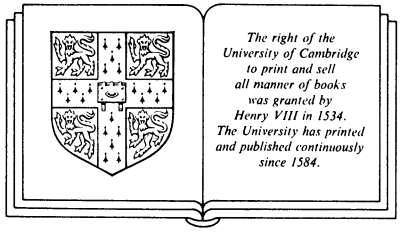
This selection of D. H. Mellor's papers demonstrates the wide-ranging originality of his work. It gathers together sixteen major papers on related topics written over the last seventeen years. Together they form a complete modern metaphysics. The first five papers are on aspects of the mind: on our 'selves', their supposed subjectivity and how we refer to them, on the nature of conscious belief and on computational and physicalist theories of the mind. The next five papers deal with dispositions, natural kinds, laws of nature and how they involve natural necessity, universals and objective chances, and the relation between properties and predicates. Then follow three papers about the relations between time, change and causation, the nature of individual causes and effects and of the causal relation between them, and how causation depends on chance. The last three papers discuss the relation between chance and degrees of belief, give a solution to the problem of induction, and argue for an objective interpretation of decision theory. Two of the papers included here have been especially written for this volume, another has been revised for it, and many have hitherto been relatively inaccessible. A substantial introduction summarises the papers and indicates the connections between them.

Matters of Metaphysics

Matters of Metaphysics

D. H. MELLOR
Professor of Philosophy, University of Cambridge



CAMBRIDGE UNIVERSITY PRESS
Cambridge
New York Port Chester
Melbourne Sydney

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo

Cambridge University Press
The Edinburgh Building, Cambridge CB2 2RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9780521411172

© Cambridge University Press 1991

This publication is in copyright. Subject to statutory exception
and to the provisions of relevant collective licensing agreements,
no reproduction of any part may take place without
the written permission of Cambridge University Press.

First published 1991

A catalogue record for this publication is available from the British Library

ISBN-13 978-0-521-41117-2 hardback
ISBN-10 0-521-41117-3 hardback

Transferred to digital printing 2005

For Tim Crane and Jamie Whyte

Contents

	Preface	*page*	xi
	Introduction		xv
	Minds		
1	Analytic philosophy and the self		1
2	I and now (1989)		17
3	Consciousness and degrees of belief (1980)		30
4	How much of the mind is a computer? (1988)		61
5	(with Tim Crane) There is no question of physicalism (1990)		82
	Properties and laws		
6	In defence of dispositions (1974)		104
7	Natural kinds (1977)		123
8	Necessities and universals in natural laws (1980)		136
9	Laws, chances and properties (1990)		154
10	Properties and predicates		170
	Causation		
11	McTaggart, fixity and coming true (1981)		183
12	The singularly affecting facts of causation (1987)		201
13	On raising the chances of effects (1988)		225
	Prediction and decision		
14	Chance and degrees of belief (1982)		235
15	The warrant of induction (1988)		254
16	Objective decision making (1983)		269
	References		288

Preface

Two of the papers in this volume, chapters 1 and 10, are published here for the first time. Chapter 1 is also due to appear in German as:

'Analytische Philosophie und das Selbst', *Orientierung durch Philosophie*, ed. P. Koslowski, 1991, Tübingen: J. C. B. Mohr (Paul Siebeck).

Chapter 10, 'Properties and predicates', is also due to appear, with a reply by D. M. Armstrong, in:

Ontology, Causality, and Mind, ed. K. Campbell et al., Cambridge: Cambridge University Press.

Chapter 9 has been significantly revised from its first publication as:

'Laws, chances and properties', *International Studies in the Philosophy of Science* **4** (1990), 159–70. Reprinted by permission of the Editors.

The other chapters are reprinted here with no substantial changes. The details of their original publication are:

Chapter 2: 'I and now', *Proceedings of the Aristotelian Society* **89** (1989), 79–94. © The Aristotelian Society 1989. Reprinted by courtesy of the Editor.

Chapter 3: 'Consciousness and degrees of belief', *Prospects for Pragmatism*, ed. D. H. Mellor, Cambridge: Cambridge University Press (1980), 139–73.

Chapter 4: 'How much of the mind is a computer?', *Computers, Brains and Minds*, ed. P. Slezak and W. R. Albury, Dordrecht: Kluwer (1988), 47–69. © 1989 by Kluwer Academic Publishers. Reprinted by permission of Kluwer Academic Publishers.

Chapter 5 (with Tim Crane): 'There is no question of physicalism', *Mind* **99** (1990), 185–206. Reprinted by permission of Tim Crane and Oxford University Press.

Chapter 6: 'In defense of dispositions', *Philosophical Review* **83** (1974), 157–81. Reprinted by permission of the publisher.

Chapter 7: 'Natural kinds', *British Journal for the Philosophy of Science* **28** (1977), 299–312. Reprinted by permission of the Editor.

Chapter 8: 'Necessities and universals in natural laws', *Science, Belief and Behaviour*, ed. D. H. Mellor, Cambridge: Cambridge University Press (1980), 105–25.

Chapter 11: 'McTaggart, fixity and coming true', *Reduction, Time and Reality*, ed. R. Healey, Cambridge: Cambridge University Press (1981), 79–97.

Chapter 12: 'The singularly affecting facts of causation', *Metaphysics and Morality: Essays in Honour of J. J. C. Smart*, ed. P. Pettit et al., Oxford: Blackwell (1987), 111–36.

Chapter 13: 'On raising the chances of effects', *Probability and Causality*, ed. J. H. Fetzer, Dordrecht: Reidel (1988), 229–39. © 1988 by D. Reidel Publishing Company. Reprinted by permission of Kluwer Academic Publishers.

Chapter 14: 'Chance and degrees of belief', *What? Where? When? Why?*, ed. R. McLaughlin, Dordrecht: Reidel (1982), 49–68. Copyright © 1982 by D. Reidel Publishing Company. Reprinted by permission of Kluwer Academic Publishers.

Chapter 15: *The Warrant of Induction*, Cambridge: Cambridge University Press (1988).

Chapter 16: 'Objective decision making', *Social Theory and Practice* **9** (1983), 289–309.

I am grateful to those concerned for permission to republish this material in this book, and to Cambridge University Press for enabling me to do so. For the material itself I am beholden to many more people than I can mention here. I acknowledge specific debts in references to works listed at the end of the book; but since the bibliographies and acknowledgements of my previously

published articles have not been reprinted, I must emphasise that I am indebted to far more of the literature than I refer to. But my deepest debts are to those with whom I have discussed these subjects for over twenty-five years. I owe much to many philosophers I have met elsewhere in Britain, Australasia (whose hard thinking and straight talking I find especially congenial), the United States, Ireland, Germany, Sweden, Russia, Yugoslavia and Holland. But I owe most to my colleagues and students in Cambridge, who have taught more than I can say about philosophy and how to do it. Though I cannot name them all, I thank them all; and I hope they will let the name of my mentor, friend and philosophical exemplar, Richard Braithwaite, who died last year, stand for all I owe to Cambridge philosophy.

Cambridge
March 1991

Introduction

The papers that follow have been written over a period of twenty years and are those which I now think most worth preserving, both individually and collectively. They contain inevitable repetitions, but also important connections, and between them constitute a metaphysics which I hope will interest other philosophers and students of the subject. As they were written at different times and for different purposes, they naturally vary in depth, detail and style, variations which I have not tried to iron out. In particular, I have kept the style of the two chapters (1 and 15) originally written as lectures. The first of these, 'Analytic philosophy and the self', was given in Leibniz' house in Hannover on 16 November 1989 in a series of public lectures on the rôle of philosophy today. It starts with a statement of what I take to be the main uses and virtues of analytic philosophy, and that made it seem a good first chapter for this book. After this, however, readers will find nothing more about philosophy itself, since the philosophy of philosophy seems to me among its least enlightening branches. 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 – 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 applying them, and my case for my method, such as it is, rests on the contents of the ensuing chapters.

