of *being green* and instantiating-at-*t* the property of *being yellow*. *Mutatis mutandis* for relations in both cases.

In conclusion, the relational theory fails to account for change, both intrinsic and relational; for change is having incompatible properties at different times or bearing incompatible relations, like *green-at* and *yellow-at*, *to the same entities* at different times. But on the relational theory incompatible relations like *green-at* and *yellow-at*, or *hot-at* and *cold-at*, etc., are borne to *different* entities, namely times. This is the simple reason why the relational theory fails to solve the problem of change.[9]

Notes

1 Incompatible in the weaker sense that they cannot both be possessed at the same time by the same thing. Perhaps 'contrary' would be a better word, but since 'incompatible' is more widely used in this context I shall stick to it. But unless we have a trivial notion of property, there may be cases of change when no incompatible properties are involved. Imagine the case of a person that changes from having one finger at time *t* to having two fingers at time *t'*. The properties of *having one finger* and *having two fingers* are not incompatible, since anything having the latter also has the second. So, unless we are prepared to admit properties like *having one finger and no more*, this seems to be a case of change without incompatible properties. Whether or not this is so, the paradigmatic cases of change are such that they consist in having incompatible properties (in the weaker sense) at different times. Surely if a theory cannot account for these paradigmatic cases of change, it is not a good theory of change. I shall henceforth speak, for simplicity, as if change always consisted in having incompatible properties at different times.

2 The phrase 'problem of temporary intrinsics' comes, as far as I know, from Lewis (1986: 203), but he is well aware that there is a problem of relational change (1999: 192–3).

3 The version I shall have primarily in mind is the canonical version, although I shall refer occasionally to the other versions. I introduce the other versions for the sake of comprehensiveness and to show that my argument against the relational theory is general and applies to all the versions that I know.

4 Haslanger (1989: 120, 122–3) says she advocates a version of what Lewis calls the adverbial version but makes clear that what she defends is hardly a version of the relational theory at all.

5 Mellor and van Inwagen describe their versions of the relational theory only with respect to properties, not relations. Similarly, Lewis describes (but does not defend) the first variant of the instantiation version only with respect to what it says about properties. This might be another feature of the generalized neglect of relational change. I have extended the versions of the relational theory to cover relations as well.

6 A new paper by Lewis (2002) on the subject has appeared, in which he gives further arguments against what he had called the *adverbial variant* of the relational theory (Lewis 1999: 188, fn. 1). But he still insists, without further argument, that certain properties are monadic and intrinsic and no relations to times: 'Even the properties *bent* and *straight* could at least sometimes be monadic: for instance, when they are properties of momentary things' (Lewis 2002: 4).

7 Why does Hawley say '*If* the distance … '? She tells me that she does not doubt it, but she was trying to be careful. In any case, if the distance between the parts of a thing is not an intrinsic part of it then the conclusion to draw is not that distance is not an external relation, but that Hawley's proposed definition of external relations should be abandoned, for, as Hawley would admit, spatial distance is a sort of paradigm of external relation.

8 There is also a problem with Hawley's definition of external relations, since it presupposes a fairly generous view about composition. Indeed, it seems to presuppose that mereological composition is unrestricted and that for every two things x and y there is a third, namely $x + y$. Perhaps she does not need such a strong thesis, but for her arguments to go through she at least needs the still strong thesis that for every two things x and y that can stand in an external relation to each other there is a third entity, i.e. $x + y$. And why must anyone admitting external relations be committed to *any* view on composition?

9 For comments on previous versions of this chapter I thank audiences at the Universities of Cambridge, Edinburgh and Sheffield and, especially, Hugh Mellor. I also thank the Leverhulme Trust, whose Philip Leverhulme Prize allowed me to find the time to finish the chapter.

References

Armstrong, D. M. (1980) 'Identity through time', in P. van Inwagen (ed.) *Time and Cause: Essays Presented to Richard Taylor*, Dordrecht: D. Reidel: 67–78.

Forbes, G. (1987) 'Is there a problem about persistence?', *Aristotelian Society* 61 (Suppl.): 137–55.

Haslanger, S. (1989) 'Endurance and temporary intrinsics', *Analysis* 49: 119–25.

Hawley, K. (1998) 'Why temporary properties are not relations between objects and times', *Proceedings of the Aristotelian Society* 98: 211–16.

Hinchliff, M. (1996) 'The puzzle of change', *Philosophical Perspectives* 10: 119–36.

Johnston, M. (1987) 'Is there a problem about persistence?', *Aristotelian* 61 (Suppl.): 107–35.

Lewis, D. (1986) *On the Plurality of Worlds*, Oxford, UK, and Cambridge, MA: Basil Blackwell.

—— (1999) 'Rearrangement of particles: reply to Lowe', in *Papers in Metaphysics and Epistemology*, Cambridge, UK: Cambridge University Press.

—— (2002) 'Tensing the copula', *Mind* 111, 441: 1–13.

Mellor, D. H. (1981) *Real Time*, Cambridge, UK: Cambridge University Press.

—— (1998) *Real Time II*, London: Routledge.

Merricks, T. (1994) 'Endurance and indiscernibility', *The Journal of Philosophy* 91: 165–84.

van Inwagen, P. (1990) 'Four-dimensional objects', *Noûs* 24: 245–55.

13 Presentism

A critique

L. Nathan Oaklander

The problems of time and change are inextricably connected for change involves time and, Shoemaker (1969) notwithstanding, time involves change, or so McTaggart (1934; 1968) has argued. That they are related is not in doubt; how they are related is. For McTaggart they are related in such a way that if there is to be time and change, then there must be an A-series, and temporal becoming, but what is the A-series? And what is temporal becoming? These are not easy questions to answer, because there are many different versions of A-time and temporal becoming, and I do not intend to discuss them all. Rather, my aim will be to focus on one version of A-time, the presentist version, and argue that, contrary to its recent proponents, it does succumb to McTaggart's paradox.[1] Even within the limited scope of this chapter, the task of refuting presentism is complicated by there being several different versions of it. One would not think that this is so because all presentists maintain that only the present exists, whereas the past and the future do not exist. Nevertheless, there are different presentist versions of the A-theory, and, although I believe that in one way or another they are *all* susceptible to McTaggart's paradox, there is only one version that I shall endeavour to refute, namely that propounded by William Lane Craig in his recent trilogy on time: *The Tensed Theory of Time*: *A Critical Examination* (2000a), *The Tenseless Theory of Time*: *A Critical Examination* (2000b) and *Time and the Metaphysics of Relativity* (2001).

I chose Craig's defence of presentism for two reasons. First, A-theorists who follow Prior in adopting a presentist 'metaphysic' are often criticized for lacking an ontology (see, for example, Oaklander 1984: 90–2; Smith 1993: 158–69; 1994a; 1999: 248–9; 2003; Tooley 1997: 165–70, 232–8). To say that the tenses do not refer to B-relations and do not ascribe A-properties is one thing, to say what then are the ontological correlates of the tenses is quite another. It is the latter task that Prior and his followers are commonly accused of shirking. Craig is an exception. He is sensitive to the 'lack of ontology' criticism of Prior-based theories (see Craig 2000a: 192–4), and attempts to 'found' or provide an ontological ground for both B-relations and A-determinations in the A-series, 'tensed facts' and temporal becoming. For that reason, he provides his readers with a metaphysical theory to be evaluated.

I have a second reason for choosing to discuss Craig's version of presentism. Presentists typically explain, promote and defend their view as being the temporal analogue of the serious actualist position with respect to possible words according to which only the actual world is real. By examining Craig's presentist metaphysic we can evaluate just how successful the marriage between presentism and actualism is. Since my ultimate goal is to argue that presentism, or at least Craig's version of it, is not immune from McTaggart's conundrum, I shall begin my discussion of Craig with an examination of McTaggart.

According to McTaggart, we ordinarily (or commonsensically) conceive of time as involving the notions of past, present and future (A-determinations) and earlier than/later than and simultaneous with (B-relations). Although McTaggart claims that the A-series (defined in terms of A-determinations) and the B-series (defined in terms of B-relations) are both essential to our *ordinary* conception of time, he believes that A-determinations and the A-series are more fundamental, more ultimate and more essential to the *ontological* nature of time than B-relations and the B-series. In fact, his view is that the B-series is dependent on the A-series, not only because there would be no B-relations unless there were A-determinations, but more fundamentally because the B-series is ontologically reducible to the A-series and the non-temporal C-series. Thus, McTaggart claims that while the A-series and the C-series are each ultimate

> [t]he B series, on the other hand, is not ultimate. For given a C series of permanent relations of terms, which is not in itself temporal and therefore is not a B series, and given the further fact that the terms of this C series also form an A series, … it results that the terms of the C series become a B series, those which are placed first, in the direction from past to future, being earlier than those whose places are farther in the direction of the future.
>
> (McTaggart 1934: 118)

I think that this passage makes it clear that, for McTaggart, there are no ontologically primitive or simple temporal relations. Metaphysically, time is entirely constituted by the A-series, and it together with the non-temporal but ordered C-series grounds the commonsense view of time as involving *both* A-determinations and B-relations.

Given his positive ontology of time, McTaggart's negative thesis can be recast by saying that, while the A-series and the C-series are necessary and sufficient for the existence of B-time, they are not sufficient for A-time or B-time, which is a contradiction. For time requires *change* and the A- and C-series cannot account for change without introducing some metaphysical correlate of temporal becoming. However, there is no consistent, non-circular way to metaphysically interpret temporal becoming so that change is not

contradictory.[2] Since, for the A-theorist, B-time requires temporal becoming, and temporal becoming is contradictory or viciously circular, it follows that there is no B-time, and without B-time there is no time at all.

With this background we are ready to turn to Craig's discussion of McTaggart's paradox and his exposition of the metaphysics of presentism. Craig has basically two responses to McTaggart. The first is to claim that McTaggart mistakenly treats temporal becoming 'as a sort of qualitative change insofar as he attempts to combine a B-theoretic ontology with A-theoretic becoming' (Craig 2000a: 179). On the pure A-theory that Craig adopts, 'past and future events/things/times are not real or existent and, hence, do not exemplify properties like pastness or futurity. Rather, entities come to be and pass away absolutely, so that the only temporal entities that there are are the present ones' (Craig 2000a: 179). Craig's response gives rise to three central questions:

(1)   If temporal becoming is not to be understood as a species of qualitative change, then how is it to be understood?
(2)   If McTaggart mistakenly combines a B-theoretic ontology with the A-theory, how then does Craig attempt to analyse temporal relations between and among items existing in the B-series?
(3)   If the past and the future do not exist, then what are the truthmakers of past- and future-tense statements?

This last question becomes particularly important given his second response to McTaggart's paradox.

Craig's second objection is that McTaggart's model of treating temporal becoming as the donning and doffing of the non-relational temporal properties of *pastness* and *futurity* is erroneous, because there are no such properties. To say that an event 'is past' or 'is future' is not to attribute a property to the event. 'Rather such ascriptions should be parsed as asserting that the entity in question did or will exist; …' (Craig 2000a: 190).[3] Of course, to parse attributions of pastness and futurity in terms of statements about what *did* or *will* exist, or in terms of what *was* or *will be* the case, does not answer the question of how we are to understand grammatical ascriptions of pastness and futurity, but just raises it once again. For that reason, the account of the truthmakers of past- and future-tense statements in terms of 'what is non-relationally present' is an ever-pressing concern.

It should be clear, therefore, that Craig's critique of Mellor's (1998: 70–8) version of McTaggart's problem is ineffectual. Craig repeats the familiar point against McTaggart, Mellor and others, that no event has all three A-determinations timelessly or simultaneously but successively, and he reflects this by saying that no matter what level we start at we get a consistent set of propositions.[4] Suppose we start with

3.   FP*e* & N*e* & PF*e*

(Craig 2000a: 203)

This is read as '*e* will be past' and '*e* is present' and '*e* was future'. Craig claims there is no contradiction in 3. Perhaps not, but we are still left with the question: What are the truthmakers for the first and last conjunct? More specifically, what is the ontological difference between FP*e* and PF*e*, given that neither 'F' nor 'P' is a predicate that ascribes properties to *e*? Unless we are told, we cannot tell.[5] Without such an account, however, the appeal to grammatically consistent tensed statements is a vacuous response to McTaggart's paradox or Mellor's formulation of it.

The need for an account of the passage of time or temporal becoming is also urgent, and for basically the same reason. To see why consider the following passage:

> In his 'McTaggart's Paradox Revisted,' *Mind* 101 (1992): 323–326, Lowe synthesizes the A-theorist's position by saying that every event is such that it is or was or will be truly describable as past, and is or was or will be truly describable as present, and is or was or will be truly describable as future, which he symbolizes as
>
> 6**. (NT 'Np' v PT 'Np' v FT 'Np') & (NT 'Pp' v FT 'Pp' v FT 'Pp') & (NT 'Fp' v PT 'Fp' v FT 'Fp').
>
> … surely (6**) does represent the passage of time, since the same tense operator in each conjunct cannot operate on the true disjunct, on pain of contradiction, so that differently tensed statements will be true in each conjunct. *This difference in tense does represent the flow of time*.
>
> <div align="right">(Craig 2000a: 205; emphasis added)</div>

The appeal to truth predicates does not avoid the need to specify the grounds of truth. The fact that differently tensed statements will be true in each conjunct cannot adequately reflect the passage of time unless we have some account of the direction of becoming. More specifically, if NT 'Np' & PT 'Fp' & FT 'Pp' then we want to know, given that the past and the future do not exist, what is the difference between PT 'Fp' & FT 'Pp'? What is the basis, in the metaphysics of presentism, for *p* being first future and then present and then past rather than the other way around? To answer that question we need some model upon which to understand temporal becoming.

Craig's explication of temporal becoming begins with an appeal to the serious actualist's conception of possible worlds as states of affairs that exist as abstract objects but are not instantiated. He then claims that 'tensed possible worlds which *did*, *do*, or *will* obtain are tensed actual worlds' (Craig 2000a: 209; emphasis added). Of course, the appeal to tensed possible worlds which *did*, *do*, or *will* obtain can hardly provide a metaphysical explanation of what the tenses stand for in propositions reflecting temporal becoming. Leaving that difficulty aside for the moment, Craig (2000a: 209) continues by saying that 'the tensed actual world at *t*, is the world which obtains when *t*'s *being present*

obtains, or more simply, when *t* is present', but *when* does *t*'s being present obtain? Judging from his comments it appears that *t*'s being present obtains *before t**'s being present obtains (for any later *t**), since Craig maintains that '[t]ensed actual worlds constitute the tensed history of the actual world α, for they are respectively comprised of all states of affairs entailed by α and each *successive t*'s being present' (Craig 2000a: 209; emphasis added). Thus Craig's view is that there are possible worlds that exist whether they are instantiated or not, and as time flows possible worlds obtain or become actual by being *successively* instantiated. That the appeal to succession is integral to Craig's account of becoming is evident from other passages as well.

For Craig, temporal becoming is modelled on the different members of the A-series coming into existence successively, as successive times become present. He says, 'the doctrine of objective becoming, … could be graphically displayed as the *successive actualization of the history of the actual world*. It is this model of a *successively instantiated*, rather than tenselessly existing, actual world that precludes the existence of a "totality of facts"' (Craig 2000a: 207; emphasis added). The appeal to succession implies the existence of temporal relations, and the appeal to possible worlds that *did* or *will obtain* implies the existence of past- and future-tense facts. Craig's prima facie commitment to B-relations and primitive past- and future-tense facts renders his version of 'presentism' subject to McTaggart's paradox unless he can provide an ontological reduction of temporal relations and past- and future-tense facts to what is presently real. Thus, we are led once again to the question: What then, on a presentist metaphysics, are temporal relations, and what are the past- and future-tense facts that are the truthmakers of past- and future-tense statements?