Minds The first five chapters concern aspects of the mind. Chapters 1 and 2 contain work on the self motivated by scepticism about recent claims for its subjectivity. In chapter 1, I attack the myth of subjectivity by showing how a wholly objective world can incorporate all our apparently subjective first-person knowledge of ourselves and of the world as it appears from our own perspective. Chapter 2, 'I and now' (1989), supplements chapter 1 by showing how the causal mechanisms of our minds enable our subjective – first-person present tense – thought and speech to refer directly and infallibly to ourselves and to the present time. This remarkable ability turns out to need neither subjective facts or selves, internal representations of ourselves or of the

present, nor any problematic capacity for (or concepts of) infallible self-reference or self-knowledge. All it needs is causal contiguity: the fact that causes are contiguous to their immediate effects. That is all the mental machinery it takes to put us and the present time into the truth conditions of our subjective statements and beliefs.

Subjective beliefs figure also in the theory of conscious belief developed in chapter 3, 'Consciousness and degrees of belief' (1980), written for a volume of essays to commemorate the fiftieth anniversary of the death of F. P. Ramsey. Our direct if fallible awareness of what we believe poses a problem for the functionalist accounts of beliefs pioneered by Ramsey, i.e. accounts which define beliefs by how our senses cause them, how they interact with each other and with desires and other states of mind, and how they combine with desires to cause actions. The problem is to say what conscious belief (which I call assent) is, how it arises, and how it relates to its unconscious counterpart. I argue that assenting is believing one believes, a 'second order' subjective belief caused by one's 'first order' beliefs *via* an inner sense. I show how this accommodates degrees of both belief and assent – and justifies taking subjective probability to measure actual, rather than merely rational, degrees of belief. Finally I tackle the apparent and obviously false consequence (C) that assenting to any proposition entails assenting to every proposition which it entails and of which we're conscious. My treatment of this problem here is inconclusive, but I now think I could accommodate (C)'s falsity by taking account of limitations in the causal mechanisms of belief formation discussed in chapter 4. But that account remains to be given, as does an answer to recent apparent counter-examples to my equation of assent with second order belief – an equation which, however, failing any other account of assent, I still think I believe.

Chapter 4, 'How much of the mind is a computer?' (1988), was written in response to some extravagant claims for computational psychology. In it I first consider what it might usefully mean to claim that mental processes are computations, i.e. causal transformations of information embodied in representations. I argue that this claim is vacuous unless computation is understood as part of a process of forming beliefs, but that this does not preclude non-trivial information-processing theories of perception and inference. However, I then argue that there can be no such theories of mental processes involving other mental states such as desire which, unlike beliefs, are not essentially truth-seeking. Most of the mind, I therefore conclude, is not a computer.

In chapter 5, 'There is no question of physicalism' (1990), Tim Crane and I attack the common but unclear assumption in much recent philosophy of mind that mental states and processes are all really physical. To avoid vacuity, this thesis needs a non-question-begging definition of 'physical' which will prevent the mental counting as physical in its own right; and we argue that no such definition exists. First we show why the physicalist cannot define the physical either as what reduces to physics, or as what has causes or effects. Then we show why he cannot define it as the non-intentional or as the law-governed: since all the problematic features of intentionality occur also in physics, and laws occur also in psychology. Finally we argue that the mental does not even supervene on the physical, and conclude that the whole question of physicalism is as trivial as the doctrine itself is false.

Properties and laws Chapters 6 to 10 concern natural properties, kinds and laws. Getting the right account of these is important not only in itself, but for the light it sheds on other metaphysical matters, including the metaphysics of the mind. Thus the misconception that laws are necessary has led many to deny the psychological and psychophysical laws discussed in chapter 5. Conversely, many attempts to reduce mental states to behavioural dispositions have been motivated by the anti-realist view of dispositions which I attack in chapter 6. In this chapter, 'In defence of dispositions' (1974), I argue that dispositions, both physical and mental, are real properties of things and people: i.e. that changes in them have real causes and effects. I argue moreover that dispositions have these causal powers in their own right, needing no non-dispositional properties to give them a real basis, such as molecular structures are supposed to provide for solubility, or brain states for mental dispositions like beliefs and desires. For these properties too are only dispositions: like inertial masses, which are no less real for being nothing but dispositions which objects have to accelerate under applied forces.

In chapter 7, 'Natural kinds' (1977), I turn from the supposed bases of dispositions to such supposedly essential properties of natural kinds as being H_2O, which essentialists say that water not only is but must be. The paper was written to rebut two well-known semantic arguments for these essences, one based on an anti-Fregean account of the extension of terms like 'water', the other on the necessary truth of identity statements like 'water is H_2O'. Both arguments fail: the first by having premises which both fail to be true of real kind terms and fail to entail essentialism; the second by begging the question, since 'water is H_2O' won't even be true, never mind necessary, unless being H_2O is an essential property of water – which, I argue, it isn't. And as for

H_2O, so in general: apparently essential properties are either not even shared by every actual thing of the kind, or their importance as properties is evidently more a feature of our theories than of the world.

Chapter 8, 'Necessities and universals in natural laws' (1980), was written for a festschrift for Richard Braithwaite, complementing a paper supporting Braithwaite's Humean view of laws by attacking two non-Humean alternatives. One of these takes laws to be metaphysically necessary, the other takes them to be second order relations between the universals (properties and relations) involved. The implausible metaphysics of these views could be justified, I maintain, only by yielding better solutions than Humeans can provide to such problems as how an uninstantiated law (like Newton's first law of motion) can hold non-trivially – which, I argue, they fail to do.

In chapter 9, 'Laws, chances and properties' (1990, revised), I supplement the previous chapter's critique of rival views with a unified account of both deterministic and indeterministic laws, based on the probabilities which, I argue, all laws contain. As in my (1971) *The Matter of Chance*, I take these probabilities to be real single-case chances, an interpretation modified and further defended in chapter 14. I then identify the 'natural necessity' of deterministic laws with the chance of 1 which they give their corresponding regularities. Laws in general I take to be embodied in the actual properties and relations (including chances) they contain, where actual properties and relations are those that would be quantified over by the Ramsey sentence of the conjunction of all true law statements.

Chapter 10, 'Properties and predicates', looks at how closely universals characterised non-semantically – as in chapter 9 – correspond to our general concepts, i.e. to the meanings of our predicates. I argue that they correspond hardly at all. I show how even if the effects on our senses of a single property fixed the extension of a predicate like 'red', the property would still not fix (though its laws would constrain) the predicate's connotations. I argue moreover that in fact few if any predicates, except some of those used to state natural laws, even correspond to single universals, let alone have them as anything like their meanings.

Causation In chapters 11 to 13 I turn from properties and laws to causation, and first to the link between causation and time. In chapter 11, 'McTaggart, fixity and coming true' (1981), I first give a tenseless account of change, using arguments which are taken further in my (1981) *Real Time*. I then use this to attack the idea that causation gets its temporal direction from the flow of time, which makes events necessary when they or their sufficient causes become

'fixed' (or the tenseless propositions saying they or their causes occur 'come true') by becoming present. I argue that these concepts of 'fixity' and 'coming true' either entail the contradictions that McTaggart showed to be inherent in the flow of time, or reduce to tenseless temporal relations which fail to give causation its temporal direction.

Chapter 12, 'The singularly affecting facts of causation' (1987), is about the relata of causation. In it I argue that singular causation primarily links facts rather than particular events. I rebut a well-known argument to the contrary, and show how causation linking events (Don's fall caused his death) follows from causation linking existential facts about them (Don died because he fell: i.e., there was a death of Don because there was a fall of Don). I argue moreover that the temporal order which (I argue in *Real Time*) causal order entails therefore also primarily links facts rather than events or spacetime regions; and infer from this that spacetime regions are not events but are individuated by the spatio-temporal relations of facts located at them.

In chapter 13, 'On raising the chances of effects' (1988), I argue that although causes need not determine their effects, they must at least raise their chances. The basis of the argument is what I call causation's 'means–end' connotation: namely, that when effects are ends, their causes are *ipso facto* means to them. I define the means–end relation without appeal to causation by using the expected utility principle of non-causal decision theory – requiring only that the probabilities involved be objective chances rather than subjective probabilities – and hence derive the required result. I then show how to read 'raise the chances' non-causally, as saying that effects' chances are greater in the circumstances with their causes than without them, dealing with apparent counter-examples by showing how these conditionals can be made true by probabilistic dispositions. Finally I show how a cause's efficacy increases the more it raises its effects' chances, thus explaining our prejudice in favour of deterministic causation.

Prediction and decision In the last three papers I defend my concept of objective chance and show how it affects what we should believe, predict and decide to do. Chapter 14, 'Chance and degrees of belief' (1982), was written in response to a critique by Wesley Salmon of my belief-based version of the so-called propensity theory of chance. In it I develop the analogy between chance and secondary qualities like colour, both of which are dispositional properties of events or objects defined by the mental states (degrees of belief, visual sensations) that constitute veridical perceptions of them. However, drawing on the arguments of chapter 6, I deny that this shows chances not to

be real properties, but argue that it does show how their existence, and the rationality of acting on them, can be compatible with an underlying determinism.

Chapter 15, 'The warrant of induction' (1988), was my Inaugural Lecture as Professor of Philosophy at Cambridge University. This seemed an appropriate occasion to work out and defend the inductive solution to the problem of induction suggested by Ramsey and further developed by Braithwaite and others. In it I defend inductive solutions against charges of vicious circularity by arguing generally that what warrants a belief is its having a very high chance of being true, whether or not the believer knows it has. I then apply this criterion to beliefs formed by inference, and show how the habit of forming inferential habits by induction warrants them by making them most likely to be formed when the conclusion of the inference has a high chance of being true if the premise is true.

Chapter 16, 'Objective decision making' (1983), defends an objective reading of prescriptive decision theory against orthodox subjective and rationalist readings. I say doctors are right to prescribe drugs for me, not because I think (or have reason to think) they will work, but because they probably *will* work. I argue therefore that one should only maximise objective expected utilities, i.e. expected utilities incorporating chances rather than merely subjective or epistemic probabilities. I defend this view against widespread ontological and epistemic objections, using arguments from chapters 6, 14 and 15 to defend objective utilities and chances, and to show that they are in fact easier to know than subjective utilities and probabilities.

1 *Analytic philosophy and the self*

1 Introduction Bishop Berkeley said in 1710, in the introduction to *The Principles of Human Knowledge*, 'Upon the whole, I am inclined to think that the far greater part, if not all, of those difficulties which have hitherto amused philosophers, and blocked up the way to knowledge, are entirely owing to ourselves – that we have first raised a dust and then complain we cannot see' (Berkeley 1710 p. 46).

Those remarks of Berkeley's seem to me just as true now as they were in 1710. Indeed matters are in some ways worse now than they were then. For one thing, philosophers today are too rarely amused by the difficulties that block the way to knowledge. They should be amused, because philosophy has to deal amongst other things with the limits of what makes sense: that is, with the boundary between sense and nonsense, which is the very stuff of humour. Take this example from Lewis Carroll's *Through the Looking-Glass*:

> 'Who did you pass on the road?' the King went on, holding out his hand to the Messenger for some more hay.
>
> 'Nobody', said the Messenger.
>
> 'Quite right', said the King: 'this young lady saw him too. So of course Nobody walks slower than you. '
>
> 'I do my best', the Messenger said in a sullen tone. 'I'm sure nobody walks much faster than I do!'
>
> 'He can't do that', said the King, 'or else he'd have been here first.' (Carroll 1887 pp. 85–6)

It takes a philosopher to see *why* this is funny, to see why it's nonsense to talk of Nobody as if he and she (Nobody is both male and female ...) were a being of some kind. The reason is, of course, that although the word 'Nobody' looks like the name of some being, it really isn't a name at all: it's a way of saying that there was *not* a being who walked either slower or faster than the Messenger. Now that's a pretty trivial piece of philosophical analysis, which anyone could do: but as we shall see, there is much more serious (and more

misleading) nonsense than Lewis Carroll's around, which it takes rather more analysis to expose and explain.

To expose nonsense, however, we must first detect it: we need a nose for nonsense. And, as Ramsey said of Wittgenstein's proposition that philosophy itself is nonsense, 'we must then take seriously that it is nonsense, and not pretend, as Wittgenstein does, that it is important nonsense' (Ramsey 1929c p. 1). Now I don't think philosophy is nonsense, but I do think it includes taking the fact of nonsense seriously and saying why it's nonsense. To do that, however, we need to be suitably amused by jokes like the one about Nobody, and to distinguish taking them seriously from pretending they're important. But not all philosophers are suitably amused. Some I fear lack the serious sense of humour, and with it the nose for nonsense, that good philosophy needs. And that is a serious defect. For without a nose for nonsense, philosophers run a real risk of talking nonsense themselves, and (unlike Lewis Carroll) of persuading themselves and others that it's important nonsense.

None of this would matter much if philosophy were read and judged only by other philosophers, as mathematics is by mathematicians, who can, on the whole, tell when their colleagues are talking nonsense. But it isn't, even though perhaps it should be, since philosophy is really no more of a spectator sport than mathematics is – by which I mean that it's not like poetry, for example, which you needn't be a poet to judge, whereas you do need to be a philosopher to judge philosophy, just as you need to be a mathematician to judge mathematics. Of course philosophy, like mathematics, is also read by outsiders who don't want to judge it, but rather to take it on trust and use it, just as physicists use mathematics. But not many outsiders want philosophy to do physics with: on the whole, they want it to provide a kind of secular substitute for religion. In other words, they want their philosophers to be gurus. And the last thing disciples want in gurus is a sense of humour: it's inimical to the air of authority which attracts disciples to gurus in the first place. So when philosophical gurus raise a dust by talking important-sounding nonsense, their disciples, far from *complaining* that they cannot see, are all the more impressed by the profound obscurity of the proffered view. In philosophy, therefore, as in religion and medicine, a gullible public will often give much fame and fortune to mystery-mongers.

What has all this to do with *analytic* philosophy? Well, to pursue Berkeley's metaphor, philosophical analysis is, as even my trivial example of it illustrates, a kind of intellectual sprinkler system, whose function is to lay the conceptual dust which obscures our view of the world. This indeed is one of its primary objects: to detect and dissipate the bogus mysteries which nonsense generates,

like Lewis Carroll's little mystery about Nobody, so that the world's real mysteries can be more clearly seen and thereby – we hope – better appreciated and understood.

In this sense, good philosophy has always been analytic. Analysis is more a matter of technique than of doctrine, and it is as evident in Plato, Aristotle, Aquinas, Leibniz, Hume, Kant and Mill as in any modern analytic philosopher. What if anything distinguishes analytic philosophy, so-called, is that it not only uses analytic techniques, but is explicitly concerned to develop and assess them: not of course as ends in themselves, but as means to philosophical understanding. But not of course the only means, since an analyst always needs non-analytic material to analyse. Analysis can no more provide a complete philosophy on its own than – for example – democracy can provide a complete politics: because, obviously, accepting the principle of majority rule doesn't tell you whom or what to vote for, or why. No political democrat, in other words, can *just* be a democrat; and in much the same way, no philosophical analyst can *just* be an analyst. Which is not of course to deny that analysis matters, just as democracy matters; nor to deny that it can conflict with philosophical nonsense (like the being of Nobody), just as democracy can conflict with political nonsense (like the one-party state).

But whereas everyone can feel that democracy matters, and can more or less see why, it is less obvious to non-philosophers why philosophical analysis matters. If philosophy in general is not really a spectator sport, what can analytic philosophy, in particular, offer to the rest of society? Well, I could say, for a start, that it offers, because it demands and encourages, a socially desirable temperament. A nose for nonsense isn't only an asset in philosophy. A sense of humour, and hence of proportion, is a powerful antidote to political and religious fanaticism. An insistence on explicit discursive understanding where it can be had, as opposed to obscure intimations of ineffable insight, is a great deterrent to charlatanism of all kinds. A commitment to truth, and hence to basing one's beliefs on evidence rather than on wishful thinking (however high-minded), is essential not only to good science, but to all serious attempts to acquire knowledge and understanding about anything, including ourselves. And the feeling for reason which analysis gratifies helps to combat a recurrent tendency to elevate feeling at the expense of reason, as if they were opposed, and as if we didn't need both.

Society, however, is not only indebted to the temperament that analytic philosophy fosters. The results of analysis too have had many uses outside philosophy itself, although I don't wish to exaggerate them, or to accept that they provide its main justification: philosophy, like mathematics, has a value of

its own, independent of its applications. Still, those applications are remarkable enough: ranging from the invention of computers (based on analyses of the concepts of mathematical proof and truth) to debates on abortion, which turn on concepts of life and of humanity whose analysis is far too important to be left to people with particular religious (or anti-religious) axes to grind.