Craig does attempt to answer these questions, and in so doing he diverges in many ways from temporal solipsism, 'an idiosyncratic doctrine associated with the views of A. N. Prior and not logically connected with the A-Theory of time' (Craig 2000a: 214). One of the main ways in which Craig deviates from Prior's version of presentism is in his holding that there are past- and future-tense facts that are the truthmakers for past- and future-tense statements. I will let Craig speak for himself:

> On the presentist semantics given here, a future-tense statement is true iff there exists some tensed actual world at *t* in which the present-tense version of the statement is true, where *t has not elapsed by the present moment*. A past-tense statement is true iff there exists some tensed actual world at *t* in which the present-tense version of the statement is true, where *t has elapsed by the present moment*. Those are the truth-conditions of past- and future-tense statements; but they are not what make the statements true. Ultimately what makes the statements true is that reality *was* or *will be* as the statements describe; when the time comes, for example, a sea battle is going on, and therefore the statement made the day before, 'There will be a sea battle tomorrow,' was true. There are tensed facts corresponding

to what tensed statements assert, but past- and future-tense facts exist because of the present-tense facts that *did* or *will* exist.

<div align="right">(Craig 2000a: 213–14; emphasis added)</div>

For Craig there are past- and future-tense facts, but they exist because purely present-tense facts, for example *a battle is being fought at Waterloo*, *did* or *will* obtain. Alternatively, a fact is a future-tense fact if the *time t* at which it is present *has not elapsed by the present moment* (that is, *t is later than* the present moment), and a fact is a past-tense fact if the *time t* at which it is present *has elapsed by the present moment* (that is, *t is earlier than* the present moment). Thus, Craig's account either presupposes the existence of irreducibly past- and future-tense facts, or it assumes the existence of B-relations, or it leaves the tenses unanalysed and so is guilty of the 'lack of ontology' objection he and others have raised against Prior and his followers.

Look at it this way. On the one hand, Craig wants there to presently exist truthmakers for past- and future-tense statements. If a statement is *true now* then it must be true in virtue of some fact that *exists now*. On the other hand, he does not want to countenance past and future existents. He attempts to avoid the contradiction that a conjunction of those two views entails by claiming that past- and future-tense facts exist at present, but they are not ultimate. However, his attempt to show that past- and future-tense facts are not ultimate is either unsuccessful or it succeeds only at a cost of reintroducing a B-theoretic ontology that he sought to avoid, thus undermining presentism and making his A-theory susceptible to McTaggart's paradox.

We can begin to see why this is so by noting that Craig claims that, if a past-tense statement is now true, then there is a present-tense fact that *did* obtain or there is a present-tense fact that exists at a time *t* that has elapsed by the present time. What, then, is involved in *t*'s having elapsed by the present moment, or a present tense fact having obtai*ned*? I can think of several possibilities:

(1)   The present moment is moving across the A-series of presently existing things/events/moments, and a present-tense fact *did exist* when the moving NOW has passed it by. So what exists now is the fact that the NOW (as a relation to a term outside the series or as a monadic property) has already passed (or has already been exemplified) by a given instantiated state of affairs, and that fact is the ground of the past-tense fact that X *was* F.

(2)   To say that a present-tense fact did or will obtain at a time that has or has not elapsed by the present moment is to countenance the existence of presently obtaining primitive past- and future-tensed facts, X *was* F and X *will be* F.

(3)   If a past-tense statement is true, then there presently obtains the fact that a present-tense state of affairs exists at a time *t earlier than* the present moment *t**. Similarly, if a future-tense statement is true, then there

presently obtains the fact that a present-tense state of affairs exists at a time *t later than* the present moment *t**.

(4)   Finally, one can eschew ontology altogether and claim that the tenses are logical operators, or that the tenses and temporal becoming are conceptually primitive, and have no ontological significance whatsoever.

Clearly, the first two alternatives are unacceptable. The first involves a view of temporal becoming that McTaggart and many others, including Craig, have found reasons to reject.[6] The second is inconsistent with Craig's presentism, since if there are ultimate past- and future-tense facts then temporal objects must exemplify the properties of *pastness* and *futurity* and therefore must, in some sense, exist. The last alternative (4) is also explicitly rejected by Craig, who construes his version of presentism as providing an 'ontological foundation' for temporal relations and the direction of time.

There remains the third interpretation, although it too raises questions. If only the present exists, then how can there presently obtain a temporal *earlier than* or *later than* relation between two temporal objects at least one of which does not exist? Nevertheless, there is reason to believe that Craig adopts the alternative (3), which analyses past- and future-tense facts in terms of what is *earlier* or *later than* the present moment, since he expresses sympathy with such a view about the ontological status of the past and future put forth by Alfred Freddoso (1983). Freddoso maintains that 'the proposition "Socrates drank hemlock" is now temporally necessary, since "Socrates drinks hemlock" is a member of a past submoment which obtains *prior* to the present in any world sharing the same history *prior* to the present with our world … .' (Craig 1991: 180; emphasis added; see also Craig 2000a: 214, fn. 140). And referring to future-tense propositions Freddoso says, 'a proposition *p* is necessary *per accidens* at *t* in world *w* just in case *p* is true at *t* and at every moment *after t* in every possible world which shares the same history … with *w* at *t*' (Freddoso 1983: 266; emphasis added, quoted in Craig 1991: 180). The appeal to 'prior' times implies a temporal relation between a past event or time and the present, and the statement 'every moment *after t'* implies a temporal relation between a later event or time and the present moment. If, however, Craig appeals to *unanalysable* temporal relations to account for the truthmaker of past- and future-tense facts, then Craig contradicts himself, since he claims that a B-theoretic ontology coupled with A-theoretical becoming renders McTaggart's paradox inescapable. It is not surprising, then, that he attempts to provide an ontological reduction of B-relations in terms of A-determinations, the A-series and temporal becoming (Craig 2000b: 149–58). In the final part of this chapter I shall critically examine Craig's attempt.

In *The Tenseless Theory of Time: A Critical Examination*, Craig (2000b) agrees with McTaggart's positive view of time that 'on the A-Theory of time, the obtaining of the temporal relations *earlier than/later than* among temporal particulars can be derived from the objectivity of A-determinations and the

A-series' (Craig 2000b: 150). Paradoxically, Craig interprets Mellor as also maintaining that 'the very temporal relations which lie at the heart of the B-Theory are derivable from the A-series [and A-determinations]' (Craig 2000b: 151). It is true that Mellor offers various possible reductions of the B-series to the A-series, but there are two important facts to note about his 'definitions.' First, they presuppose the existence of McTaggart's A-series and A-determinations. More specifically, on Mellor's interpretation of the A-theory, and temporal becoming, 'Futurity, temporal presence, and pastness are all supposed to be real non-relational properties that everything in time successively possesses, changing objectively as it exchanges each of properties for the next' (Mellor 1981a: 89–90). Craig explicitly rejects this interpretation of the A-theory, arguing, as does Mellor, that it leads inevitably to McTaggart's paradox.

Second, Mellor has argued that McTaggart has shown that A-change is contradictory and thus the A-theory of temporal becoming is absurd. As he puts it, 'What disproves all A-theories is a contradiction inherent in their concept of change' (Mellor 1998: 70). Thus, although Mellor would agree that *if* the A-theory is true then B-relations could be defined in terms of the A-series, the point is moot since the A-theory is false. Clearly, Mellor does not believe that B-relations can be defined in terms of A-determinations. Craig, on the other hand, claims to be defending an ontological reduction of temporal relations that 'goes all the way back to McTaggart' (Craig 2000b: 150), but the analyses that McTaggart and Mellor propose imply that A-determinations are either properties of events/moment/things or relations to some term outside the temporal series. Craig (2000a) denies the existence of A-determinations as characteristics of events/things/moments, whereas in Craig (2000b) the definitions of B-relations he offers require those properties. Thus, his appeal to Mellor's definitions to support an ontological reduction of B-relations to A-determinations is inconsistent with his presentist metaphysic according to which there are no such properties. Furthermore, I shall argue that, in adopting the A-account of B-relations endorsed by McTaggart and spelled out by Mellor, Craig's analysis of temporal relations does not avoid the difficulties McTaggart raises since he is committed to a theory that is contradictory, circular or vacuous.

Craig's first ontological reduction of *earlier than/later than* relations is as follows:

D$_1'$:   *e is* earlier than e^* $\equiv e$ is more past or less future than e^*.
      e is later than e^* $\equiv e$ is more future or less past than e^*.

<div align="right">(Craig 2000b: 153)</div>

According to Craig, *more past/future than* are A-relations and not monadic properties. They are relations that presently obtain between terms that occupy different positions in the A-series.

> Thus, for example, if *e is* earlier than *e** and it happens that *e* is present, then *e* is less future than *e**. Similarly, in the case that one of the events is past and the other future, we should think of each one as having none of the A-determinations of its *relatum*. Thus, for example, if *e* is past and *e** is future, then *e is* earlier than *e** just because *e* is more past than *e**.
>
> (Craig 2000b: 153)

Craig (2000b: 154) maintains that *more past* and *more future* are primitive concepts. What, then, are its relata? And if the *relatum* of an A-relation are *e* is past and *e* is future, then what is the ontological status of those relata? Clearly, if *being future* and *being past* are non-relational properties of past and future events, then his view is inconsistent with presentism and, by his own lights, susceptible to McTaggart's paradox (see Craig 1998). On the other hand, if past and future events have no ontological status, so that neither *e*'s *being past* nor *e*'s *being future* exists, then we have an A-relation without relata, which is absurd. Finally, if Craig attempts to analyse *e* is past and *e* is future in accordance with the possible worlds analysis he offered previously (Craig 2000a), then the truthmaker for, say, 'It was raining' is that the present-tense fact *It is raining* obtains at a moment of time *t* that has elapsed by the present moment. In that case, however, his ontological reduction of so-called B-relations is obviously circular, since, as we have seen, there is no acceptable account of 'time *t* has elapsed by the present moment,' other than that time *t is earlier than* the present moment.

Furthermore, his account of relations, sketchy as it is, raises serious problems concerning his notion of A-relations. He says:

> relations are abstract objects which plausibly do not exist in time at all. Non-contemporaries stand in a relation at their respective times and the timelessly existing relation reaches across time to relate the two individuals. As for the individuals themselves, we could ascribe to them relational properties: Socrates, at the time he existed, had the property of *going to be referred to by William Craig* or the property of *being referred to by William Craig at t_n*. He no longer has that property, but I now have the property of *referring to Socrates*. The *relation between us can be analyzed in terms of such relational properties or said to exist timelessly in virtue of such properties*.
>
> (Craig 2000a: 212; last emphasis added)

The first problem with this account of relations is that it is incompatible with his account of A-relations, and his presentist ontology. Craig claims that 'the A-theorist is at liberty to stipulate that the above concepts [*more past* and *less future*] are among his theoretical primitives …' (Craig 2000a: 154). Perhaps so, but if A-relations are theoretical primitives, then they cannot be analysed in terms of relational properties, or be said to exist in virtue of such properties. On the other hand, if A-relations are analysable in terms of *relational* properties, then what else could they be if not parasitic on the B-relations that he is

attempting to analyse. Finally, the very notion that we could treat *being past* and *being future* as *relational properties* of the terms of A-relations contradicts his previous claim that 'The construal of pastness and futurity as relational predicates *should not be taken to mean that these are relational properties inhering in events*' (Craig 2000a: 190; emphasis added). Given the inconsistency between Craig's accounts of relations in general and temporal A-relations in particular, it is debatable whether or not there are any terms of A-relations or, indeed, whether there are any A-relations at all.

To see what is involved in this last point, note that Craig claims that 'If the relations *earlier than/later than* can be truly and tenselessly ascribed, it is because and only because the A-relations *more past/future than* and *less past/future than* can be truly and presently ascribed' (Craig 2000a: 152). My question is this: What is the ground of the truth of statements that assert the existence of an A-relation between events/things/moments, and what are A-relations presently ascribed to? What presently exemplifies those relations? I do not think that Craig has a consistent set of answers to these questions. Since past and future things/events/moments do not exist, they cannot be the terms of A-relations nor can A-relations ascribe a tense to *them*. For the presentist, what does not presently exist cannot presently exemplify properties, including tensed properties. But then, since A-relations, like all relations, are timeless, there is nothing that presently exists that could provide an ontological foundation for affirming the existence of the (tenseless) temporal relations of *earlier than/later than*. To put the difficulty otherwise, Craig is faced with a dilemma. If past and future temporal objects do not exist, then there is nothing for A-relations to presently ascribe objective tense to. If past and future temporal objects do exist, then presentism is false. Thus, given Craig's version of the A-theory, whether there are past and future temporal objects or not, there are no A-relations, there are no B-relations and there is no time. For these reasons, his first reductive analysis of B-relations is unsuccessful.

Craig's second definition or reduction of B-relations to A-determinations is also unsuccessful. His second definition is as follows:

D_2'': *e is* earlier than e* $=_{df}$ There is some time *t* such that at *t* it is an objective fact that *e* has presentness and e* is future.

(Craig 2000b: 156)

e is later than e* $=_{df}$ There is some time *t* such that at *t* it is an objective fact that *e* has presentness and e* is past.

(Craig 2000b: 156)

Richard Gale has claimed that to relativize A-determinations to times in this way is circular because 'The predicates "___ *is* past at ___" and "___ *is* future at ___" … express a timelessly true or false statement about a B-relation between

two events, i.e. they make B-statements' (Gale 1968: 90–1; quoted in Craig 2000b: 155). Craig dismisses Gale's claim on the grounds that (1) a tenselessly true statement such as the *definiens* of D_2'' can refer to an A-determination and (2) being 'at a time' does not 'illicitly smuggle in the so-called B-relation of *simultaneous with* … [since] being 'at a time' is foundational to the notion of simultaneity, rather than the other way around' (Craig 2000b: 156, fn. 22). Craig concludes that the '*definiens* thus should not be construed in terms of the ascription of any so-called B-relations' (Craig 2000b: 156–7).

There are two problems with Craig's second definition and his response to the objections. First, Craig never specifies what are the ontological correlates of '*e** is past' and '*e** is future' in the two *definiens*' of D_2''. He does say that 'According to (D_2''), at *t e* has the premier A-determination of presentness, and the *definiens* in each case refers to an objective tensed fact … [and] therefore refers to an A-determination' (Craig 2000b: 156), but what are the objective tensed facts in this case? If '*e** is past' and '*e** is future' attribute A-determinations (properties) to *e**, then the past and future must exist in order to exemplify those properties, and that is incompatible with his professed presentism. On the other hand, if the past and the future do not exist so that past- and future-tense facts are not ultimate, but analysable in terms of present-tense facts that have or have not yet elapsed, then his analysis is circular, since there is no analysis of 'time *t* has elapsed by the present moment' that is both consistent and does not reintroduce B-relations, and this leads to another problem with his second reductive analysis of B-relations.

Craig misses the main point of Gale's charge of circularity. The introduction of time and, in Craig's case, absolute time or moments is crucial if we are to avoid a contradiction.[7] For if the *definiens*' are, as Craig says, tenselessly true statements, then, in order to avoid a contradiction, time must be included in what the *definiens*' of D_2'' express. Otherwise we would get

> *e is* earlier than e* $=_{df}$ It is an objective fact that *e* has presentness and *e** is future.

and

> *e is* later than *e** $=_{df}$ It is an objective fact that *e* has presentness and *e** is past.

Obviously, those two objective facts contradict each other and fail to account for whether the direction of time is from *e* to *e** or from *e** to *e*. The introduction of some time *t* and *t** at which the objective facts mentioned in the *definiens* are at provide such an account *if and only if t* and *t** are members of a *temporal* sequence, that is a sequence with an intrinsic direction. Thus, even if the predicates 'is past at', 'is present at' and 'is future at' do not presuppose the existence of B-relations, Craig's analysis is still circular because the existence of 'at time *t*' in his analysis does presuppose the existence of B-relations.

Finally, let us turn to Craig's third attempt to ground the existence of B-relations on the reality of tensed facts.

D_3': *e is* earlier than *e** $\equiv$ *e* becomes present first and *e** becomes present second.

e is later than *e** $\equiv$ *e** becomes present first and *e* becomes present second.

Craig considers one objection to this account and his reply is telling.