But besides all this, I think analytic philosophy serves society most distinctively when it increases our understanding by clarifying concepts that concern everyone, whether they are philosophers or not. And in the rest of this lecture I should like to illustrate how it can do that by discussing one such concept, the concept of the self.[1]

2 The self We are all selves: that is, we are not just objects in the world, like sticks and stones, but also subjects, who experience the world, including ourselves and others; who learn about it, try to understand it, and value or disvalue various aspects of it; and who interact with it, both individually and collectively, because we want things to happen in it and to it, and in particular in and to ourselves and other selves.

Being a self is obviously the most important, or at least the most interesting, fact about each of us; and to understand ourselves we must understand that fact. What is it to be a self, and in particular, what is it to be *one*self: what distinguishes *me* from the rest of the world, including other selves, which I experience and with which I interact? Those are questions which concern all reflective people, not just philosophers: we all want to know and understand what makes us the selves we are. And answering those questions is a perennial project for all schools of philosophy. The project is perennial because our understanding of the self is inevitably affected by developments not only in philosophy but in society, in religion, in morality and in the sciences, natural, psychological and social. But it is the job of philosophy, above all, continually to assimilate and assess the significance of such developments for our conception of our selves.

On this occasion I can discuss only a small part of this perennial project. But it is an important part, and one that I think shows both how easy it is to raise a conceptual dust which effectively blocks our view, but also how such dust can be laid by a sprinkling of analysis. The part of the project which I have in mind is the part that deals with self-knowledge, by which I mean first-

[1]Versions of what follows have been discussed in lectures and seminars at the Universities of Cambridge, Edinburgh, Sheffield, Perth, Belgrade, Leuven and Nottingham, and at the 1989 Annual Conference of the Australasian Association of Philosophy in Canberra. I have been much helped in revising it by comments made on those occasions and also (in my discussion of the 'knowledge argument' in **3**) by an unpublished paper of Paul Teller's.

person knowledge, knowledge about oneself. This includes of course our first-hand knowledge of our own experiences, and in particular of what those experiences are like, a peculiar kind of knowledge that has itself attracted recent analytic attention (see e.g. Jackson 1986; Lewis 1989). But I want to discuss something more general, an aspect of all first-person knowledge, not only of our own experiences, but also of who we ourselves are, and of how the world is, or seems to be, from our own 'perspective' or 'point of view' – to use terms made fashionable recently by Thomas Nagel (e.g. 1983, 1986).

First-person knowledge is commonly and naturally called 'subjective' because, as we shall see, it is, in an important sense, relative to the subject: that is, to the person who has the knowledge. But it is also important to realise that, in another equally common and legitimate sense of 'subjective', this kind of knowledge needn't be subjective at all. This is the sense in which to be subjective is to be a mere matter of opinion, with any opinion (that is, belief) on the matter in hand as good as any other, because there's no fact of that matter to make some one belief true and contrary ones false.

Now in this sense, obviously, knowledge cannot be subjective: one can't know something unless there is something to know, some fact of the matter which does make some belief about it true and contrary ones false. And that is the sense, incompatible with knowledge, in which some philosophers hold, for example, that moral and aesthetic values are subjective. They hold, in other words, that no belief about such values (except this one!) is made true by any fact: because, as a matter of fact, the world contains no objective values, and hence no moral knowledge.

Well, maybe our moral and aesthetic beliefs are subjective in this sense, in which 'subjective knowledge' is a contradiction in terms. I don't think so, but I'm not going to press that point here. The point I am going to press is that most, if not all, of our non-evaluative first-person beliefs are clearly not subjective in this sense. Take my beliefs that I'm Hugh Mellor and that I live in Cambridge. Those are by no means mere matters of opinion, with one opinion as good as any other. For the fact is that I *am* Hugh Mellor, and I *do* live in Cambridge, and any contrary beliefs, say that I'm the Pope and live in Rome, would be as false as the belief that the Earth is flat. These beliefs of mine, about who I am and where I live, are as objectively true as any beliefs can be. Moreover, whatever it takes to make true beliefs knowledge, no one (other than professional philosophical sceptics) would deny that these beliefs of mine qualify. Not only am I Hugh Mellor, I know I am; and not only do I live in Cambridge, I know I do.

So first-person knowledge, knowledge about myself, isn't subjective in the sense of being mere opinion, with no facts to make it true. But it *is* subjective in the sense of being relative to the subject, to the person whose knowledge it is. It is, to use Nagel's terms, 'perspectival' knowledge: knowledge from the knowing subject's 'point of view'. But these expressions – 'perspective' and 'point of view' – are as misleading as the word 'subjective' is, and for the same reason: because they too suggest that knowledge of this kind is subjective in my other sense. For what the phrases 'from my perspective' and 'from my point of view' really mean is 'as I see it' or 'in my opinion'. And adding either of those phrases to an assertion either adds nothing – what, after all, is any sincere assertion but a giving of one's opinion? – or it appears to admit that contrary opinions could be just as good. But in examples like the ones I've just given, I admit no such thing. Who I am, and where I live, are not mere matters on which one opinion is as good as any other. They are matters of plain, observable fact, as knowable to me as the shape of the Earth, or the fact that 2+2=4.

What makes first-person knowledge peculiar, and problematic, is not, as the expressions I've just jibbed at suggest, that it's subjective in the sense that would prevent it being knowledge at all. The peculiar problem it presents is this. When I know that I live in Cambridge, I know something which *no one else can know*. I don't mean that no one else can know that Hugh Mellor lives in Cambridge: many people know that. Nor do I mean that, even if you knew that Hugh Mellor lived in Cambridge, you might not know that I do, because you might not know that I am Hugh Mellor. That possibility just exemplifies a well-known feature of knowledge – its non-extensionality – which is indeed a problematic feature, but not the one I want to discuss.

The problem I want to discuss shows up not in the fact that you *might* not know that I live in Cambridge, but in the fact that you *can't* know this, not in the first-person form in which I know it. In other words, you can't know the first-person fact that appears to be what makes this belief of mine, that I live in Cambridge, true. The reason you can't know this is that there's no first-person belief of yours which this first-person fact could make true, not even if you live in Cambridge too. For even if you do, the fact that *I* live in Cambridge isn't what makes *your* first-person belief, that *you* live there, true. The only belief of yours which the fact that I live there could make true is your belief that *I* live there. But for you that belief is a second-person or third-person belief, not a first-person one. So the fact which makes it true for you can't be a first-person fact, and *a fortiori* it can't be the first-person fact that I live in Cambridge. And if that fact doesn't make that belief of yours

true, it certainly can't make any other belief of yours true. But what this means is that, for you, the first-person fact that I live in Cambridge isn't a fact at all. The fact for you is the corresponding second-person or third-person fact: closely related to my first-person fact, but not the same.

In short, the first-person fact that I live in Cambridge is only a fact for me. For anyone else, it isn't a fact at all. And that's why you can't know it: because for you, there's no such fact to know. First-person knowledge seems therefore to be subjective in the sense not only that the *knowledge* is relative to the subject, but that the *facts* which are thereby known are relative to the subject. And that appears to present a serious ontological problem, whose solution could profoundly affect our conception of ourselves and of the world we live in.

The problem, I must emphasise, only arises because this first-person knowledge really is *knowledge*: that is, because it isn't just a matter of opinion. Moreover, whatever facts make my first-person beliefs about who I am and where I live true clearly don't depend on my knowing or believing them. They would still be facts even if I lost my memory, and had no idea of who I was or where I lived. And as in these examples, so in general. First-person facts are not generally reducible to first-person knowledge or belief: they are not subjective in that sense either. Whether I know it or believe it or not, it is a fact for me, if only for me, that I am Hugh Mellor and that I live in Cambridge. What makes my knowledge of these facts subjective, therefore, is not just that I have some kind of epistemically privileged access to them which other people lack. That may be true of Hugh Mellor's objective knowledge of Hugh Mellor's experiences. But my first-person knowledge, whether of my experiences or of who I am or where I live, is not just epistemically restricted to me. It isn't only *knowledge* of these facts that's confined to one person, it's the facts themselves.

And this is what poses the ontological problem. For if there really are subjective facts, which are not facts for everyone, but only for the subject, then since there are many different subjects, there are correspondingly many different worlds. My world contains the subjective facts that I am Hugh Mellor, and that I live in Cambridge. Your worlds contain neither of those facts, not even if it's a fact in your world that you live in Cambridge too. So it seems that we don't in fact live in a single objective world. Our worlds have an objective overlap, of course: they all contain the objective facts that Hugh Mellor lives in Cambridge, and that the Pope lives in Rome. But the Pope's world doesn't contain the subjective fact that I live in Cambridge, or that I'm

Hugh Mellor; any more than my world contains the subjective fact that I'm the Pope, or live in Rome.

The world of objective facts – the facts which, being facts for everyone, are common to all our subjective worlds – seems therefore to be only a small part of any of our worlds. And not the most important part either. For although my subjective facts include some very trivial ones (like how many eyelashes I have), they also include all the facts that are most important to me: who I am, whom I know and am related to, when I live, and where; how old I am; what my experiences, abilities, beliefs, desires, values and projects have been, are and will be. And all these fundamental facts about me lie completely beyond the reach of any objective science, whether it be physics or psychology: for objective science, by mere definition, concerns itself with facts that are facts for everyone, which these are not.

I shall call the account of first-person knowledge I've just given, which takes it to be knowledge of subjective first-person facts, 'subjectivist'. The subjectivist account presents a picture of the world – or rather of many worlds, one for each of us – about whose most important factual aspects, including who, where, when and what we ourselves are, no objective science can ever tell us anything at all. That is a very striking metaphysical thesis, and if true, a very serious limitation on the factual scope of science.

Not surprisingly, therefore, the subjectivist picture appeals most to those who are most resistant to the pretensions of science and, in particular, to its pretensions to encompass, if not all values, then at least all facts. While for the same reason, the picture appeals least to those who think the world just is the one objective world of physics and the other public sciences. Those who think that will tend therefore to dismiss these subjective facts just because they are subjective in the present sense. Russell (1940 p. 108), for example, claims that what are now called 'indexicals' – words like 'I', 'now', 'here' and 'this' – 'are not needed in any part of the description of the world, whether physical or psychological' (see also Quine 1954). In other words, there are no first-person facts: which seems to make my first-person beliefs subjective in the other sense of being mere matters of opinion, with any one opinion as factually good as any other. But that is incredible: as Nagel (1986 p. 57) says, 'if it is not a fact about the ... world that I am [Thomas Nagel], then something must be said about what else it is, for it seems not only true but ... one of the most fundamental things I can say about the world'. And even if we lack Nagel's sense of self-importance, we may still agree with him, against Russell, that 'it provides a clear example of the ineliminability of indexicals from a complete conception of the world'.

Whom then shall we follow? The subjectivist, who gives us our first-person facts and selves, but traps us in our own subjective worlds? Or the objectivist, who puts us all together in a single world but denies the objective truth of all our first-person beliefs? Anyone with a nose for nonsense will surely say: neither. There must be a more sensible way to go; and there is. But to see it we need a sprinkling of analysis to lay the conceptual dust that has so far obscured our view of it.

3 Subjective truths, objective facts The trick is to distinguish more carefully than I have so far done between truths and facts. So far I have simply assumed that true beliefs are made true by facts which correspond to them. This doesn't, I should say, commit me to the controversial correspondence conception of truth, because I am not using this concept of correspondence to *define* truth. If anything, I would use it to define facts (as whatever makes beliefs true), leaving truth to be defined in some other way, which I needn't discuss here. What matters here is not what, if anything, such a concept of correspondence can define, but simply that, if truths are to correspond to facts at all, they must do so in a more complicated way than I have so far implied.

Take Lewis Carroll's Messenger, who believes he passed nobody on the road, and suppose this belief of his is true. As I've already said, what makes it true is not the apparently corresponding fact that he passed some being, namely Nobody. There's no such fact, because there's no such being. There can't be, as we can see by supposing that it's also true that nobody passed the Messenger. But no being can both pass and be passed by the same person at the same time: that's nonsense, which is the whole point of Lewis Carroll's joke.

This particular nonsense, as we've seen, has a very simple explanation. The fact which really makes the Messenger's belief true is the fact that there was *no* being that he passed: a fact which is perfectly consistent with there also being no being who passed him. End of analysis: dust laid. And now we can see quite clearly how beliefs which are apparently about Nobody are really made true by facts which incorporate no such impossible being.

It is perhaps less obvious how first-person beliefs can be made true by facts that incorporate no first-person self, no *I*. But they can. All first-person beliefs can be made true by purely objective facts. Take my belief that I live in Cambridge. That obviously corresponds to, and can therefore be made true by, the fact that Hugh Mellor lives in Cambridge, because I *am* Hugh Mellor. And this true belief (that I'm Hugh Mellor) can in turn be made true by the

fact that it's held by Hugh Mellor, a fact which likewise contains no first-person constituent.

Now if all true first-person beliefs can be made true in this way by objective facts, we can escape the dilemma I presented earlier. We can have real first-person knowledge in a single, wholly objective world. And if there are facts about that world which our public sciences can't discover, it will only be because they're practically or physically impossible to discover, not because they're first-person and therefore not facts for everyone.

But can this be so? Can all true first-person beliefs really be made true by objective facts? To show that they can, we must first produce enough objective facts to correspond to all our true first-person beliefs. Now that, oddly enough, is both quite easy to do and not especially controversial. No one denies that the truth of my belief that I live in Cambridge does at least correspond to the fact that this belief is held by someone who lives in Cambridge. And as for me, so for everyone. Any belief of the form 'I live in town *T*' will be true if and only if, as a matter of fact, it is held by someone who lives in town *T*.

No one disputes this (see e.g. Nagel 1983 p. 216). Nor does anyone dispute the objectivity of these corresponding facts. Everyone accepts that it's a fact for everyone, and not just for me, that my beliefs are held by Hugh Mellor, who lives in Cambridge. And similarly for all first-person beliefs – including, incidentally, beliefs about one's own experiences. The truth of any belief whose content is 'I'm in pain' will always correspond to the objective fact that its owner is, at that very time, in pain. None of this is seriously denied.

But why then would anyone deny that these objective facts are what make our first-person beliefs true? Why should anyone continue to believe in such mysterious first-person facts as *my* being Hugh Mellor and *my* living in Cambridge when they are as clearly redundant as they are ontologically problematic?

The main reason is the following so-called 'knowledge argument'. Take my knowledge that I live in Cambridge. I say that the belief which constitutes this knowledge is made true by the objective fact that the believer, Hugh Mellor, lives in Cambridge. But that isn't what I know by knowing that *I* live in Cambridge. For suppose I forget who I am, or go mad and believe that I'm the Pope and that Hugh Mellor – whoever he may be – doesn't live in Cambridge at all. I can still know that *I* do. But that means that I can know this without knowing, or even believing, the very fact – that Hugh Mellor lives in Cambridge – which I say is what makes my belief about where I live true.

And even if I do know that Hugh Mellor lives in Cambridge, that will have nothing to do with my knowing that I do if I don't believe I'm him.

Knowing that I live in Cambridge cannot therefore be equated with knowing the objective facts which I say make that belief of mine true, since I can easily know the one without knowing the other. And as for this piece of first-person knowledge, so, obviously, for all others. But then first-person knowledge looks after all as if it must be what it seems to be, namely knowledge of subjective first-person facts, which must therefore exist: since one can't very well know what isn't the case.

What's wrong with this argument? As an argument for a difference between subjective and objective beliefs, nothing. My first-person belief that I live in Cambridge must indeed differ from my objective belief that Hugh Mellor does so. But this, as we shall now see, doesn't mean that my first-person belief must be made true by a first-person fact.

To show why not, I must first draw another distinction: between belief *types* and *tokens* of those types. A belief type is defined by a belief's *content* (what is believed), regardless of who believes it or when. A token of a belief type is a particular person holding a belief with that content at a particular time. (Take for example the belief that we're human: one belief type with, at any instant, millions of tokens.)