> Oaklander objects that the use of 'first' and 'second' conceal so-called B-relations (Oaklander 1996: 211); but a moment's reflection shows that this is not the case. There are ordinal numbers that are wholly atemporal and can characterize spatial or abstract objects as well as temporal particulars. Given the order of their temporal becoming, the temporal ordering of the two events in question necessarily follows.
>
> (Craig 2000b: 157)

Admittedly, *spatial* or *abstract objects* can be characterized as 'first' and 'second' without presupposing temporal relations, but it does not follow that 'first' and 'second' can characterize *temporal objects* without presupposing temporal relations. Indeed, temporal relations between and among particulars are intrinsically different from all other instances of one-dimensional order, such as that of points on a line and numbers in order of magnitude, in that only a temporal series has an intrinsic direction. The terms 'first', 'second', 'third', and so on, can give a spatial series an *order*, but they cannot give spatial objects a *direction*. For that reason, to say that *e* becomes present *first* and *e** becomes present *second* is either irrelevant to determining their B-relation to one another or assumes that *e* and *e** become present in a given direction; it does not account for it. Thus, Craig's third account of B-relations is circular unless we eliminate the 'first' and 'second' from it. In that case, however, Craig's analysis is inadequate since from *e* becomes present and *e** becomes present we cannot infer that e *is* earlier than *e** or vice versa.

One last, related, criticism. Suppose that *e* becomes present first, *e** becomes present second, *e*** becomes present third, *e**** becomes present fourth, and so on. Since this conjunction is tenselessly true the *definiens* in D_3' leaves out the information about which event is present NOW? Indeed, what could be the ground of the *definiens* being true now, unless there is the further fact that *e** *is present NOW*. However, since to become present is an act of a temporal being, it follows that all of the terms in the A-series obtain (present tense) at some time. But what accounts for their direction? Which events become present *before* the others? Unless the NOW moves *successively* along the series of events that obtain (present tense) at some time or other, there is no change, and without change there is no time. Unfortunately, the notion of the *successive* actualization of the terms of the A-series presupposes precisely what Craig is attempting

to analyse, namely B-relations. For that reason, Craig's analysis is viciously circular and the circularity cannot be avoided by positing another A-series of events or times at which the terms of the first series undergo becoming on pain of a vicious infinite regress.

According to Craig, 'the A-theorist can account for the existence of so-called B-relations by *founding them on the reality of tensed facts*; thus far, *McTaggart's argument seems to be vindicated*' (Craig 2000b: 157). The problem is twofold: first, McTaggart's view on B-relations implies the existence of the A-series either as a series of terms that have an A-relation to a term outside the series or as a series of terms that have the A-properties of *pastness*, *presentness* and *futurity*. Thus, in so far as McTaggart's view is vindicated, Craig's presentist metaphysics is refuted, since the two are incompatible, and if his reductive analysis of temporal relations depends on McTaggart's positive view of time being vindicated, then his analysis is refuted once again. Second, if Craig rejects McTaggart's view of the A-series and temporal becoming, then it is unclear how he has accounted for the existence of B-relations by founding or ontologically grounding them on the reality of tensed facts, because it is not clear what tensed facts exist on a presentist metaphysics. If the only tensed facts there are are present-tense facts – those that exist NOW – then there are no present-tense facts that could ground the truth of statements about what is earlier or later than now or about what is past or future. Clearly, the appeal to present-tense facts that *did* obtain, or *will* obtain, or to present-tense facts that obtain at a time that *has or has not elapsed by the present moment* is either to eschew ontological commitment altogether or to appeal to precisely those past- and future-tense facts or B-relations that Craig sought to avoid. In any case, on Craig's presentist version of the A-theory, time is unreal.

Craig claims to 'have an ontological foundation in [his] metaphysic of time for affirming the existence of the (tenseless) temporal relations *earlier than/later than*' (Craig 2000b: 159). On the basis of my critique of Craig's metaphysics of presentism, it would appear that he has not provided an ontological foundation for temporal relations. I conclude that Craig's A-theoretical account of time, change and becoming *is* subject to McTaggart's paradox and must therefore be rejected.

Acknowledgement

Notes

1   For recent defences of presentism see, for example, Prior (1968), Chisholm (1981), Bigelow (1991; 1996), Hinchliff (1996), Craig (1998), Zimmerman (1998), Ludlow (1999) and Percival (2002). For criticisms see Oaklander (1984; 1994a; 1999; 2002), Le Poidevin (1991; 1999), Smith (1993; 1999; 2003), Tooley (1997) and Mellor (1998; 2001), among others.

2 For interpretation and defence of McTaggart's negative argument against A-time see Le Poidevin and Mellor (1987), Mellor (1998) and Oaklander (1996).

3 Craig also argues that presentness is not a property: 'Since presentness is identical with temporal existence (or occurrence) and existence is not a property, *neither is presentness a property*. Presentness is the act of temporal being.' (Craig 2000a: 202; emphasis added; compare with Craig 2000a: 191–201.) Note, however, that Craig is not consistent on this matter since he also claims 'one need not use tensed statements alone to talk about tense; for example, 'The A-determination *presentness* is an absolute property, not a mere relation' is tenselessly true (or false), but refers to an objective A-determination' (Craig 2000b: 156; compare with Craig 2000a: 222).

4 This common critique misses the mark because McTaggart does *not* start off by *assuming* that every event is past, present and future. On the contrary, McTaggart begins by *insisting* that an event or moment in time can have *one and only one* A-determination. Thus, he says:

> And we *must* say that a series is an *A* series when each of its terms has, to an entity *X* outside the series, *one*, *and only one*, of three indefinable relations, pastness, presentness, and futurity….
>
> (McTaggart 1968: 2, 20; emphasis added)

And again in 'The unreality of time',

> Past, present, and future are incompatible determinations. *Every event must be one or the other*, *but no event can be more than one*. … And, if it were not so, the A series would be insufficient to give us, in combination with the C series, the result of [B-] time.
>
> (McTaggart 1934: 123; emphasis added)

    The further claim that every event/thing/moment has all three A-determinations is not assumed but is implied by the view – endorsed by A-theorists – that change requires temporal becoming.

5 For the most carefully worked out A-theoretical account of the ontological significance of the tenses see Smith (1993; 1994a,b). For a critique of Smith see Oaklander (1996). Smith (2002) has recently modified his views.

6 For my criticism of this model of becoming see Oaklander (1984: Chs 2 and 3) and essays in Oaklander and Smith (1994: Part II). For criticisms of other non-presentist accounts of becoming see Oaklander (1994b,c; 2001) and Mellor (1981b; 1998).

7 Craig (2001) argues for absolute time.

References

Bigelow, J. (1991) 'Worlds enough for time', *Nous* 45: 1–20.

—— (1996) 'Presentism and properties', in James Tomberlin (ed.) *Philosophical Perspectives*, 10, *Metaphysics*, Cambridge, MA: Blackwell.

Chisholm, R. M. (1981) *The First Person*, Minneapolis: University of Minnesota Press.

Craig, W. L. (1991) *Divine Foreknowledge and Human Freedom*: *The Coherence of Theism*: *Omniscience*, New York: E. J. Brill.

—— (1998) 'McTaggart's paradox and the problem of temporary intrinsics', *Analysis* 58: 122–7.

—— (1999) 'Oaklander on McTaggart and intrinsic change', *Analysis* 59: 319–20.

—— (2000a) *The Tensed Theory of Time*: *A Critical Examination*, Dordrecht: Kluwer.

—— (2000b) *The Tenseless Theory of Time: A Critical Examination* Dordrecht: Kluwer.

—— (2001) *Time and the Metaphysics of Relativity*, Dordrecht: Kluwer.

Freddoso, A. J. (1983) 'Accidental necessity and logical determinism', *Journal of Philosophy* 80: 257–83.

Gale, R. M. (1968) *The Language of Time*, London: Routledge, Kegan & Paul.

Hinchliff, M. (1996) 'The puzzle of change', in J. Tomberlin (ed.) *Philosophical Perspectives*, 10, *Metaphysics,* Oxford: Basil Blackwell.

Le Poidevin, R. (1991) *Time*, *Cause and Contradiction*, Basingstoke: Macmillan.

—— (1999) 'Egocentric and objective time', *Proceedings of the Aristotelian Society*, 99: 19–36.

Le Poidevin, R. and Mellor, D. H. (1987) 'Time, change and the indexical fallacy', *Mind* 96: 534–38.

Ludlow, P. (1999) *Semantics*, *Time and Tense: An Essay in the Metaphysics of Natural Language*, Cambridge, MA: MIT Press.

McTaggart, J. E. M. (1934) [1908] 'The unreality of time', *Mind* 18: 457–74, repr. in S. V. Keeling (ed.) *Philosophical Studies*, London: Edward & Arnold & Co.

—— (1968) [1938] 'Time', in C. D. Broad (ed.) *The Nature of Existence*, Vol. 2, Cambridge, UK: Cambridge University Press, repr. Grosse Pointe, MI: Scholarly Press.

Mellor, D. H. (1981a) *Real Time*, Cambridge, UK: Cambridge University Press.

—— (1981b) 'McTaggart, fixity and coming true', in R. Healey (ed.) *Reduction*, *Time and Reality*, Cambridge, UK: Cambridge University Press.

—— (1998) *Real Time II*, London: Routledge.

—— (2001) '*Real Time II*: *Replies to Hinchliff, Paul and Perry*', in L. N. Oaklander (ed.) *The Importance of Time*, Dordrecht: Kluwer.

Oaklander, L. N. (1984) *Temporal Relations and Temporal Becoming*, Lanham, MD: University Press of America.

—— (1994a) 'Bigelow, possible worlds and the passage of time', *Analysis* 54: 244–8.

—— (1994b) 'McTaggart, Schlesinger, and the two-dimensional time hypothesis', in L. N. Oaklander and Q. Smith (eds) *The New Theory of Time*, New Haven, CT: Yale University Press.

—— (1994c) 'Zeilicovici on temporal becoming', in L. N. Oaklander and Q. Smith (eds) *The New Theory of Time*, New Haven, CT: Yale University Press.

—— (1996) 'McTaggart's paradox and Smith's tensed theory of time', *Synthese* 107: 205–21.

—— (1999) 'Craig on McTaggart's paradox and the problem of temporary intrinsics', *Analysis* 59(4): 314–18.

—— (2001) 'Tooley on time and tense', in L. N. Oaklander (ed.) *The Importance of Time*, Dordrecht: Kluwer.

—— (2002) 'Presentism, ontology and temporal experience', in C. Callender (ed.) *Time*, *Reality*, *and Experience*, Cambridge, UK: Cambridge University Press.

Oaklander, L. N. and Smith, Q. (eds) (1994) *The New Theory of Time*, New Haven, CT: Yale University Press.

Percival, P. (2002) 'Mellor on time', in C. Callender (ed.) *Time*, *Reality*, *and Experience*, Cambridge, UK: Cambridge University Press, pp. 91–118.

Prior, A. N. (1968) *Time and Tense*, New York: Oxford University Press.

Smith, Q. (1993) *Language and Time*, New York: Oxford University Press.

—— (1994a) 'The infinite regress of temporal attributions', in L. N. Oaklander and Q. Smith (eds) *The New Theory of Time*, New Haven, CT: Yale University Press.

—— (1994b) 'The logical structure of the debate about McTaggart's paradox', in L. N. Oaklander and Q. Smith (eds) *The New Theory of Time*, New Haven: Yale University Press.

—— (1999) 'The "sentence-type version" of the tenseless theory of time', *Synthese*, 119: 233–51.

Smith, Q. (2002) 'Time and degrees of existence', in C. Callender (ed.) *Time, Reality, and Experience*, Cambridge, UK: Cambridge University Press.

—— (2003) 'Reference to the past and future', in A. Jokic and Q. Smith (eds) *Time, Tense and Reference*, Cambridge, UK: MIT Press.

Tooley, M. (1997) *Time, Tense and Causation*, Oxford: Clarendon Press.

Zimmerman, D. (1998) 'Temporary intrinsics and presentism', in D. W. Zimmerman and P. van Inwagen (eds) *Metaphysics: Big Questions*, Cambridge, MA: Basil Blackwell.

14 *Real Metaphysics*
Replies

D. H. Mellor

Introduction

To be offered a festschrift is a great honour; to have such editors and contributors makes the honour greater still; and to be given the last word makes the offer irresistible. The only drawback is that, as saying how much I agree with everyone would take too long, and be less useful to readers, than saying where and why we disagree, my replies may seem ungraciously combative. Still, since we all know how debate can advance philosophy, I hope no one will infer any disrespect from my disagreements. On the contrary, it is to the work and friendship of these colleagues, mentors and students that I owe much of the understanding and pleasure that philosophy has brought me. For that I am very grateful to them all.

Truthmaking, truth and success

1   David Armstrong

In my theories of causation (1995) and of time (1998), I invoke the concept of truthmaking to resolve an ambiguity in 'giving a proposition's truth conditions'. This phrase may mean saying what makes a proposition true. But it may also just mean using a metalanguage to say when an object language sentence expressing the proposition *is* true; and that may tell us nothing about what *makes* it true.

Take time for example. Advocates of tenseless time habitually use a tenseless metalanguage to say when tensed sentences are true, while their opponents use a tensed one to say when tenseless sentences are true. Each side then attacks the other's metalanguage, one giving tenseless (e.g. indexical) accounts of 'past', 'present' and 'future', the other saying that 'earlier' really means 'less future or more past'. And on this semantic issue both might be right: each might be able to say in its own terms when any temporal sentence is true. But they cannot both be right about what *makes* such sentences true, i.e. about whether time itself is tensed. And both are certainly wrong if they think the semantics of time – or of any other contingent subject matter – fixes its ontology. That is why we must distinguish the two and why, because the

expression 'truth conditions' blurs the distinction, I avoid it in my (1998) in arguing for a tenseless ontology.

David Armstrong has of course never reduced metaphysics to semantics, as his theory of the mind (1993) shows. He may start with a behavioural account of mental *concepts*, but this is not what makes him identify mental *states* with physical states of the central nervous system. And even if some of us dispute that identity (Crane and Mellor 1990), few today still think that the meanings of mental terms suffice to tell us what makes psychological propositions true.

This is why I share David's belief in truthmaking, understood as his 'cross-categorial' link between a non-propositional entity S and a proposition 'P' that S makes true, a link whose paradigm is, as he says, that between S and 'S exists'. This does not of course reduce truthmaking to the entailment of 'P' by 'S exists', since that is not a link between S and 'P', but it does mean that no one who grasps the concept of existence can credibly claim not to know what truthmaking is or whether there is any.

David and I do however differ on details, and in particular on two of his initial claims: that 'every truth has a truthmaker' and that 'the determination of a truth by a truthmaker is a necessitation'. I disagree: I think that many truths do *not* have truthmakers, and also that some truthmakers do *not* necessitate what they make true. Let me take these points in turn.

First, because the identity of a necessary proposition entails its truth, I cannot see why any other entity must exist to make it true. So, in particular, since any contingent proposition 'P' is *necessarily* contingent, I, unlike David, see no need of a truthmaker for the necessary truth that P is contingent and hence that, in this sense, ~P is possible. However, some modal propositions do need truthmakers, because they are contingent: their identity does not entail their truth. These include truths about chances, such as the chance $ch(H)$ of a coin toss landing heads, which I take to measure how possible some fact, say about how the coin is tossed, leaves that outcome; and I discuss these further in Section 11.

Second, even some contingent truths need no truthmakers, notably true truth-functions, whose truth follows from the truth values of their constituents. We may *say* of course that 'P&Q' and 'P∨Q' are 'made true' by the truth of 'P' and 'Q'; but this is just the entailment of one proposition by others, not the 'cross-categorial' link between propositions and other entities that concerns us here. That is what true truth-functions do not need and therefore, I claim, do not have.

The fact is that only *atomic* propositions, and such non-truth-functional compounds of them as '*a* believes that P', 'If P were the case Q would be' and '$ch(H)=p$' need truthmakers. In particular, negative propositions do not need them, since if 'P' is made true by S, all it takes to make 'P' false and hence '~P' true is that S *not* exist. (I do not of course claim that we can always tell which if either of two *sentences* 'P' and '~P' expresses an atomic proposition: if either does, it will be the one that *does* have a truthmaker.) To postulate a

distinct 'falsemaker' for 'P', say ~S, to be a truthmaker for '~P' only raises the gratuitous question of why S and ~S, like an ontological Cox and Box, cannot coexist. It also, as David admits, makes it hard to explain how there could be *nothing*: for what entity could make it true that there *are* no entities? Once we see that negative truths need no truthmakers, that problem disappears.