Now what the knowledge argument I've just given shows is that my belief that I live in Cambridge differs in content from my belief that Hugh Mellor does so, even though I am Hugh Mellor. How is this supposed to show that my first-person belief must be made true by a first-person fact? Well, suppose we accept the idea that the content of a belief type is its 'truth condition': that is, the condition in which it's true. (This idea is controversial in various ways, but none of them affects the present argument.) For example, the content of our belief that we're human is obviously the condition in which that belief is true, namely that we are indeed human. But then mustn't the first-person content of my belief that I live in Cambridge also be its truth condition? In other words, mustn't the first-person fact that I live in Cambridge be what makes this first-person belief true?

Not at all. All that follows from my assumptions is that the truth conditions of first-person belief types must differ from those of objective belief types. But this doesn't mean that the former can't be objective: they can be; and I say they are. The only difference between the two kinds of truth conditions is that the first varies from person to person and the second doesn't. For example, the objective truth condition of the objective belief that Hugh Mellor lives in Cambridge, namely that *Hugh Mellor* lives in Cambridge, is the same for

everyone. Whereas the objective truth condition of anyone's token of the first-person belief 'I live in Cambridge' is that *that person* lives in Cambridge.

In other words, we can identify the content of the belief type 'I live in Cambridge' with a *function* (in the mathematical sense) from any person who has this belief to the objective truth condition of his or her token of it: namely, that this person lives in Cambridge (see Kaplan 1979). And this is a completely objective function: from objective people, like Hugh Mellor or the Pope, to objective truth conditions, like Hugh Mellor living in Cambridge or the Pope doing so. It requires the existence of neither first-person selves nor first-person facts. In particular, it lets my belief that I live in Cambridge be made true by the objective fact that Hugh Mellor lives there. But because it doesn't identify the *content* of my belief with that fact, it needn't identify the belief itself with the belief that Hugh Mellor lives in Cambridge.

In short, it simply doesn't follow, from the evidently different contents of subjective and objective beliefs, that the subjective ones are made true by subjective rather than by objective facts. As an argument for first-person facts and selves, the knowledge argument is a simple *non sequitur*.

Moreover, this objective account of how first-person beliefs differ from objective ones shows how I may not believe that I'm Hugh Mellor even though I *am* Hugh Mellor and even though I obviously do believe that I'm me and that Hugh Mellor is Hugh Mellor. The reason of course is that these three belief types have different contents: that is, different functions from any person *P* who has those beliefs to the truth conditions of his or her tokens of them. The content of 'I'm me' is the function from *P* to *P*'s being *P*, and that of 'Hugh Mellor is Hugh Mellor' is the constant function from *P* to Hugh Mellor's being Hugh Mellor, neither of which anyone can doubt. That doesn't prevent anyone, even Hugh Mellor, doubting the content of 'I'm Hugh Mellor', namely the function from *P* to the truth condition that *P* is Hugh Mellor – a function which incidentally shows at once why only I, Hugh Mellor, can *know* that I'm Hugh Mellor, since only Hugh Mellor's tokens of this belief type can satisfy its truth condition.

This account thus immediately explains, and explains away, the idea of subjective first-person facts. They have been invalidly inferred from the way in which the truth of first-person beliefs depends on who holds them; just as Lewis Carroll's Nobody is invalidly inferred from the way in which the truth of the Messenger's belief depends on his not having passed anyone. Equally importantly, this account also explains, as we've just seen, why my *knowing* who I am is not an automatic and therefore trivial consequence of my *being* who I am. And finally, it shows clearly how knowing who I am differs from

knowing any objective facts, even if those are all the facts there are. It therefore shows how, even in a single objective world, knowledge of one's own identity can be a real and irreducibly extra piece of knowledge, which may matter to people as much as it does to Nagel or as little as it does to me. And this in turn enables this account to reconcile Russell's claim that we need no indexicals to describe the *world* with Nagel's claim that we do need them to express our first-person *conception* of it.

In short, this objective account of the content of first-person beliefs lays all the conceptual dust which subjectivism raises about the nature of first-person knowledge in general, and knowledge of one's own identity in particular. And dissipating the apparent mystery about the first-person self in this way enables us to see more clearly the real problems which our concept of the self presents: namely, how do selves differ from sticks and stones and computers and lower forms of life? In other words, what does it take to be *a* self, capable of agency, consciousness, knowledge – including first-person knowledge – and understanding? That is where the real mysteries lie, which we will understand much better and faster once we distinguish them from the bogus mystery of what it takes to be *this* self. Once we know what *a* self is, there need be no more mystery about that than there is about what it takes to be *this* room when we know what *a* room is.

4 Subjective selves This however is not quite the end of the matter. So far I have shown only that first-person knowledge *needn't* be knowledge of first-person facts, not that it *can't* be. To show that, we must look more closely at what follows if it is: namely, that there are such facts. So let's suppose, for the sake of the argument, that there are first-person facts, and that my living in Cambridge is one of them, and hence different from the fact that Hugh Mellor lives there.

How can these two facts differ? Well, obviously, they can differ only if I differ from Hugh Mellor: that is, only if I, my first-person self, and Hugh Mellor are distinct beings. And so perhaps we should be, from a subjectivist's point of view. For one thing, I, as a constituent of all my first-person facts, exist only in my own subjective world; whereas Hugh Mellor, as a constituent of many objective facts, exists in everyone's world. How then can he and I be identical? And if we aren't identical, then we do have (both of us!) a very simple explanation of how I can know first-person facts about myself without knowing the corresponding objective facts about Hugh Mellor: namely, that they are quite different facts, about quite different beings. And that includes, not just external facts about where I live, but internal facts about my experi-

ences, beliefs, desires, hopes, fears, emotions – in short, all the first-person facts which constitute my inner life.

All this makes me, my first-person self, as opposed to Hugh Mellor, look rather like the immaterial ego of Descartes: a 'thinking thing' defined by its introspectible attributes, and distinct from the spatially extended body which it inhabits, and whose demise it could conceivably survive. In fact there are important differences, both between my first-person self and a Cartesian ego, and between Hugh Mellor and my body. For one thing my first-person self, unlike my Cartesian ego, is just as embodied and located in space as Hugh Mellor is; and for another, Hugh Mellor, unlike my Cartesian body, thinks just as much as I do. But as these differences, though striking, don't really matter for present purposes, I propose, for convenience, to call the subjectivism which denies my identity with Hugh Mellor 'Cartesian'.

Cartesian subjectivism, like Descartes himself, can conceive of me, my first-person self, surviving the death of the objective Hugh Mellor, who lives in Cambridge, and to whose body I am now, for better or worse, so literally attached. That is, I now get all his (and only his) hangovers and other bodily pains, and pleasures; I see only what falls within his field of view; I hear only what strikes his ear; I touch only what he reaches; and so on. But when Hugh Mellor dies ... ah, then my life could be something else again, whether reincarnated in this terrestrial world, or freed of all bodily constraints in some celestial one.

Cartesian subjectivism can appeal to all these more or less enticing Cartesian possibilities, but only at the cost of inheriting all the many well-known problems that face Descartes' conception of the self. How for example does my first-person self interact causally with Hugh Mellor's objective body, as it clearly does, since what hits him causes me to get hurt, and what I decide to do causes him to do it? Or again, what constitutes my self's identity through time, if not the causal continuity of Hugh Mellor's body? And so on. Those are indeed serious problems, and consequently serious objections to Cartesian subjectivism. But they aren't the decisive ones. It's not the causal links between Hugh Mellor and me that really reveal the nonsense in Cartesian subjectivism: it's the logical links.

The first thing to notice is that subjectivists can only take this Cartesian way out by denying the very first-person fact that gives subjectivism its appeal in the first place: namely the fact that I am Hugh Mellor. Now I too, of course, deny that this identity is a *fact*. I say there are no such first-person facts of identity: there is just Hugh Mellor. But I don't thereby deny the truth of my belief that I'm Hugh Mellor. On the contrary, as we've seen, I can admit and

account quite easily for the truth of this belief, and give it a content which only I can know, without appealing to any first-person facts at all. Whereas a Cartesian subjectivist who denies the fact that I'm Hugh Mellor must also deny the truth of my belief that I'm Hugh Mellor: a denial which I must say I find completely incredible.

Next, by making me a being distinct from Hugh Mellor, the Cartesian subjectivist reduces to causal links what are obviously logical ones. It is, for example, quite obvious that by living in Cambridge, Hugh Mellor doesn't just *cause* me to live there too. His doing so *entails* my doing so: because I *am* him. Similarly, when he takes the train to London, he doesn't just drag me along causally, a first-person free rider on his objective ticket, two passengers for the price of one. I *have* to go with him, logically have to, because I am him. British Rail are quite right to issue only one ticket; and when we (Hugh Mellor and I) travel on it, we aren't swindling them at all, because there really is only one of us. At least that is what I say, and I'm an objectivist. And if I can say it – and say it truly – then a subjectivist had better be able to say it truly too.

In short, Cartesian subjectivism makes no sense. If there are first-person facts, then the fact that I'm Hugh Mellor must be one of them. But admitting that fact just impales our subjectivists on the other horn of a dilemma. For if I really am Hugh Mellor, then my living in Cambridge, travelling to London, having various experiences, and so on, cannot be different facts from Hugh Mellor's doing and experiencing those very same things. But then since those facts about him are all objective, the same facts about me must also be objective. In other words, our first-person beliefs can no more be about first-person selves, constituents of whole arrays of first-person facts, than the Messenger's beliefs can be about Nobody. So not only *can* first-person beliefs be made true by objective facts, in the way I've outlined, they *must* be. Our first-person beliefs can only refer to the objective people, the ordinary selves, whose beliefs they are, and who are constituents of nothing but objective facts. (For a detailed account of how people's first-person beliefs refer to them, see chapter 2.)

I conclude therefore that the subjective mystery of the first-person self not only can be dissolved, it must be. Like Lewis Carroll's Nobody, the idea of the first-person self is nonsense; and so, as Ramsey said, we must take seriously that it is nonsense, and not pretend, as Nagel does, that it is important nonsense. Our being subjects does not make us subjective beings, trapped in our own subjective worlds. No such worlds exist. There is just one objective world, a world of which each of us, every self, every subject, is equally – and wholly – a part.

To reject the first-person self in this way is, however, as I have tried to show, by no means to deny that we all have our own different first-person knowledge of ourselves, and of how things are, or seem to be, from our own point of view. Nor is it to deny that we have this knowledge by having true first-person beliefs: beliefs whose contents are quite different from those of any objective beliefs. For there is, as we've seen, no objective mystery about these differences: they follow simply and solely from the way in which the truth of first-person beliefs depends on whose beliefs they are, which the truth of objective beliefs does not.

That is why no first-person facts or selves are needed to enable us all to have our very different but equally true beliefs about who we are, where we live, and what we want, experience and are doing. This manifest diversity in our first-person knowledge of ourselves and of other things shows neither that we all live in different worlds, nor that our supposed knowledge is really mere opinion, corresponding to no objective facts. That is a completely false dichotomy. The differences in our true first-person beliefs are simply consequences of our being objectively different people, with different objective histories and experiences, and differing in our objective relations to the rest of the one objective world we all inhabit. Those objective differences are quite enough to enable all our different first-person beliefs to be equally true, thus enabling them to constitute the very different first-person knowledge which we all know that each of us has. End of analysis: dust laid.

2 *I and now*

1 Introduction Many philosophers overrate the present subject. *Pace* Nagel (1983), there are no subjective facts or selves; nor, *pace* many others, 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. That's the message of this paper.

I start with some underrated truisms about the token-reflexive truth conditions of 'subjective' (by which I mean first person present tense) sentences. I then ask how 'I' and 'now' refer to their users and to when they're used. To answer that question, I turn to the beliefs we use these words to express: what it takes to have them, to have them consciously and to put them into words. None of this involves subjective facts or selves, internal representations of oneself or the present, or any problematic capacity for, or concept of, self-reference or self-knowledge.

2 Subjective truths, objective facts 'I' and 'now' are often called 'indexicals'. I shall call them 'token-reflexives', which describes their relevant feature: the way they make the truth of sentence tokens depend on facts about those tokens.

Suppose some man, *K*, faces food at some time, *T*; and take the sentence type

S(*KT*) *K* faces food at *T*,

tenselessly (i.e. as implying nothing about whether *T* is past, present or future)[1] unlike the semantically present tense

S(*KN*) *K* faces food now,

which implies that *K*'s facing food is temporally present.

S(*KT*) is not token-reflexive: the truth of *S*(*KT*)'s tokens – *s*(*KT*)s – does not depend on when they occur. But that of *S*(*KN*)'s tokens does: an *s*(*KN*) is

[1]So *T* must be a date, like 1 January 1988, not a tense like *now*; *K* similarly is someone like Napoleon, not someone like *I*; and '*K*' and '*T*' mustn't themselves be token-reflexive.

true if and only if it occurs while *K* faces food. *S*(*KN*) is temporally token-reflexive.

Now consider the tenseless but first person sentence

S(*IT*) I face food at *T*,

which says of whoever I am that I face food at *T*. So a token, *s*(*IT*), of *S*(*IT*) is true if and only if produced by someone who faces food at *T*: its truth depends on who produces it. *S*(*IT*) is personally token-reflexive.[2]

Finally, the subjective sentence

S(*IN*) I face food now,

which implies that I (whoever I am) face food now (whenever that is), is both temporally and personally token-reflexive: its tokens, *s*(*IN*)s, are true if and only if produced by people who are simultaneously facing food.

What then are the truth conditions of *s*(*KN*)s, *s*(*IT*)s and *s*(*IN*)s? We know their truth depends on when or by whom they're produced. So either (1) their truth conditions vary from time to time or person to person or (2) whether those conditions obtain varies.

The temporal alternatives are well known. (1) is the 'tenseless' view, on which any *s*(*KN*) is true if and only if *K* faces food at whatever time *Y* it occurs: a condition that varies with *Y*. (2) is the 'tensed' view, on which an *s*(*KN*)'s truth condition is always that *K* faces food *now*. Only this condition doesn't always obtain: sometimes *K* faces food now, sometimes he doesn't.

Which view is right? The tenseless view: because, as McTaggart (1908) showed long ago, there is really no such time as *now*. There can't be. If there were, then since all sentence tokens of all the types 'The time is now *Y*' are true if and only if they occur at *Y*, every time *Y* would have to be both *now* (to make the true tokens true) and not *now* (to make the false ones false), which it can't be (see McTaggart 1927 ch. 33; Mellor 1981 ch. 6). Reality can no more include a time (*now*) or a property (*being now*) satisfying this impossible condition than it can include a barber who shaves all and only those who don't shave themselves.

In short, though there are many tensed *truths*, there are no tensed *facts*: where by a 'fact' I mean a truth condition that obtains. Just as 'Nobody shaves all and only those who don't shave themselves' can't be made true by a fact involving a real *Nobody*, so no *s*(*KN*) can be made true by a fact involving a

[2]Not all personal (i.e. personally token-reflexive) sentences are first person, of course (e.g. 'My mother faces food', 'You face food', 'He faces food'), just as not all tensed sentences are present tense. But the first person ones are all that need concern us here.

real *now*. All *s*(*KN*)s, like all other tensed sentence tokens, have purely tenseless truth conditions.

But what then do tensed sentence types and tokens mean?[3] Their meanings can't be identified with their truth conditions, since those vary from time to time, which their meanings don't. And *S*(*KN*) never means what *S*(*KT*) means, not even at *T*, when *S*(*KN*) and *S*(*KT*) have the same truth condition: that *K* faces food at *T*. So what does *S*(*KN*) mean?

I say with Kaplan (1979) that *S*(*KN*)'s meaning is a semantic function, $f_{KN}(Y)$, from the time *Y* when any *s*(*KN*) occurs to its truth condition: *K* facing food at *Y*. Of course $f_{KN}(Y)$'s values vary with its argument *Y*, but $f_{KN}(Y)$ doesn't: nor therefore does *S*(*KN*)'s meaning. And its meaning always differs from *S*(*KT*)'s, even when *Y*=*T*. For *S*(*KT*) means a constant function, f_{KT} (= *K* facing food at *T*), which *S*(*KN*) doesn't. $f_{KN}(Y)$ and f_{KT} are never the same function, even when *Y*=*T* and they have the same value: so even then, *S*(*KN*) and *S*(*KT*) differ in meaning, as they should.