And so does the otherwise intractable problem of saying what makes generalizations true. Imagine a world with just two particulars, *a* and *b*, to both of which a contingent predicate '*F*' applies. If, as David assumes, truthmakers must *necessitate* what they make true, it will take more than the truthmakers of '*Fa*' and '*Fb*' to make 'everything is *F*' true, since '*Fa&Fb*' does not entail this, because it does not entail that there are no other particulars. But as 'there is no particular that is neither *a* nor *b*' is a negative truth, it needs no truthmaker. All it needs is that no truthmaker for its negation exists, i.e. that no particular other than *a* or *b* exists. So if *a* and *b* are indeed the only particulars, whatever makes '*Fa*' and '*Fb*' true will also make true 'everything is *F*', even though it will not necessitate it.

Similarly for properties. Suppose there are N properties, $F_1 \ldots F_N$, for some finite or infinite N. David says in his Section 3 that we need 'a truthmaker … for the truth that [this] class of properties is the class of all the properties'. But not if all the negative truth 'there are no properties other than $F_1 \ldots F_N$' needs is the non-existence of a truthmaker for its negation, i.e. the non-existence of any property other than $F_1 \ldots F_N$. Here again we have a generalization made true by entities, $F_1 \ldots F_N$, which fail to necessitate it.

In short, David's necessitation principle fails for generalizations, which are not entailed by the conjunction of all their instances, since that conjunction does not entail that there *are* no other instances. But this should not make us reject his principle altogether, only when a truth requires certain entities not to exist.

And once we allow this harmless (because principled) exception to necessitation, we may as well allow another: that where it takes several entities to necessitate a proposition, we may as well call any of them, given the others, a truthmaker for it. Take truths about what is visible in a mirror. To necessitate these we need both the mirror and the objects reflected in it, not to mention the reflected light and the laws of reflection. Yet *given* the mirror, the light and the laws, we may as well say that propositions about what is visible in it are made true by the objects it reflects.

Similarly with truths about David's beliefs, for example, that he is an Australian. For even physicalists will admit that it takes more than David's brain states to necessitate propositions about what he believes. It also takes laws linking his brain states to how he behaves, and perhaps his living in Australia and not in some 'twin Australia' elsewhere in the universe. Yet given all that, it is an innocuous abbreviation of physicalism to say that propositions about David's beliefs are made true by states of his brain.

2   *David Lewis*

'Any proposition has a subject matter, on which its truth value supervenes', says David Lewis: a proposition 'P' can be true in one possible world and false in another only if those worlds differ in its subject matter. Thus if 'P' is 'there is (actual) styrofoam', then in any world with styrofoam 'P' is true, and in any world without it 'P' is false and '~P' true.

It follows that, as we have just seen, the negative existential proposition '~P' needs no truthmaker in any world, merely the absence from that world of the truthmaker for 'P', namely styrofoam. Yet in their postscript to his paper Gideon Rosen and David argue that propositions like '~P' do in fact have a truthmaker of the kind he offers, namely the world 'qua just as it is'.

I disagree, for the following reason. David's world is the mereological sum of all its parts, $S_1 \ldots S_N$, for some finite or infinite N. But $S_1 + \ldots + S_N$ will only be the sum of *all* the world's parts if the world has no *other* parts and, in particular, none that, by being styrofoam, would make 'P' true. Calling $S_1 + \ldots + S_N$ 'the world' only begs that question: it does not enable it to make '~P' true.

David's own paper offers truthmakers not for negative existentials but for predications of intrinsic properties. These truthmakers assume his theory of possible worlds, containing only counterparts of particulars in other such worlds (Lewis 1968), a theory which actualists such as me and David Armstrong reject. What can we offer instead?

Since, as David Lewis admits, truths can depend on 'whether something is, and … how something is', the truth of '*Fa*' may depend on *a*'s properties as well as its existence. Even so, as he shows, such propositions can still be made true by particulars, if properties are sets of particulars. But this, we may all agree, is credible only if merely possible as well as actual particulars exist. For truthmakers we may therefore have a choice of package deals: David's many worlds of particulars versus an actual world of what in my (1995: Ch. 13.4) I call 'facta' and David Armstrong (1997) calls 'states of affairs', entities which contain properties that are not just sets of particulars.

To our actualist package deals David objects that he does not understand the 'unmereological composition' of our facta – 'unmereological' because a particular *a* and a property *F* can exist without *a* being *F*, hence the notorious regress of instantiation relations: *I* linking *a* and *F*; *I'* linking *a*, *F* and *I*; and so on. But I do not face this regress, since I, like Wittgenstein (1922), take ours to be a world of facta, not of particulars. Only my facta are not simples, i.e. tropes: they are structured, because they instantiate laws. Thus, if it is a law that everything is *G*, I say that its instances are not the *G*-particulars *a*, *b*, … but the *G*-facta *Ga*, *Gb*, … , where *G* is what these facta share and *a*, *b*, … are their differentiae. In short, particulars for me exist only in facta, which they therefore need not combine with universals to constitute. All that follows from the possibility of *a* and *F* existing without *a* being *F* is that laws including *F* may have no instances that coincide or overlap in spacetime with *Ga*, which to contain *a* they would have to do.

In both my theories of change (Mellor 1981: Ch. 7; 1998: Ch. 8) I divide temporally extended particulars into *events* (e.g. speeches), which have temporal parts, and *things* (including people), which do not. This means that whereas things can change, events cannot, since temporal variation in an event (e.g. a speech getting louder) is just a difference between distinct entities, namely distinct temporal parts of it. But if a is a thing, F is a changeable property and t is a time, I say now (Mellor 1998: Ch. 8.6) that 'a is F at t' is made true not by a temporal part of a – a-at-t – being F but by an Fa-factum located at t. What I failed to see is that, as David rightly assumes, this factum must be *essentially* located at t, to enable it to necessitate the truth of 'a is F at t'. This however is no objection, since it does not imply that a itself must be F at t, merely that, if it is not, that individual Fa-factum will not exist.

David's other upgrade to my theory of change, giving it proxies for his temporal parts, I find less congenial. The idea is that a's history, a_H, is an event with a temporal part a_H-at-t that, by having a property F^* related to F, can provide a truthmaker for 'a is F at t'. But for me, as for Davidson (1970), events are particulars, whose parts, if any, are also particulars, whereas a_H's parts are not particulars but facta containing a. These of course I accept, but not their mereological sum, which is what a_H must be. For just as David does not understand unmereological composition, so I, for reasons I cannot go into here, reject the unrestricted mereological composition which he does understand and accept. That is to say, I deny that any two or more entities automatically compose another of which they are parts and, specifically, that facta containing a compose any such entity as a_H. In short, *pace* David, I do wholeheartedly reject the temporal parts he offers me, by denying the existence of the whole he thinks they are parts of.

3   *Peter Smith*

I have long endorsed what Peter Smith follows Jamie Whyte (1990) in calling 'success semantics'. This is the thesis that the truth of our beliefs is what makes the actions they combine with our desires to cause succeed in achieving the objects of those desires. Unfortunately, I also mistook this thesis to require truths to correspond to facta, or to facts in some other non-trivial sense of 'fact'. Peter and David Lewis (2001) have now persuaded me that this is wrong. Neither success semantics nor the fact that some truths need truthmakers, either is or needs any such correspondence theory of truth. All they need is the equivalence principle, that any proposition 'P' is true if and only if P, a principle which (with Peter's qualifications) I now think tells us all we need to know about truth.

I also agree with Peter that only contingent propositions need truthmakers, since their identity does not entail their truth, as we noted in Section 1 that the identity of necessary propositions does. Contingent truths that are not truth-functions of other propositions must therefore be *made* true by what our world contains. Peter and I agree moreover that most truthmaking entities

are facta, containing contingent particulars (as opposed, for example, to numbers) and contingent 'natural' properties (including relations) whose sharing entails objective resemblance and a similarity of causal powers (see, for example, Shoemaker 1980) – these from now on being what I shall mean by 'properties'. Properties so understood may be universals (as I think), sets of exactly resembling particulars or tropes, or something else again (see Mellor and Oliver 1997: *passim*); but that is another issue, which we need not settle here.

For what matters here is not what properties *are*, but what properties, and hence what facta, our world contains, and this I say depends on laws, as follows. First, I extend the idea of a law statement's so-called 'Ramsey sentence' by taking it to replace *all* that statement's predicates, not just its theoretical ones, with existentially bound variables. Then, calling the Ramsey sentence of the conjunction of all actual laws 'Σ', I say that our world's properties are those over which Σ's second-order quantifiers must range in order to make Σ true. This is what, in my (1995: Ch. 15.4–7), I call 'Ramsey's test' for what properties there are.

But how on this view, Peter asks in his Section 4, can my facta make propositions like his 'the ice-cream is in the freezer' true, given that neither being ice-cream nor being in a freezer is a property so understood? My answer is, as he says, that they can do so because for me the world contains more properties than those that figure in 'the ultimate laws of fundamental physics'. Thus, for reasons given in Section 9 of my (2000b), I say, for example, that temperatures are properties, distinct from the micro-properties, such as the mean kinetic energies of gas molecules, to which laws of nature link them (see Section 6 below). The freezer, a, and the ice-cream, b, can then be identified by their thermal, chemical and spatio-temporal properties and relations (and those of their parts), since these, together with the relevant laws, suffice to make it true that a is a freezer, that b is ice-cream and that a contains b.

4   Chris Daly

Concepts

I am grateful to Chris Daly for telling me how I do philosophy, and I own up to much of what he says. But not to all. For a start, I take our concepts to be less well defined than he does. Before relativity, for example, we all took simultaneity to be a two-term transitive relation. Nowadays, most of us think that it is either a three-term relation (the third term being a reference frame) or not transitive. I would put this by saying that we have discarded one of simultaneity's connotations, which Chris thinks means we have changed our concept, since he takes that to be a connotation of 'connotation'. I disagree, because all I mean by calling an inference we feel entitled to draw from applying a predicate 'F' a 'connotation' of 'F' is that it is one of several such inferences that matter to us and which we think preserve truth.

I therefore deny that, when we find that one such inference sometimes fails when the others always succeed – as when indeterministic causes fail to 'necessitate' their effects – we must always infer not that we have discovered something about causation, or simultaneity, but that our concept of it has changed. So I do not think, as Chris implies, that all 'folk' utterances about what is happening now must be false, just because most folk do not know that what is happening 'now' at a spatial distance is relative to a frame of reference.

Of course a concept *may* change, if we find that too much of what we have habitually inferred from applying a term fails to be true; and it may be hard to say how much is too much. But it may still be clear enough in a given case that *not* too much fails; and in showing this it helps to be able to explain the appeal of the connotation we are discarding. With causation I do this, as Chris notes, by exploiting the fact that most of its connotations come by degrees, measured I say by how much a cause C *raises* the chance of an effect E (not just, as Chris assumes, by E's chance with C). This, by providing a measure of what I call C's *efficacy*, explains why deterministic causation, where C raises E's chance from 0 to 1, is the ideal: because it meets causation's other connotations as fully as they can be met. This in turn shows why, before the rise of indeterministic theories, determinism was itself an important connotation of causation. It also shows how and why causation can fall short of that deterministic ideal and still be causation, and reduces the question 'How far short?' to one of how to map a qualitative concept (causation) onto a quantitative one (efficacy) – a question with as uninterestingly context-dependent answers as 'How hot is hot?'.

Truth and belief

Chris's four objections to the success semantics I endorse in Section 3 may be met as follows. First, since, as Chris says, causes always precede their effects, beliefs that combine with desires to cause actions will always include beliefs about the future: in his example, Toad's belief that there will still be honey in his pot when he opens it.

Second, success semantics does *not* say that all combinations of beliefs and desires cause actions whose success the beliefs' truth will ensure, precisely because the causation required is indeed contingent, e.g. on the agent's not being paralysed. All it says is that any actions these combinations *do* cause will succeed if all the beliefs involved are true. However, I do now think that, as human beliefs and desires generally cause actions indirectly, by causing intentions to act, we should call their effects decisions, i.e. the forming of intentions. This moves much of the contingency that bothers Chris into the link between intention and action. (But not of course all, since the causation of decisions by beliefs and desires is still contingent on, for example, the absence of stronger conflicting desires.)

Third, the fact that some causation is indeterministic does not rule out deterministic theories like (a) Newtonian mechanics and (b) decision theories

which say that decisions are caused by combinations of belief and desire. All the contingency of these theories shows is that (a) accelerations *might* not be caused by forces acting on masses and (b) decisions *might* not be caused by beliefs and desires. But just as accelerations that *are* caused by forces in the way Newton says *will* therefore be proportional to the net forces that cause them, so decisions caused by full beliefs and desires in the way decision theories say will, if carried out, succeed if all those beliefs are true.

Chris himself gives the answer to his fourth objection, that true beliefs can cause actions which fail, as when the pot Toad opens to get the honey he truly believes it contains is booby-trapped. The answer, taken from Jamie Whyte (1997), is that this is not the only belief Toad needs to make him decide to open the pot. He must also believe that if he opens the pot he will get what it contains, and this is the belief whose falsity makes his action fail. Chris says that invoking beliefs like this makes success semantics trivial; but it is not, any more than it is trivial that objects have masses that satisfy Newton's laws of motion. Nor therefore is it trivial that we have states of mind (beliefs) with contents that both make them combine with other such states (desires) to make us decide to act in specific ways and that can also, by being true, ensure that the objects of those desires are achieved by acting in those ways.

Communication

Chris discusses two claims that he says I make in my (1990), namely that to tell you that P I must (1) make you believe that I *believe* I believe P and (2) consciously believe P; he also conjectures that I believe (2) because I believe (1). I am afraid he is wrong on all counts. What I do say is that to tell you that P I must (1*) make you believe I believe P and – ignoring degrees of belief – (2*) have a conscious belief either in P or in not-P, depending on whether I want to tell you the truth or a lie. Nor is (1*) my reason for believing (2*): (2*) seems to me an observable fact, explained by my (1980) thesis that to believe any P consciously is to believe one believes it.

I do however agree with Chris that we can make statements, just as we can act in other ways, without the beliefs that cause us to do so being conscious, and I give an example of this in my (1990). What I deny is that these are cases of *communication*, i.e. of *telling* someone that P, as opposed to *showing* that P, or unconsciously revealing one's belief that P, revelations which may of course cause others to believe P too. But this need not be why, in Chris's example, a husband's absent-minded reply 'P' to his wife fails to tell her that P in my sense of 'tell'. For what the husband lacks in this case need not be a conscious belief that P, but any intention or expectation of making his wife believe P itself, which is after all the main point of telling someone that P. What makes his reply to his wife absent-minded, I suggest, is that he is not even *trying* to tell her that P.

The property of truth

Finally, Chris says that my Ramsey Test for what properties there are makes truth itself a property as well as a concept. Not so, and Chris is right to conjecture that my 'talk about the property *truth* is intended as a *façon de parler*'. Maybe the Ramsey Test would make truth a property if success semantics told us what truth *is*. But as I say in Section 3, I now see that all success semantics tells us about is *belief*, with *truth* being sufficiently defined by the principle that, for all 'P', 'P' is true if and only if P. This being so, the relevant parts of statements of the laws that success semantics requires need only read 'if *x* believes P, and P, then …', in which the predicate 'is true' does not occur. So since it follows that, whatever laws there are, the Ramsey sentence Σ of their conjunction *need* not quantify over truth in order to be true, my Ramsey Test does not make truth a property.

Mind and causation

5   Tim Crane

When Frank Jackson's (1986) Mary leaves the black-and-white room she was brought up in, she sees a red tomato and thereby learns what red looks like. Tim Crane endorses Frank's view that this is learning a fact, namely that (as Tim puts it) 'red looks like this', which only those who have seen something red can know. In my (1993b) I followed David Lewis (1988) and Laurence Nemirow (1990) in denying this, arguing that all Mary acquires is an ability to imagine and recognize red things; arguments which, despite Tim's rebuttals, I still accept.

I, however, unlike many 'ability theorists', am not a physicalist. So, as Tim admits, physicalism is not my motive for denying the undeniably non-physical fact that he and Frank think Mary comes to know. I believe in many non-physical facts, not only in the weak sense of 'truths' but also as truthmaking facta, such as those containing the visual sensations Mary's tomato – call it Rudy – gives her. I also believe that seeing Rudy may well teach Mary four relevant truths, namely that both Rudy and its colour both are and look red. What I deny is that, besides these four unproblematic truths, there are any truths about *what red looks like*, i.e. about *what it is like to see red*.

I deny this largely because I still cannot see why, if there are such truths, we cannot state them. For, as I say in my (1993b: 8), although we

> have words for properties of experiences, like a 'loud' noise, a 'sweet' taste, a 'warm' feeling … we can say nothing about what they are like. What does [a loud noise] sound like, sugar taste like, warmth feel like? We cannot say. All we can say is that these experiences are more or less like, i.e. similar to, certain other experiences. But that does not tell us what, in the relevant non-relational sense of 'like', any one of a set of similar experiences is *like*.

Tim disagrees: he says that Mary's 'red looks like this' says what red looks like, at least to her. I deny this: I think that what Mary says means either 'this looks red', where 'this' fixes the reference of 'red' by referring to Rudy (or to its colour), or 'red looks like this looks', which uses the irrelevant relational sense of 'looks like' to say that red looks like Rudy or its colour, whatever that looks like.

Tim does say that no book could 'express the proposition … Mary expresses when she says "red looks like this" ', but that seems to me both false and irrelevant. For Mary might as well have said 'red looks like (the colour of) Rudy', a proposition any book could state, illustrated perhaps by a colour picture of Rudy, to whose colour it too could then use 'this' to refer. But that is irrelevant if, as I claim, 'red looks like this' does not state what Mary learns when she learns what red looks like.

This however only reinforces Tim's main point, that to learn what red looks like Mary must see something red: black words on white paper will not do. But that too is irrelevant if what she learns is not a truth but an ability. And, anyway, it is not a necessary truth that Mary cannot learn what red looks like without seeing something red: not everyone need be as unimaginative as she was in her black-and-white room. Sculptors, for example, who can plan work in their heads, can tell what a sculpture will look like before anyone sees it, while the score of (say) Berlioz's *Symphonie Fantastique* can make some musicians hear in their heads orchestral textures they have never heard in reality, thereby coming to know what these sound like before they hear them. This is not of course to assert what Tim would deny, namely that these musicians have knowledge of a proposition expressed by the score. They do not; but whereas Tim thinks this is because they know a fact that cannot be so expressed, I think it is because there is no such fact: the score teaches them not a truth but how to imagine and recognize Berlioz's orchestral sounds.

Mary's knowledge differs therefore from Tim's other example of knowledge that books cannot express, namely the knowledge his Vladimir expresses by pointing to (say) Thetford Forest on a map and saying 'I am here!'. For what makes Vladimir's knowledge inexpressible by any movable map with an 'I am here!' sign fixed to it is not his (in effect) affixing that sign by pointing, but the indexicality of 'here', which requires his map to be where he is for its 'I am here!' sign to express what he knows. But there is nothing indexical about what Frank (1986) says Mary learns: namely, *what red looks like*, i.e. looks like *to everyone* (with normal eyesight), not just to her. And if the truth that Tim and Frank say Mary knows is not indexical, why can it not be expressed – unless, as I claim, there is no such proposition?

My know-how view does, as Tim says, make Mary's knowledge 'irreducible to propositional knowledge', but not therefore as 'completely different' from it as he thinks. For the success semantics I espouse makes knowing that P, for many contingent P, special cases of knowing-how, namely of knowing how to act to get what we want. Take the example in Section 4, of Toad opening a pot to get the honey it contains. His action succeeds because the belief

(P_H) that combines with his desire for honey to cause this action, namely that the pot contains honey which he will get if he opens it, is true. So if, as I assume, for any P, knowing P entails believing P, and P, then by knowing P_H Toad knows how to get the honey he wants. (Even if he cannot in fact get the honey, because he cannot open the pot, he still knows *how* to get it.) Similarly in all other cases. Mary's knowing how to imagine and recognize red is not so different from Toad's knowing how to get his honey by knowing that his pot contains it. All Mary's case shows is that propositional knowledge is not the only form of know-how.

But if Mary's knowledge is not propositional, it is not knowledge of a fact even in the trivial sense given by the principle that a proposition 'P' is true if and only if it is a fact that P. So in particular it is not knowledge of a subjective fact. But because indexical knowledge, like Vladimir's knowledge of where he is, *is* propositional, it is of a fact in this trivial sense. But it does not follow that this fact is what makes Vladimir's belief in it true, and I say it is not: Vladimir's indexical belief is made true by the non-indexical (and so for Tim objective) fact that Vladimir is in Thetford Forest. It is in this truthmaking sense of 'fact', for which I coined the term 'factum', that I say there are no indexical facts.

Here, however, my denial that truth-functional truths have corresponding facta poses a problem, as Tim notes. For Vladimir, by believing he is in Thetford Forest, also believes among other things that he is not in Russia, i.e. that (for him) Russia is not here. But this belief of his is not made true by the negative factum that he is not in Russia, since there are no such facta, any more than there are indexical facta. In what sense then can Russia's not containing Vladimir be a fact when its not being here for him is not? The answer is that our world's facta, by fixing which atomic (and other non-truth-functional) propositions are true, thereby fix which truth-functions of those propositions are true. We may therefore extend the concept of facts as facta in an innocuous but non-trivializing way by saying that these truths too state facts. It is in this sense that I say it is a fact that Vladimir is not in Russia. But if, as I claim, there are no indexical *facta*, it is not a fact even in this extended sense that for Vladimir Russia is not here. This is the sense in which I deny Tim's subjective facts.

6   *Frank Jackson*

Frank Jackson argues from physicalism to 'the a priori passage principle' that 'for each true statement concerning our world, there is a statement in physical terms that a priori entails [it]'. The validity of his argument I accept, but not its physicalist premise, for reasons Tim Crane and I gave in Mellor and Crane (1990) and I think Frank has not refuted. Specifically, I still think that physicalism faces a fatal dilemma: either all sciences (including psychology) count as physical and it is trivially true, or it is false that, as Frank (1998: Ch. 1) puts

it, 'the kinds of properties and relations needed to account for the exemplars of the non-sentient are enough to account for everything … contingent'.

Why does Frank think they are enough? After all, his own examples, the microphysics of water and of heat, do not account for anything sentient. Still, they do use microscopic facts to account for macroscopic ones and, as he says (Frank 1998: 7), 'the mind is a macroscopic phenomenon'. That, however, is, as he might admit, a pretty weak induction, even if his examples work; and the fact is that they do not work. For, despite what he and many others, misled by Kripke (1972) and Putnam (1975), have said, water is *not* H_2O, and heat is *not* molecular kinetic energy: in neither case does microphysics account in Frank's sense for the macroscopic phenomena.

In Section 9 of my (2000b) I gave several reasons for denying that heat is molecular kinetic energy, one of which may be summarized as follows. First, temperatures pass my Ramsey Test for being real properties, being quantified over in many laws: the laws of thermodynamics itself, the laws linking them to the masses, pressures and volumes of samples of given gases; to the mean kinetic energies of gas particles; to the rates of chemical reactions; to the frequency distributions of emitted radiation; and so on. Second, suppose that we take the laws of thermodynamics, and those linking temperatures to such other properties of macroscopic objects as their pressures and volumes, to specify what Frank and others call the 'heat role'. Then, *pace* Frank, this role has at least two 'occupants': not only the mean kinetic energy E of gas particles, but also the energy flux X of black-body radiation. But neither of these can *be* the temperature T to which different laws of nature link them: for, as I show in my (2000b), the way in which gas and radiation initially at different temperatures in the same vessel must interact to reach thermal equilibrium requires X, E and T to be distinct properties.

But what if the laws linking X, E and T are necessary, as I shall reluctantly admit in Section 10 that they might be? Certainly, if X and T are correlated necessarily, any energy flux X of black-body radiation will entail that its temperature is the corresponding T. But also vice versa: the supervenience is symmetrical, as it would be between states of mind and brain correlated by deterministic and necessary laws. There is no sign here of the asymmetrical supervenience that physicalism needs. And there is certainly no sign of it with the law linking E and T, which advocates of $T=E$ must pretend is deterministic even though they know very well it is not. For since the real law only links any T to a chance of the corresponding E, which, although high, is always less than 1, it will, even if it is necessary, positively prevent T supervening on E.

In short, the non-thermal 'kinds of properties and relations needed to account for' gas particles are *not* enough to account for the thermal behaviour of gases, which they do not even entail, never mind a priori. Similarly, although for different reasons, with water and H_2O. First, suppose we again take the laws that link water's macroscopic properties – its solvent powers, density, freezing and boiling points, latent heats, and so on – to define the

'water role', then to be water cannot possibly be to be H_2O, since, even if we count ice and steam as water, these allegedly identical properties have quite different extensions. In particular, no single H_2O molecule can be water, since it instantiates hardly any of water's laws, having no solvent powers, density, freezing or boiling points, or latent heats. Water's relation to H_2O is at best that of a heap of sand to its grains; to say therefore that it *is* H_2O is as absurd as saying that people are not human bodies but human cells.

Moreover, unlike a temperature, water is not for me a property at all, since the Ramsey sentence of all laws need not quantify over it. What 'water' names is not a single property but a natural kind, a congeries of macroscopic properties, such as those listed above. And the microphysics of the H_2O (and other) molecules that water contains is not, as Frank supposes, enough to account for this congeries: if only because, as we have just seen, it cannot account for the temperature of water (nor, for example, for its pressure), on which most of its other properties depend. But if even a mature microphysics cannot account in Frank's sense for the most important macroscopic properties of water, I see no reason to share his faith that the sciences of the non-sentient will one day account in his sense for all mental phenomena. On the contrary, to me it seems obvious that peculiarly psychological 'kinds of properties and relations' will always be needed to do that, just as peculiarly thermal and other macroscopic kinds of properties are needed to account for the phenomena of heat and of water.

7   *Paul Noordhof*

Epiphenomenalists owe us a theory of causation to explain why non-physical mental entities can have causes but not effects, a debt that I agree with Paul Noordhof they cannot discharge. All serious theories of causation link causes to effects (or their chances) in one or both of Hume's (1748: §VII) two ways: by counterfactuals or as instances of generalizations. And nothing about either way stops mental entities figuring as easily in their antecedents as in their consequents. If you would (probably) not have thought it was cold out had you not seen the snow, why might you not have (probably) gone out had you not thought it was cold? If everyone in brain state B (and …) feels embarrassed, why may not everyone who feels embarrassed (and …) blush?

Paul discusses the stock answer to such questions, the 'causal closure' principle that all effects have *only* physical causes, and accepts my and Tim Crane's (1990) objection that their all *having* physical causes does not entail this. However, Paul thinks our argument requires non-physical causes to overdetermine their effects, and notes that an unwillingness to admit 'systematic overdetermination … is the major reason why most philosophers of mind have become physicalists'. But, as I note in Section 8, an effect's physical and non-physical causes will *not* overdetermine it when they are linked, as they usually are, by laws that make both present or absent together. What Paul calls the 'a priori implausibility of systematic overdetermination' is as bad an argument for epiphenomenalism as it is for physicalism.

There being no other argument for epiphenomenalism that I know of, the inefficacy of the mental can be only an axiom. But, as Paul says in his Section 1, it is an axiom that rules out both obvious examples of mental causation and good causal theories of how we know and refer to our own states of mind. I find these objections to it stronger than Paul does, since I deny that they need more defence just because epiphenomenalists can explain them away. Compare Kripke's (1971) proof that laws of nature which we cannot know a priori could still be metaphysically necessary: this is no reason to think that such laws *are* necessary, given other arguments for their contingency; and similarly here. It is epiphenomenalists, not their opponents, who should be on the defensive, since it is they who need independent arguments for the inefficacy of the mental, to set against all the apparent examples of mental causation, and the independent arguments for causal theories of knowledge and reference.

This is of course no objection to Paul's new argument against epiphenomenalism: it does no harm to make a strong case stronger. I do however jib at the changes he thinks he needs to make to my theory of causation, for the following reasons.

First, I do indeed think it is metaphysically necessary that, as Paul puts it, 't precedes t' if there is some fact C at t which causes some fact E at t''. The necessity of this is clearly consistent with there being possible worlds where spacetime is not dense, or where special relativity is false, or where – as in our world – all parts of a solid object can move together at the same uniform velocity; and I cannot see why Paul says in his Section 3 that it is not.

Second, my argument against simultaneous causation does not stop two facts coinciding when – as in Paul's example of Jim's being both the fittest and the shortest man – neither causes the other. Moreover, I show in my (1995: Ch. 17.2) how to accommodate facts which coincide and *do* seem to interact, as when a gas sample's pressure at t is apparently caused by its volume at t and vice versa.

Third, Paul misreads my argument against the possibility of simultaneous causation at a distance. If backward causation is impossible, then simultaneous causation between non-coincident facts must also be impossible if there are *any* possible worlds where it would yield backward causation. But there are, since it does so in all worlds, like ours, where special relativity is true. So what my argument needs is not, as Paul thinks, that special relativity be necessary, merely that it be possible.

Fourth, Paul disputes my *reductio* proof that no two facts C and E can cause each other, and hence that causal loops, and thus backward causation, are impossible. This proof assumes that, *if* C and E can interact, any individually possible values of E's chances with and without C, and of C's chances with and without E, can coexist. Yet elsewhere, as Paul notes, I rule out combinations of laws that would impose incompatible time orders on spacetime points which instantiate them. Why then, he asks, instead of ruling out causal loops, should we not rule out the combinations of E's and C's chances that generate contradictions?

To this good question I have a four-part answer. First, in the time-order case we have no choice: there is no other way of ruling out incompatible time orders. But the contradictions that backward causation seems to make possible may be ruled out in two ways. Either backward causation is impossible, or all and only the members of an infinite, unspecifiably complex and totally ad hoc set of combinations of individually possible chances are impossible. The former is a vastly simpler theory, which is my second reason for believing it.

My third reason is that as the chances that C and ∼C give E are located in different possible worlds, I do not see how they can constrain each other; and similarly for the chances that E and ∼E give C. And my fourth is that, on my theory of chance, the facta that are the chances which C gives E and E gives C not only follow from different laws, but also have different locations: this chance of E being where C is, and this chance of C being where E is. This means that Paul's theory must postulate necessary links both between otherwise independent laws and between regions of spacetime that may be widely separated. Such links contradict attractive Humean assumptions about the independence of laws and of spacetime regions; and while my theory also violates the latter, since I say that any $ch(E)=1$ entails E, mine is a single and independently argued exception (2000a), not a farrago of ad hoc expedients.

That is the case for my theory that backward causation is impossible. It is not a logical knockout – the rival theory contains no contradiction – but then it does not need to be. In philosophy, as in boxing and science, one can still win decisively on points. It is a mere dogma of analysis, which I reject, that metaphysical theories, like theories in logic and mathematics, can be established only by showing that all their rivals entail contradictions.

Finally, to Paul's own case against epiphenomenalism I have only one objection: I do not see why we cannot credit mental facts that lack effects with temporal locations. For, since these facts do have causes, the principle of no unmediated action at a distance, which we can all accept, will locate them at the earliest time that is later than all their causes. They must still, as Paul says, coincide with facts that have effects, in order to make them earlier than those effects. But why should epiphenomenalists not buy this consequence of our view of spacetime? For even on Paul's theory that view need not, as he thinks, make any law 'take the form it [does] because of the presence of mental facts', since on that view it matters neither which facts with effects coincide with which facts without them, nor what those effects are, just so long as there are some.

8   *Peter Menzies*

My views on causation owe as much to Peter Menzies as his do to me. And we agree on more than he supposes, if not on whether causation is a relation. But before showing that, I must tackle his objections to my claim that causes must raise the chances of their effects.

Peter (Menzies 1989) showed how denying unmediated action at a distance

deals with all cases of what, following David Lewis (1986b: §E), he calls 'early' pre-emption, where the 'process that … brings about the effect cuts short all alternative processes before the effect'. But this, as he says, does not dispose of 'late' pre-emption cases, like his Case l, where 'the effect itself [the falling of the victim, Tony (say)] cuts short all alternative processes'. But the problem here is not, as he implies, that no *immediate* cause can raise the effect's chance: for effects need never have immediate causes if spacetime is dense and (as I argue in my 1998: Ch. 10) causes must precede their effects, since causation must then be as dense as time is.

Peter is however right to say that coping with his Case 1 means making the effect – Tony's falling – a different entity if caused by assassin A than if caused by assassin B. Now I have an initially tempting way of doing this, based on an identity criterion for facts with causes and effects (about which I have been less reticent than Peter claims): namely that, for any such facts D and D′, D=D′ iff D and D′ have all the same actual causes and effects (Mellor 1995: Ch. 9.3). This, however, being only a criterion of actual, not of counterfactual identity, does not show that a given fact *could* not have had different causes or effects, nor therefore that B's firing would not have caused (a counterpart of) the very effect that is actually caused by A's firing.

How then in Case 1, if causes must raise their effects' chances, can Tony fall because A fires? Well, as no one thinks A's firing causes Tony to fall on any other occasion, the effect here must really be (as Peter notes that I say in another case) that Tony falls roughly when he does, say at time *t*. And A's firing *must* raise the chance of that, simply because, to make the pre-emption late, B must fire only if and therefore *after* he sees Tony fail to fall at *t*. That is how, on a chance-raising theory, 'Tony falls (when he does) because A fires' can be true. And if it is, then it follows on my theory that *A's shot causes* (or at least *affects*) *Tony's fall*, by causing it to occur earlier than it otherwise would (Mellor 1995: Chs 11.3–12.2) – where *A's shot* and *Tony's fall* are particular events with the same identity criterion as my facts, namely that for 'any events d and d′, d=d′ iff d and d′ have all the same causes and effects' (Davidson 1969; Mellor 1995: Ch. 9.3). And similarly for all other cases of late pre-emption, with which, as Peter conjectures, I therefore believe I can deal, as I do in another case in my (2001) reply to Laurie Paul (2001).

I can also cope with Peter's Case 2, where A and B fire together, provided that each fires only if the other does: since that makes each assassin raise Tony's chance of falling from its value if neither fires to its value if both do. This is how, I argued in my (1995: Ch. 8.7), mental and physical facts linked by psychophysical laws can determine the same effects without overdetermining them (thus refuting another bad argument for physicalism: see Section 7 above). But this will not work if it is only a coincidence that A and B fire together, and such coincidences, although rare, do indeed pose a problem for most counterfactual theories of causation. But not for those who, like me, take causes and effects to be facts in the extended sense of Section 5, because those facts, unlike particular events, can be disjunctive. For even if neither A's nor

B's firing, given the other, raises the chance that Tony falls, their disjunction does: his chance of falling would have been less had neither fired.

If disjunctive causes sound odd to those who think of causes as particulars, so much the worse for that view; anyway, hard cases make bad law. It is no mere intuition that makes me require causes to raise the chances of their effects, but the fact that, as I show in my (1995: Ch. 8), key connotations of causation – that causes explain, are evidence for and means to their effects – require them to. Still, given that requirement on the basic concept, we can soften some hard cases by extending it (as I extended 'facta' to 'fact' in Section 5). Hence, for example, David Lewis's (1986b: §B) extension of what I call 'causation' and he calls 'causal dependence' to its ancestral, in order to make it transitive. If, as I think, his extension loses too many connotations, at least the extension is clear and easily reversed. Similarly with Peter's extension of causation to processes, like those linking A's and B's firing to Tony's falling, each of which would, without the other, have raised the chance of that effect. Calling disjuncts of disjunctive causes 'causes' too is no big deal.

However, if hard cases make bad law, unclear ones, such as Peter's cricket ball examples, are worse. Still, they do support my chance-raising require-ment, as is obvious from the way our response to them depends on whether we take their 'backup' walls or hands to be included in the causal set-ups. This certainly supports Peter's view that causation is embodied in intrinsic properties of law-based systems, and that what we think causes what depends on what we hold fixed in assessing the relevant counterfactuals. Both of these claims of Peter's I accept.

All I deny is that Peter's process view of causation requires it to be a *relation*. Of course, the evolution and consequent effects of law-based systems depend on their intrinsic properties and relations, but those relations need not include causation itself. Thus, of course, when hooking a cricket ball causes it to go for six in one piece, that system's – the ball's – holding together when hit depends on the law-governed properties and relations of it and its parts. Nevertheless, this causation only requires such facta to make true certain conditionals, about the ball's chances of having various trajectories if it is or is not hit in various ways. That, for the reasons given in my (1995: Ch. 13), does not make causa-tion a relation, and Peter fails to show that it does. The instances of possibly negative properties F and G in Statement (9) (p. 130), for example, need only be facts in the ontologically vacuous sense of Section 5 above. They need not be the real facta – like a cricket ball's mass – whose existence is what makes G-instances depend causally on F-instances.

Peter therefore could and should accept the arguments he cites against a relation of causation. In particular, he need not reject one of mine because it rests on 'the dubious principle that if some fact P is entailed by, but does not entail, some other fact Q, then P cannot be a genuine factum'. That I now agree must be wrong, since any atomic factum P is entailed by but does not entail its conjunction with any independent fact, and is none the worse for that. But, in my example, Sue's pulling her drive, P, is *not* a conjunction of her

driving, D, with an independent fact, but a disjunct of the disjunction D of various ways of driving a golf ball. So the reason that only P could be a factum is not the 'dubious principle' above, given I confess in my (1995: 165), but the fact that there are no disjunctive facta. If Peter will accept that – together with my thanks for correcting me so politely! – I hope he will also then accept that causation is not, after all, a relation.

Dispositions and laws

9   Isaac Levi

I am less immune than Isaac Levi thinks to his views on dispositions and conditionals. Although we do disagree on many points, on much that matters to us we agree in substance, if not in our interests nor hence in our idioms. Take the conditional which I said above that Chris Daly's Toad needed to believe: 'If I open the pot I'll get what is in it'. I say, and Isaac denies, that this conditional has a truth value. But I do agree that it differs from any unconditional proposition, since I say that Toad's belief in it is his disposition to believe its consequent if he believes its antecedent (Mellor 1993a). And what matters to both of us is that this disposition be truth preserving, not whether its content has a truth value.

Still, I do think Toad's conditional has a truth value, partly because it can occur within other conditionals, like 'If I'll get what is in the pot if I open it, I'll buy it' and 'If it is safe, I'll get what is in it if I open it'. I also think, despite Isaac's objections, that Toad could easily 'suspend judgement concerning the truth or falsity' of such conditionals (and of modal variants like 'If I open the pot I *may* get what is in it'), 'judge them probable to varying degrees … desire that they be true in varying degrees and the like'.

This is why I can say that conditionals are *entailed* by beliefs about dispositions, like the one I believe Toad has (to believe that he'll get what is in his pot if he believes he'll open it), whereas Isaac can only say that this belief of mine *supports* that conditional (that if Toad believes he'll open his pot he'll believe he will get what is in it).

But these are just different idioms: what matters is *which* conditionals are supported or entailed by such beliefs, and here too we can agree in substance. Take the two hard cases Isaac offers me in his Section 5:

(1)   an object *a* can have a surefire disposition (*D*) to *R* if *S*'d and yet fail to *R* if *S*'d, because being *S*'d causes *a* not to have *D*; and

(2)   a coin *c* can land heads if tossed (*T*), but not if tossed by Sydney Morgenbesser, by whom it is in fact tossed.

How can I cope with these? (1) I say that *a*'s disposition *D* is 'surefire' if and only if *a*'s chance $ch(Ra)$ of being *R* is 1 when *a* is *S*'d *if it remains D*. So what '*Da*' entails is not 'If *a* were *S*, $ch(Ra)$ would be 1' but 'If *a* were *S* and *D*, $ch(Ra)$ would be 1'. So if *a* is both *S* and *D*, *Ra* may have more than one actual chance:

the $ch(Ra)=1$ that is a fact about $Sa\&Sd$, and a smaller chance that is a fact about Sa. But that is no problem for me, since the theory of chances in my (1995: Ch. 2.1) lets a proposition have many actual chances of being true, each one a fact about a different earlier fact.

This means that, in Isaac's case (2), it can be a fact about how a coin c is tossed (by someone other than Sydney) that, as it is tossed, its chance $ch(H)$ of landing heads is 0.5; and it can also be a fact about c's facing heads up as it lands that $ch(H)$ is then 0.99. Naturally not *every* prior fact about c gives $ch(H)$ a value, any more than Sa has to give $ch(Ra)$ one. In particular, while the way Sydney tosses the coin c does give $ch(H)$ a value, the mere fact that c is tossed (Tc) does not. But then Isaac's

> (a) 'On the supposition that c is tossed, it might land heads.'

is ambiguous in the way I discuss below in Section 11. For if his 'might' means merely that Tc does *not* make $ch(H) = 0$ – because it gives $ch(H)$ *no* value, high or low – then (a) is true; whereas if (a) means that, as tossed, $ch(H)$ has a value, greater than 0, then (a) may or may not be true, depending on how c is tossed.

This is not of course what Isaac says about these cases, but it is consistent with what he says. It is also immune to his objections to David Lewis's (1973) 'closest worlds' theory of the relevant conditionals, which in my (1995) I too reject as an account both of their semantics (Ch. 1.7) and of what makes them true (Ch. 14.1).

As with chances, so with dispositions. Isaac and I agree that propositions like 'Da', which ascribe dispositions to things, have truth values. But when I add that not all dispositional predicates correspond to properties, Isaac claims to have 'neither understanding of nor interest in an 'ontological' distinction between predicates characterizing properties and predicates that do not' (p. 142). Well, that is his prerogative, as it is mine not to understand or be interested in American football. But that does not mean there is no such game, or no answer to the question of what properties there are. And in fact a distinction which Isaac does draw, between 'problem-raising' and 'problem-solving' predicates, fits my answer to that question quite well. For what makes predicates 'problem-solving' for him is their 'integration into adequate theories', whereas my Ramsey Test makes properties correspond to simple predicates in statements of laws of nature. So when our theories really *are* adequate, i.e. are true, his problem-solving predicates will be those that I say correspond to properties.

10   Alexander Bird

As Alexander Bird knows, I think 'dispositional' applies primarily to predicates, namely those, like 'is fragile', whose extension is given by a conditional, something like 'would break if dropped'. And, as I have observed in Section 3 and Section 9, most predicates do not correspond to properties in my sense.

In particular, the extension of 'is fragile', like that of 'is red', will certainly differ from that of any property ranged over by the Ramsey sentence Σ of all laws (see Mellor1997).

On the other hand, every property F does correspond to an actual or possible predicate 'is F'. So we can transfer the epithet 'dispositional' from predicates to properties by applying it to F if and only if all F-things satisfy one or more conditionals, i.e. (as Alexander puts it) have 'certain conditional powers'. In this sense I, like Popper (1990), think that all properties are dispositional, since my Ramsey Test makes them all occur in laws, which say that all Fs are Gs (or vice versa), so that anything would be G if it were F (or F if it were G) or – if the law is indeterministic – would have a certain chance of being G if it were F (or vice versa).

Does this make my properties 'categorical' in Alexander's sense, i.e. such that they 'confer, of themselves alone, no … causal powers … but [do so] only because there is a law relating [them] to some other property'? I cannot tell, because for me this is a false contrast, since I say that for any property F to exist *is* for laws to relate it to other such properties. However, I do take properties to be categorical in two more usual senses. First, I have just agreed with Isaac Levi that, for any dispositional predicate 'F' (whether F is a property or not), 'a is F (at t)' is a categorical statement, i.e. has a truth value, even if the conditionals that give 'F' its extension do not. All ascriptions of dispositions are categorical in this semantic sense, just as all actual properties are categorical in the ontological sense – i.e. real – whether they are dispositions or not.

In short, I think the war between Alexander's 'categorical' and 'dispositional' 'monists' is a phoney war, since all properties, including triangularity, are both. I largely endorse Alexander's defence of the view that triangularity is as dispositional as it is real; but I do have three comments to make about what he says. First, even if it is trivially analytic that a figure's triangularity is what makes counting its corners correctly give the answer '3', its having this property can still be what makes my counting its corners *cause* me to get that answer. Second, since machines can count corners as well as people can, triangularity is indeed 'independent of any power to produce effects in human observers'. Third, since I think that occurring in laws is what makes triangularity a property, I agree in substance with Alexander's claim that its 'conditional characterization [needs] appropriate generality' to show it 'to be genuinely dispositional', i.e. to be a real property.

However, the interesting question about a dispositional property F remains, as Alexander says, whether it is *essentially* dispositional, i.e. whether nothing could be F while lacking the 'conditional powers' that the laws F occurs in give it. This however is ambiguous, since properties occur in many laws, like all those containing temperatures listed in Section 6, and each law that F occurs in will give F-things a distinct conditional power. So something might have been F while lacking *some* of these powers, if not while lacking most or all of them. Thus, just as Alexander might have been a Labour Member of Parliament but not perhaps a microbe, so our relativistic masses (which

acceleration increases) might perhaps have been Newtonian (not increased by acceleration) but not temperatures.

Alexander thinks, however, that some individual laws, and hence powers, *are* essential to some properties, and he may be right. Indeed, a truthmaking consideration tempts me to the even stronger claim in Stephen Mumford's (1998: Ch. 10), that *all* properties necessitate *all* the laws they occur in. Take the example, in Section 1 above, of truths about what is visible in a mirror. To necessitate these we need not only the mirror, the objects it reflects and the light by which it does so, but also the laws of reflection. Yet, as I say in Section 7 of my (2000b),

> the ontology of laws is notoriously problematic, with candidates ranging from Humean regularities to relations between properties … It is tempting therefore to bypass the problem … by taking the existence of factual properties to entail the laws they occur in. For then we can dispense with laws as truthmakers, even for law statements, which can all be made true by the existence of the properties and relations they refer to.

However, while I feel this temptation, I have not yet succumbed to it. I cannot yet believe, for example, that masses *could* not be as unaffected by acceleration as Newton thought; and I do not despair of saying what in the world contingent laws of nature are. But if in the end no credible account of what laws are lets them be contingent, I may then have to follow Oscar Wilde's advice that 'the only way to get rid of a temptation is to yield to it'.

11 Arnold Koslow

The range of cases covered by Arnold Koslow's logic of natural possibilities is a revelation. Its removal of the concept's common restriction to truths and worlds is especially welcome to my reply to Tim Crane in Section 5, by making knowing-how even more like knowing-that. For although Arnie does not give the example, his theory shows how abilities are as much natural possibilities for know-how as intelligible truths are for propositional knowledge.

As a logic of possibility and necessity, Arnie's theory has one obvious defect, of which he is well aware, namely that on it 'necessarily x' does not always imply 'possibly x'. Its always doing so when x is a single natural possibility (i.e. a singleton of the power set N^* of the set N of such possibilities) seems to me not enough, since this does not cover every possibility we would naturally call 'natural', such as getting an odd number (1, 3 or 5) on a throw of a die. If, however, this is (as Arnie conjectured in an email) 'an artefact of the construction [he] gave for these possibilities', it should be remediable, and I hope it is.

But whether it is or not, one question that Arnie's list of kinds of possibilities prompts is what distinguishes them from each other. What, in particular, distinguishes the contingent and quantitative physical possibilities that I call 'chances' (Mellor 2000a), like a chance $ch(H)$ of a coin toss landing heads? I

think the answer is that, being contingent, simple statements of chance like '*ch*(H) = 0.4' need truthmakers, which most of Arnie's other possibilities, being necessarily possible, do not. I said in Section 1 that because 'P is contingent' and hence '~P is possible' are necessary if true, they need no truthmakers. Similarly for the sense in which truth and falsity are the possible truth values of any 'P' and '~P'. Similarly again for the necessary possibility of possible worlds, and of possible cases invoked in mathematical proofs.

Still, not all of Arnie's other possibilities are necessarily possible. Take the possible states and transitions ascribed by theories to systems, such as the possible orbits ascribed to planets by Newton's theory of gravity. If the theory is contingent, so are these possible orbits. However, given whatever makes the theory true, nothing more is needed to make just these orbits *possible*. It is statements of the *actual* orbits of planets that need something more to make them true. And so do statements of their *chances* of being actual, whether these be 1, on a deterministic theory, or something less, on an indeterministic theory: for no contingent 'P' or value of p are propositions of the form '*ch*(P)=p' complete truth-functions of 'P'.

This is why propositions like '*ch*(H) = 0.4' need to be made true by chances. Or, rather, since $ch(P) = 1 - ch(\sim P)$ for all P, by chance *distributions*, in this case the distribution ⟨0.4,0.6⟩ over ⟨H,~H⟩. But not all propositions about chances need truthmakers, because, for reasons already given, no truth-function of '*ch*(P) = p' needs one. In particular, therefore, '~(*ch*(H) = 0)' needs no truthmaker. But then, as I said in Section I of my (2000a), this can be true, i.e. (as I noted in Section 9) H can be made possible, 'not by there being a *ch*(H)>0 but by there being no *ch*(H) at all, zero or otherwise': a coin toss can land heads simply because nothing prevents it, whether or not it has any positive chance of doing so.

Arnie is therefore wrong to say that for me 'any B is a possibility if and only if it has a non-zero chance'. A non-zero chance is not necessary for this kind of possibility. Nor is it sufficient, since B may have more than one actual chance, as I also noted in Section 9, and one of its chances may be zero, which I say entails ~B. But provided we distinguish B's being *left* possible (i.e. not being ruled out) by a fact A from its being absolutely possible (i.e. ruled out by *no* fact), then, as I say in Section IV of my (2000a),

> there is no contradiction here either, even on the view that non-zero chances are real possibilities. For a toss's landing heads can easily be left possible to some extent by one fact about the toss, to a different extent by another, and made either necessary or impossible by a third.

None of this affects Arnie's case for his two main claims, that laws and explanations rule out possibilities. However, his argument for the latter claim does make one assumption I reject. This is that I require any B explained by any A to have chances with and without A, $ch_A(B)$ and $ch_{\sim A}(B)$, such that $ch_A(B) > ch_{\sim A}(B)$. Not so: some B (e.g. laws) have no chances, high or low, and

some A (e.g. least action explanations of trajectories) do not work by raising chances. It is only *causes* that I require to raise the chances of their effects, and many explanations are not causal. Yet these too rule out possibilities, and for Arnie's reason. For all his argument really needs is what he calls explanation's 'facticity', which makes the mere existence of any explanation A of B entail B, even if A itself does not. And this, as Arnie says, on my account rules out $ch(B)=0$, thereby ruling out a possibility, namely a possible value of $ch(B)$.

Change and time

12   Gonzalo Rodriguez-Pereyra

Asking how things can be, with no reason to think that they cannot be, is a bad habit to which many otherwise sensible philosophers are oddly prone. Knowledge is a case in point; and so is change. We all know what change is: things having at different times different properties or relations (different shapes, temperatures, distances, etc.) that they could not have together at a single time. Why is this a problem? Why should the inability of things to be simultaneously hot and cold at once stop them heating up or cooling down? I see no reason why it should, nor therefore any reason to ask how change is possible. That question would only make sense if we had no consistent theory of change, but we have several: what we face is not a famine but a glut; not a paradox but a problem of choice.

   The theory of change Gonzalo Rodriguez-Pereyra discusses is the relational theory, which makes changeable properties relations to times. He and I both reject many objections to this theory, notably the unargued denial of David Lewis (1986a: Ch. 4.2) and others that properties such as temperatures *are* relations. For this looks obvious only in present-tense statements like '*a* is hot', meaning '*a* is hot now', and on the B-theory that David, Gonzalo and I all accept, what makes this true at any B-time t is that a is hot at t. But then, given the relational form of '*a* is hot at t' and '*a* is cold at t'', a denial that temperatures are relations needs arguing, and I find the arguments Gonzalo cites as weak as he does. For example, the relational theory alters our ontology less than a theory of temporal parts, which makes what is hot at t (a-at-t) neither a itself nor what is cold at t' (a-at-t'). Nor does the theory make duplicates as hard to define as Mark Johnston (1987) thinks: at any time t' a duplicate of a at t is anything with all the same relations to t' that a has to t. And I agree with Gonzalo's answer to Katherine Hawley's (1998) objection, namely that a's temperature relation to t need not be entailed by any other properties of a, t, or their fusion.

   Gonzalo and I reject these and other objections to a relational theory of change, but we also reject the theory itself, only for different reasons. Mine is that, as things can be related at a distance, the theory fails to explain why things must be *at* any spacetime points where they have changeable properties. To this Gonzalo replies first that 'instantiation versions' of the theory do entail

this. But the first version he gives does not. For suppose a and b are events, with a earlier than b ($a<b$). This, on an 'adverbial' instantiation theory, requires a three-place instantiation relation I to link a, $<$ and b. But while that may make $<$ share a's and b's locations, it cannot make a and b coincide, or a would not be earlier than b. But then, if what makes a hot (H) at t is that I links a, H and t, this too cannot entail, as it should, that a is at t.

Gonzalo's other instantiation theories may do better, by building t into H or into I: I's linking a to H-at-t, or I-at-t's linking a to H, may well make a be at t. But they face other objections: for example, that building t into H, by making H-at-t differ from H-at-t', masks the difference between a's changing and its staying the same, and makes no sense of, for example, a's being hotter at t than b is at t'; while building t into I makes no sense even of $a<b$ – as we can see by asking at what time a is earlier than b.

In any case, whatever the relative merits of these theories, I have other reasons, indicated in Section 2, for rejecting all instantiation relations and hence any theory that invokes them. And Gonzalo's other reply to my objection to relational theories of change seems to me to miss the point. Of course, as he says and I admit, some relations require their relata to coincide in time and/or space. But most relations derived from changeable properties do not: a can be hotter than, or share the shape or colour of, objects anywhere in space and time. These properties imply nothing about coincidence. Why then, if they are relations to times, must their possessors be at those times, as we know they must? Building that necessity into these relations by definition is not an explanation, merely a restatement of the fact to be explained, hence my preference for a theory (outlined in Section 2) which, by keeping properties like H monadic, automatically gives a the spacetime location of any atomic fact whose constituents are a and H.

Gonzalo's own objection to the relational theory is that, by replacing incompatible properties like H (hot) and C (cold) with compatible relations to times, it denies rather than explains change. I cannot see this. H and C are if anything *less* compatible as relations than as properties, since no ordered pair $\langle a,t \rangle$ can be both H and C, whereas any a can be both H and C, albeit at different times. I think the relational theory explicates perfectly the concept of incompatibility between H and C that makes a's being first H and then C both possible and a case of change, by showing how anything can be H and C at different times but never at the same time. But since Gonzalo disagrees, and thinks for this reason, if not for mine, that the relational theory fails to solve 'the problem of change', I am happy to offer him my theory instead.

13   Nathan Oaklander

Nathan Oaklander shows in detail why the presentism of William Lane Craig (2000a,b; 2001) is as subject to McTaggart's (1908) contradiction as any other A-theory of time. This matters because it is from the flow of time, which makes events change their A-series locations from future to present to past, that

McTaggart derives his contradiction, by arguing that it requires all events to have all these mutually incompatible locations. Many of those who accept his argument believe therefore that presentists escape it, by holding that only what is present exists, and hence that nothing in reality ever has any other temporal location. This is why presentism is widely held to be the safest as well as the most radical A-theory of time.

I have nothing to add to Nathan's demolition of this delusion, and also of Craig's canard that I think B-relations, such as being earlier than, can be derived from the A-series. Instead I shall amplify their shared criticism of Arthur Prior's failure to add an ontology to his semantics of time, by showing how it makes his presentism both vacuous and question-begging.

In Prior's (1957: Ch. II) system, temporally unqualified sentences like '*Ga*' are taken to be present tense, i.e. to say that *a* is *G now*. To these the iterable operators '**P**' ('it has been the case that') and '**F**' ('it will be the case that') may be prefixed to make statements about the past or the future. Thus, '**P***Ga*' says that *a was G*, '**F***Ga*' that it *will be G*, '**FP***Ga*' that it *will have been G*, and so on. But all such sentences, however complex, are also about the present, since they all say that it is *now* the case that *a* was *G*, will be *G*, will have been *G*, etc. These sentences could therefore always be made explicitly present tense by prefixing an operator '**N**' ('it is now the case that') without changing their meanings or truth values. This shows, as Prior says, that '**N**' is redundant, and that *all* tensed truths are truths about the present; from which it follows, on Prior's (1971) view of facts as true propositions, that all temporal facts are present facts: hence Prior's (1970) presentism.

I indicated in Section 1 how a semantics for time can fail to settle its ontology, as Prior's does, by not saying what makes it true, and hence for Prior a fact, that *a* was *G*, will be *G*, will have been *G*, etc. Prior's failure stems from his purely semantic conception of a fact, which stops his system entailing the ontological doctrine, that whatever *makes* it true that *a* was *G*, will be *G*, will have been *G*, etc., is in the present. That is what makes his presentism vacuous.

It is also question-begging. For if, as it needs to assume, all truths are present tense, all B-truths must reduce to A-truths. In particular, the meanings of B-predicates like 'earlier' must follow from those of Prior's primitive operators '**P**' and '**F**' rather than vice versa, which Nathan, I and other B-theorists deny for many reasons, of which the following is one.

All facts or events, like Queen Anne's birth and death, must become more past at the same rate: otherwise, as time passes, Queen Anne's age at death could vary, which is absurd. Hence, for all A-propositions α and β, and for all real numbers M and N, the following past-tense version of Prior's (1957: Ch. II) axiom 5,

$$\text{if } \mathbf{P}_M \mathbf{P}_N(\alpha \,\&\, \beta), \text{ then } \mathbf{P}_{M+N}(\alpha \,\&\, \beta),$$

where '$\mathbf{P}_M$' means 'it was the case M units ago that', must be necessarily true. But what makes it so: Why can different facts and events not become more

past at different rates? The answer is obvious: past time intervals between any two facts or events are mere logical consequences of the interval between their becoming present, i.e. – for a presentist – between their coming to exist or occur. But then, to stop *that* interval varying over time, any statement of it must be a temporally invariant B-statement, like 'Queen Anne's birth occurs 49 years *earlier than* her death'. And as in this example, so in others. The axioms of Prior's presentist system, which express the semantics of his operators '**P**' and '**F**', cannot derive their necessity just from A-concepts: they must also invoke the irreducibly B-concept of some facts and events becoming present more or less *earlier* than others.

References

(Where two dates occur in a reference, the first is the date of first publication and the second that of the publication referred to.)

Armstrong, D. M. (1993) *A Materialist Theory of the Mind* (revised edition), New York: Routledge.
—— (1997) *A World of States of Affairs*, Cambridge, UK: Cambridge University Press.
Craig, W. L. (2000a) *The Tensed Theory of Time: A Critical Examination*, Dordrecht: Kluwer.
—— (2000b) *The Tenseless Theory of Time: A Critical Examination*, Dordrecht: Kluwer.
—— (2001) *Time and the Metaphysics of Relativity*, Dordrecht: Kluwer.
Crane, T. and Mellor, D. H. (1990) 'There is no question of physicalism [plus Postscript]', in P. K. Moser and J. D. Trout (eds) *Contemporary Materialism*, London: Routledge (1995).
Davidson, D. (1969) 'The individuation of events', in *Essays on Actions and Events*, Oxford: Clarendon Press (1980).
—— (1970) 'Events as particulars', in *Essays on Actions and Events*, Oxford: Clarendon Press (1980).
Hawley, K. (1998) 'Why temporary properties are not relations between objects and times', *Proceedings of the Aristotelian Society* 98: 211–16.
Hume, D. (1748) 'An enquiry concerning human understanding', in L. A. Selby-Bigge (ed.) *Enquiries concerning the Human Understanding and concerning the Principles of Morals*, Oxford: Clarendon Press (1902).
Jackson, F. (1986) 'What Mary didn't know', *Journal of Philosophy* 83: 291–5.
—— (1998) *From Metaphysics to Ethics*, Oxford: Clarendon Press.
Johnston, M. (1987) 'Is there a problem about persistence?' *Aristotelian Society* (Suppl.) 61: 107–35.
Kripke, S. A. (1971) 'Identity and necessity', in M. K. Munitz (ed.) *Identity and Individuation*, New York: New York University Press.
—— (1972) 'Naming and necessity', in D. Davidson and G. Harman (ed.) *Semantics of Natural Language*, Dordrecht: Reidel.
Lewis, D. (1968) 'Counterpart theory and quantified modal logic', *Journal of Philosophy* 65, 113–26.
—— (1973) *Counterfactuals*, Oxford: Basil Blackwell.
—— (1986a) *On the Plurality of Worlds*, Oxford: Basil Blackwell.

—— (1986b) 'Postscripts to "causation"', in *Philosophical Papers*, Vol. II, Oxford: Oxford University Press.

—— (1988) 'What experience teaches', in W. G. Lycan (ed.) *Mind and Cognition: a Reader*, Oxford: Basil Blackwell (1990).

—— (2001) 'Forget about the "Correspondence Theory of Truth"', *Analysis* 61: 275–9.

McTaggart, J. M. E. (1908) 'The unreality of time', *Mind* 18: 457–84.

Mellor, D. H. (1980) 'Consciousness and degrees of belief', in *Matters of Metaphysics*, Cambridge, UK: Cambridge University Press (1991).

—— (1981) *Real Time*, Cambridge, UK: Cambridge University Press.

—— (1990) 'Telling the truth', in D. H. Mellor (ed.) *Ways of Communicating*, Cambridge, UK: Cambridge University Press.

—— (1993a) 'How to believe a conditional', *Journal of Philosophy* 90: 233–48.

—— (1993b) 'Nothing like experience', *Proceedings of the Aristotelian Society* 93: 1–16.

—— (1995) *The Facts of Causation*, London: Routledge.

—— (1997) 'Properties and predicates', in D. H. Mellor and A. Oliver (ed.) *Properties*, Oxford: Oxford University Press.

—— (1998) *Real Time II*, London: Routledge.

—— (2000a) 'Possibility, chance and necessity', *Australasian Journal of Philosophy* 78: 16–27.

—— (2000b) 'The semantics and ontology of dispositions', *Mind* 109: 757–80.

—— (2001) '*Real Time II*: replies to Hinchliff, Paul and Perry', in L. N. Oaklander (ed.) *The Importance of Time: Proceedings of the Philosophy of Time Society 1995–2000*, Dordrecht: Kluwer.

Mellor, D. H. and Oliver, A. (eds) (1997) *Properties*, Oxford: Oxford University Press.

Menzies, P. (1989) 'probabilistic causation and causal processes: a critique of Lewis's, *Philosophy of Science* 56: 642–63.

Mumford, S. (1998) *Dispositions*, Oxford: Oxford University Press.

Nemirow, L. (1990) 'Physicalism and the cognitive role of acquaintance', in W. G. Lycan (ed.) *Mind and Cognition*, Oxford: Basil Blackwell.

Paul, L. A. (2001) 'Comments on D. H. Mellor's *Real Time II*', in L. N. Oaklander (ed.) *The Importance of Time: Proceedings of the Philosophy of Time Society 1995–2000*, Dordrecht: Kluwer.

Popper, K. R. (1990) *A World of Propensities*, Bristol: Thoemmes.

Prior, A. N. (1957) *Time and Modality*, London: Oxford University Press.

—— (1970) 'The notion of the present', *Studium Generale* 23: 245–8.

—— (1971) 'Propositions and facts', in P. T. Geach and A. J. P. Kenny (eds) *Objects of Thought*, Oxford: Clarendon Press.

Putnam, H. (1975) 'The meaning of "meaning"', in *Mind, Language and Reality*, Cambridge, UK: Cambridge University Press.

Shoemaker, S. (1980) 'Causality and properties', in P. van Inwagen (ed.) *Time and Cause*, Dordrecht: Reidel.

Whyte, J. T. (1990) 'Success semantics', *Analysis* 50: 149–57.

—— (1997) 'Success again: reply to Brandom and Godfrey-Smith', *Analysis* 57: 84–8.

Wittgenstein, L. (1922) *Tractatus Logico-Philosophicus*, London: Routledge.

D. H. Mellor

A bibliography

1965

'Connectivity, chance and ignorance', *British Journal for the Philosophy of Science* 16: 209–25.
'Experimental error and deducibility', *Philosophy of Science* 32: 105–22.
Review of L. Brillouin, *Scientific Uncertainty and Information*, *Nature* 205: 325.

1966

'Inexactness and explanation', *Philosophy of Science* 33: 345–59.
Review of P. Caws, *The Philosophy of Science: A Systematic Account*, *Journal of the Royal Society of Arts* 114: 332–3.

1967

'Connectivity, chance and ignorance [Note]', *British Journal for the Philosophy of Science* 18: 235–8.
'Imprecision and explanation', *Philosophy of Science* 34: 1–9.
Review of B. Ellis, *Basic Concepts of Measurement*, *British Journal for the Philosophy of Science* 17: 323–6.
Review of R. Harré, *The Anticipation of Nature*, *Ratio* 9: 93–5.

1968

'Models and analogies in science: Duhem *vs* Campbell?', *Isis* 59: 282–90.
'On literary truth', *Ratio* 10: 150–68.
'Two fallacies in Charles Taylor's *Explanation of Behaviour*', *Mind* 77: 124–6.
Review of A. Pfeiffer, *Dialogues on Fundamental Questions of Science and Philosophy*, *Cambridge Research* 3: 24–5.

1969

'Chance I', *Aristotelian Society* 43 (Suppl.): 11–36.
'God and probability', *Religious Studies* 5: 223–34. [Reprinted in K. E. Yandell (ed.) (1973) *God, Man and Religion*.]

'Physics and furniture', in N. Rescher (ed.) *Studies in Philosophy of Science*, Oxford: Basil Blackwell, pp. 171–87.

Review of M. Sherwood, *The Logic of Explanation in Psychoanalysis*, *Nature* 224: 621–2.

Review of R. G. Swinburne, *Space and Time*, *British Journal for the Philosophy of Science* 20: 366–70.

1970

Review of I. Lakatos, *The Problem of Inductive Logic*, *Philosophical Quarterly* 20: 405–6.

1971

The Matter of Chance, Cambridge, UK: Cambridge University Press.

Review of W. Sellars, *Science and Metaphysics*, *Ratio* 13: 93–5.

1972

Review of G. H. von Wright, *Explanation and Understanding*, *Times Higher Education Supplement*, 25 February: 16.

1973

'Materialism and phenomenal qualities II', *Aristotelian Society* 47 (Suppl.): 107–19.

'Do cultures exist?', in C. Renfrew (ed.) *The Explanation of Culture Change*, London: Duckworth, pp. 59–72.

'On some methodological misconceptions', in C. Renfrew (ed.) *The Explanation of Culture Change*, London: Duckworth, pp. 493–8.

'Reply to Mr Holborow', *Mind* 82: 106–7.

Review of A. J. Ayer, *Probability and Evidence*, *Philosophical Quarterly* 23: 272–4.

Review of R. G. Swinburne, *Introduction to Confirmation Theory*, *Times Literary Supplement*, 19 October: 1286.

1974

'In defense of dispositions', *Philosophical Review* 83; 157–81. [Reprinted in R. Tuomela (ed.) (1978) *Dispositions*, and in *Matters of Metaphysics* (1991).]

'Religious and secular statements', *Philosophy* 49: 33–46.

'Special relativity and present truth', *Analysis* 34: 74–7.

'New archaeology for old', *Cambridge Review* February, 71–2. [Review of C. Renfrew (ed.) *The Explanation of Culture Change: Models in Prehistory*.]

Review of P. Achinstein, *Law and Explanation*, *Mind* 83: 146–9.

1975

'Comment [on Wesley C. Salmon, 'Theoretical Explanation']', in S. Körner (ed.) *Explanation*, Oxford: Blackwell, pp. 146–52.

'Mind over materialism?', *Listener*, 9 January: 46–7. [Text of BBC Radio 3 broadcast, 24 December 1974.]

Review of G. N. Schlesinger, *Confirmation and Confirmability*, *Times Literary Supplement*, 14 February: 176.

Review of I. Hinckfuss, *The Existence of Space and Time*, *Times Literary Supplement*, 5 September: 993.

Review of J. L. Mackie, *The Cement of the Universe*, *Ratio* 17: 251–4.

Review of P. Suppes, *Probabilistic Metaphysics*, *Philosophical Books* 16: 30–2.

1976

'Probable explanation', *Australasian Journal of Philosophy* 54: 231–41.

1977

'Natural kinds', *British Journal for the Philosophy of Science* 28: 299–312. [Reprinted in *Matters of Metaphysics* (1991).]

'The Popper phenomenon', *Philosophy* 52: 195–202. [Review of P. A. Schilpp (ed.) (1974) *The Philosophy of Karl Popper*.]

1978

'Better than the stars', BBC Radio 3, 27 February. [Radio portrait of Frank Ramsey.]

'Conscious belief', *Proceedings of the Aristotelian Society* 78: 87–101.

'Reply to Professor Fetzer's "On Mellor on dispositions"', *Philosophia: Philosophical Quarterly of Israel* 7: 661–6.

Review of L. J. Cohen, *The Probable and the Provable*, *Times Literary Supplement*, 23 June: 708–9.

Review of S. P. Schwartz, *Naming*, *Necessity and Natural Kinds*, *Philosophy* 53: 126–7.

1979

'The possibility of prediction', *Proceedings of the British Academy* 65: 207–23.

1980

(editor) *Prospects for Pragmatism: Essays in Memory of F. P. Ramsey*, Cambridge, UK: Cambridge University Press.

(editor) *Science*, *Belief and Behaviour: Essays in Honour of R. B. Braithwaite*, Cambridge, UK: Cambridge University Press.

'Consciousness and degrees of belief', in *Prospects for Pragmatism*: 139–73. [Reprinted in *Matters of Metaphysics* (1991).]

'Necessities and universals in natural laws', in *Science*, *Belief and Behaviour*: 105–25. [Reprinted in *Matters of Metaphysics* (1991).]

'On things and causes in spacetime', *British Journal for the Philosophy of Science* 31: 282–8.

'The self from time to time', *Analysis* 40: 59–62.

'Comment [on Sydney Shoemaker, 'Properties, causation and projectibility']', in L. J. Cohen and M. B. Hesse (eds) (1980) *Applications of Inductive Logic*, Oxford: Oxford University Press, pp. 326–8.

1981

Real Time, Cambridge, UK: Cambridge University Press. [Chapter 6 reprinted revised in R. Le Poidevin and M. Macbeath (eds) (1993) *The Philosophy of Time*; Chapters 5 and 6 reprinted in L. N. Oaklander and Q. Smith (eds) (1994) *The New Theory of Time*.]

'"Thank goodness that's over"', *Ratio* 23: 20–30. [Reprinted in L. N. Oaklander and Q. Smith (eds) (1994) *The New Theory of Time*.]

'McTaggart, fixity and coming true', in R. Healey (ed.) *Reduction*, *Time and Reality*, Cambridge, UK: Cambridge University Press, pp. 79–97. [Reprinted in *Matters of Metaphysics* (1991).]

'Some problems about solving problems', in P. D. Asquith and I. Hacking (eds) *PSA 1978*, Vol. 2, Philosophy of Science Association: 522–9.

1982

'The reduction of society', *Philosophy* 57: 51–63.

'Chance and degrees of belief', in R. McLaughlin (ed.) *What? Where? When? Why?*, Dordrecht: Reidel, pp. 49–68. [Reprinted in *Matters of Metaphysics* (1991).]

'Probabilities for Explanation', in C. Renfrew (ed.) *Theory and Explanation in Archaeology*, London: Academic Press, pp. 57–63.

'Counting Corners Correctly', *Analysis* 42: 96–7.

'Chance (II)', in *Encyclopedia of Statistical Sciences* Vol. 1, New York: Wiley, pp. 405–11.

Review of H. Putnam, *Reason*, *Truth and History*, in *Times Literary Supplement*, 16 July: 774.

Review of W. H. Newton-Smith, *The Structure of Time*, *Philosophical Books* 23: 65–9.

1983

'Objective Decision Making', *Social Theory and Practice* 9: 289–310. [Reprinted in *Matters of Metaphysics* (1991).]

'MacBeath's soluble aspirin', *Ratio* 25: 89–92.

'The eponymous F. P. Ramsey', *Journal of Graph Theory* 7: 9–13.

Review of B. Skyrms, *Causal Necessity*, *British Journal for the Philosophy of Science* 34: 97–104.

1984

'What is computational psychology? II', *Aristotelian Society* 58 (Suppl.): 37–53.

Review of D. M. Armstrong, *What is a Law of Nature?*, *Times Literary Supplement*, 19 October: 1197.

Review of H. E. Kyburg, Jr, *Epistemology and Inference*, *British Journal for the Philosophy of Science* 35: 175–9.

1985

Review of R. G. Swinburne, *Space*, *Time and Causality*, *Philosophical Books* 26: 243–5.

1986

'Tense's tenseless truth conditions', *Analysis* 46: 167–72.

'History without the flow of time', *Neue Zeitschrift für Systematische Theologie und Religionsphilosophie* 28: 68–76.

1987

'Fixed past, unfixed future', in B. Taylor (ed.) *Michael Dummett: Contributions to Philosophy*, Dordrecht: Nijhoff, pp. 166–86.

'The singularly affecting facts of causation', in P. Pettit, R. Sylvan and J. Norman (eds) *Metaphysics and Morality*, New York: Blackwell, pp. 111–36. [Reprinted in *Matters of Metaphysics* (1991).]

'Indeterministic causation', *African Philosophical Inquiry* 1: 157–75.

'Statistical causality [in Russian]', *Voprosy Filosofii* 80–90.

(with R. Le Poidevin) 'Time, change and the "indexical fallacy"', *Mind* 96: 534–8.

1988

The Warrant of Induction, Cambridge, UK: Cambridge University Press. [Inaugural Lecture. Reprinted in *Matters of Metaphysics* (1991).]

'How much of the mind is a computer?', in P. Slezak and W. R. Albury (eds) *Computers, Brains and Minds*, Dordrecht: Kluwer, pp. 47–69. [Reprinted in *Matters of Metaphysics* (1991).]

'I and now', *Proceedings of the Aristotelian Society* 89: 79–94. [Reprinted in *Matters of Metaphysics* (1991).]

'On raising the chances of effects', in J. H. Fetzer (ed.) *Probability and Causality*, Dordrecht: Kluwer. pp. 229–39. [Reprinted in *Matters of Metaphysics* (1991).]

'Crane's waterfall illusion', *Analysis* 48: 147–50.

1989

'Induction is warranted', *Analysis* 49: 5–7.

'Subjektivno znanje', *Filozofske Studije* 21: 3–18.

1990

(editor) *F. P. Ramsey: Philosophical Papers*, Cambridge, UK: Cambridge University Press. [Japanese translation by K. Ito (1996).]

'Introduction' to *F. P. Ramsey: Philosophical Papers*: xi–xxiii.

(editor) *Ways of Communicating*, Cambridge, UK: Cambridge University Press. [Italian translation *La Comunicazione*, Bari: Edizioni Dedalo (1992); Portuguese translation *Formas de Comunicação*, Lisbon: Teorema (1993).]

'Telling the truth', in *Ways of Communicating*: 81–95. [Italian translation 'Dire la verità'; Portuguese translation 'Falar Verdade'.]

(with T. Crane) 'There is no question of physicalism', *Mind* 99: 185–206. [Reprinted in Mellor (1991) *Matters of Metaphysics*, and with Postscript in P. K. Moser and J. D. Trout (eds) (1995) *Contemporary Materialism*. Romanian translation 'Nu Poate fi Vorba de fizicalism', in A. Botez (ed.) (1993) *Realism si Relativism in Filosofia Stiintei Contemporane*.]

'Laws, chances and properties', *International Studies in the Philosophy of Science* 4: 159–70. [Reprinted revised in *Matters of Metaphysics* (1991).]

1991

Matters of Metaphysics, Cambridge, UK: Cambridge University Press.

'Analytic philosophy and the self', in *Matters of Metaphysics*: 1–16. [Translations: 'Analytische Philosophie und das Selbst', in P. Koslowski (ed.) (1991) *Orientierung durch Philosophie*, Tübingen: J. C. B. Mohr (Paul Siebeck); 'Filozofia Analityczna i Jazn', in P. Gutowski and T. Szubka (eds) (1998) *Filozofia Brytyjska u Schylku XX Wieku*, Lublin: Towarzystwo Naukowe Kul.]

'Properties and predicates', in *Matters of Metaphysics*: 170–82. [Reprinted with reply by D. M. Armstrong in J. Bacon, K. Campbell and L. Reinhardt (eds) (1993) *Ontology, Causality and Mind*, and revised in D. H. Mellor and A. Oliver (eds) (1997) *Properties*.]

'Causation and the direction of time', *Erkenntnis* 35: 191–203.

'Dispositions', in H. Burkhardt and B. Smith (eds) *Handbook of Metaphysics and Ontology*, Vol. 2, Munich: Philosophia, p. 221.

1992

'Probability and the evidence of our senses', in A. P. Griffiths (ed.) *A. J. Ayer: Memorial Essays*, Cambridge, UK: Cambridge University Press, pp. 117–28.

'There are no conjunctive universals', *Analysis* 52: 97–103.

1993

'How to believe a conditional', *Journal of Philosophy* 90: 233–48.

'Nothing like experience', *Proceedings of the Aristotelian Society* 93: 1–16. [Translation in *Filozoficky Casopis* 42 (1994): 985–99.]

'Supervenience? No chance! Reply to Menuge', *Analysis* 53: 236–9.

'An interview with Hugh Mellor', *Cogito* 7: 3–10.

1994

'Philosophe de la Science à Cambridge en dialogue avec Angela Botez', *Revista de Filosofie (Romania)* 41: 377–9.

Review of L. E. Hahn (ed.) *The Philosophy of A. J. Ayer*, *Philosophy* 69: 107–10.

1995

The Facts of Causation, London: Routledge.

'Cambridge philosophers I: F. P. Ramsey', *Philosophy* 70: 243–62. [Based on the broadcast 'Better than the Stars' (1978).]

1997

(editor, with A. Oliver) *Properties*, Oxford: Oxford University Press.

(with A. Oliver) 'Introduction' to *Properties*: 1–33

'Reply to Hinckfuss', *Philosophical Books* 38: 8–11.

1998

Real Time II, London: Routledge.

'Campbell, Norman Robert', in E. J. Craig (ed.) *Routledge Encyclopedia of Philosophy*, Vol. 2, London: Routledge, pp. 193–4.

'Events', in *Routledge Encyclopedia of Philosophy*, 3: 461–3.

'Ramsey, Frank Plumpton', in *Routledge Encyclopedia of Philosophy*, Vol. 8: 44–9.

'Transcendental tense I', *Aristotelian Society* 72 (Suppl.): 29–43.

Review of M. Tooley, *Time*, *Tense*, *and Causation*, *Philosophy* 73: 631–4.

1999

Review of D. Cockburn, *Other Times: Philosophical Perspectives on Past*, *Present and Future*, *Philosophical Review* 108: 428–30.

2000

'Possibility, chance and necessity', *Australasian Journal of Philosophy* 78: 16–27.

'The semantics and ontology of dispositions', *Mind* 109: 757–80.

'Equally effective causes', *Analysis* 60: 71–3.

'The point of refinement', *Analysis* 60: 243–6.

2001

'The time of our lives', in A. O'Hear (ed.) *Philosophy at the New Millennium*, Cambridge, UK: Cambridge University Press, pp. 45–59.

'Transparent and Opaque Causation', in W. Spohn, M. Ledwig and M. Esfeld (eds) *Current Issues in Causation*, Paderborn: Mentis, pp. 9–20.

'*Real Time II*: Replies to Hinchliff, Paul and Perry', in L. N. Oaklander (ed.) *The Importance of Time: Proceedings of the Philosophy of Time Society 1995–2000*, Dordrecht: Kluwer, pp. 95–102.

'Realistic Metaphysics: An Interview with D. H. Mellor', *Theoria* 67: 4–21.

2002

'Time travel', in K. Ridderbos (ed.) *Time*, Cambridge, UK: Cambridge University Press, pp. 46–64.

'Too many universes', in N. A. Manson (ed.) *God and Design: The Teleological Argument and Modern Science*, London: Routledge.

2003

'For facts as causes and effects', in J. Collins, N. Hall and L. A. Paul (eds) *Causation and Conditionals*, Cambridge, MA: MIT Press.

'Time', in F. Jackson and M. Smith (eds) *Oxford Handbook of Contemporary Philosophy*, Oxford: Oxford University Press.

Index