As for times and *S*(*KN*), so for people and *S*(*IT*). (1) is what I shall call the 'impersonal' analogue of the tenseless view, which says that an *s*(*IT*) is true if and only if produced by someone, *X*, who faces food at *T*: a condition that varies with *X*. Whereas on (2), the 'personal' view, every *s*(*IT*)'s truth condition is the same, whoever produces it: namely, that (as its producer would say) *I* face food at *T*. Only this condition doesn't obtain for every *I*: some *I*s face food at *T*, some don't.

Again (1) is right. The personal view can be disproved by an analogue of the argument against the tensed view. There can't be personal truth conditions. If there were, then since all sentence tokens of the form 'I am *X*' are true if and only if produced by *X*, every person *X* would have to be both *I* and not *I*, which no one can be. Reality can no more contain a person (*I*) or a property (*being me*) satisfying such a condition than it can contain a time *now* or a property of *being now*. All *s*(*IT*)s, like all other personal sentence tokens, have purely impersonal truth conditions. Though there are many personal truths, there are no personal facts.

But what then do *S*(*IT*) and its tokens mean? Again, they can't mean their truth conditions, because those vary from person to person, which their meaning doesn't. And no *s*(*IT*) ever means what an *s*(*KT*) means, even when they're both produced by *K* and have the same truth condition: *K* facing food at *T*. So what does *S*(*IT*) mean?

---

[3]A sentence type has a meaning only if all its tokens have the same meaning. So I ascribe meanings indifferently to both.

It means another semantic function, $f_{IT}(X)$, from whomever, X, produces a token $s(IT)$ to its impersonal truth condition: X facing food at T. For again, although $f_{IT}(X)$'s value varies with its argument X, $f_{IT}(X)$ doesn't: nor therefore does $S(IT)$'s meaning. Nor does $S(IT)$ ever mean what $S(KT)$ means: having the same value for X=K doesn't make $f_{IT}(X)$ the same function as f_{KT}.

And as for tensed and personal sentences, so *a fortiori* for subjective ones. They can't be made true by subjective facts, which would have to be both tensed and personal and can't be either. So no $s(IN)$ is ever made true for me, whoever I am, by the subjective fact that I face food *now*. There's no such fact, because there's no such person as *I* and no such time as *now*. An $s(IN)$ can only be made true by the objective fact that whoever (X) produces it, at whatever time (Y), simultaneously faces food: a semantic function, $f_{IN}(X,Y)$, which is indeed obviously just what $S(IN)$ always means.

3 The referents of 'I' and 'now' So much for meaning. What about reference? The terms 'K' and 'T' refer to the man K and the time T. That is, they make any $s(KT)$ say that it's K who faces food, and faces it at T. They make $s(KT)$'s truth depend on a fact about that man and that time.

What do 'I' and 'now' refer to? What person and what time do they make part of an $s(IN)$'s truth condition? Not the person *I* or the time *now*, since (as we've seen) no such person or time exists. The type words 'I' and 'now' have no one referent. But their tokens do. They refer to those who produce them, and to when they're produced: because they make each $s(IN)$ say about its producer, and about when it's produced, what every $s(KT)$ says about K and T: namely, that that person faces food at that time.

How are these references made? There's a familiar answer for terms like 'K' and 'T': their referents are functions of known 'senses', semantic properties given (say) by definite descriptions. 'K' and 'T' refer to whatever man (K) and time (T) satisfy the descriptions that we would use to give the senses of those terms.

I think this answer is inadequate even for 'K' and 'T', and it's hopeless for 'I' and 'now'. For since their referents vary from user to user and time to time, so must their senses. But how then do we know what they are? Not by knowing definite descriptions, since I could forget almost everything about myself and the present time and still know that *I* had done so *now*. Anyone, X, at any time, Y, can use 'I' to refer to X and 'now' to refer to Y, without knowing descriptive senses for these words remotely sufficient to distinguish X and Y from all other people and times.

But perhaps tokens of 'I' and 'now' have senses that are ineffable, like tastes and smells? But that doesn't explain how we know them, because unlike tastes and smells, senses aren't self-intimating. I may have no sense at all of the time: my 'now' will still refer to it; nor of who I am: my 'I' will still refer to Hugh Mellor. How can these references be fixed by senses available only now, and to me, but of which even I need now know nothing?

And not even self-intimating senses would make these references as infallible as they are. The 'now' in a token *s(IN)* can't fail to refer to when it's produced; nor can my 'I' tokens fail to refer to Hugh Mellor. Self-intimation isn't that infallible. Even I can mistake the taste of Glenfiddich for that of Glenmorangie, and misapply their names accordingly. I can't do that with 'I' and 'now'.

These and related problems of self-reference have no generally accepted solution. They have even made Anscombe (1975) and others (see Diamond and Teichman 1979) deny that 'I' refers at all. But 'I' and 'now' do refer – or rather their tokens do, since they make the truth conditions of token sentences include their producers and when they're produced. But to see how, we must turn away from sentences like *S(IN)* to the beliefs those sentences express.

4 Belief, truth and utility I can believe that I face food now without saying so, and even without being able to say so: animals need no public language to believe *IN*. (They may need a 'language of thought' (Fodor 1975), but that's another matter.) Knowing when to eat comes before knowing how to talk: language use is an advanced activity, not a basic one. Agents needn't be able to say what they're doing, let alone what beliefs and desires make them do it. Those beliefs and desires needn't even be conscious: my belief that what I'm facing now is food can make me eat it absent-mindedly while my conscious mind is miles away. Agency entails neither linguistic ability nor conscious belief.

In particular, it doesn't entail self-conscious belief. So we mustn't infer from our own self-conscious linguistic use of 'I' that animals need a self-conscious or linguistically expressible 'concept of the self' to believe *IN*. They don't. Having a concept of the self means having a conception of it: an idea of what distinguishes selves from stones, and oneself from other selves. But that means having beliefs of the form 'I am *F*', where *F* is something that selves are and stones aren't. So we need some subjective beliefs *before* we can have a concept of the self. And even they won't be the first subjective beliefs we acquire. For first, simply to survive, we must be able to believe truly that we now face food, or predators, or a mate. Those are the first subjective beliefs

our animal ancestors must have evolved the ability to acquire when facing food, predators, etc. Concepts of the self (and of the present) are a much later luxury.

But how can the above semantics for consciously used sentences like *S(IN)* be applied to beliefs that need be neither conscious nor linguistically expressed? First, we apply the type–token distinction thus: a token, *b(P)*, of a belief type *B(P)* is a fact of the form: *X* believes *P* at time *Y*.[4] Next, the content *P* of any such belief is the meaning of a sentence, *S(P)*, that would express it: i.e. a semantic function (constant or otherwise) from *X* and *Y* to the truth conditions which a token *S(P) would* have if *X* produced one at *Y*. Thus for *P=IN*, the content of *X*'s belief *B(IN)* at any time *Y* is the meaning of *S(IN)*, namely the function $f_{IN}(X,Y)$ from *X* and *Y* to the truth condition that *X* faces food at *Y*.

But beliefs are distinguished one from another not only by their contents but by how they affect our behaviour. I may be able to believe *IN* without doing so consciously, or saying so. But not without being disposed to act in various ways: for example, to eat what I face when I'm hungry. Of course different desires will make this belief make me act differently. It will make me act one way when I want to eat food, and in other ways when I want to cook it, freeze it or throw it away. And a different belief about where the food is would make the same range of desires make me do a quite different range of things.

In short, believing *IN* makes me embody a causal function from desires to actions, just as a desire for food makes me embody a causal function from beliefs to actions. In other words, beliefs and desires affect each other's causal powers: believing *IN* affects how hunger makes me act; hunger affects how believing *IN* makes me act.

The beliefs and desires that have these effects are of course tokens, not types: e.g. *K* believing *IN*, and wanting food, at *T*. Now these tokens, I have said, are facts, corresponding to the truth of sentences: '*K* believes *IN* at *T*', '*K* is hungry at *T*'. But Davidson (1967a) has notoriously argued that causes and effects must be events, not facts: particulars, corresponding to singular terms, not to sentences. But he is wrong, as I and others have argued (see chapter 12). Most causation connects facts, and is rightly reported by a connective: e.g. (*C*) '*K* eats what he faces at *T* because *K* gets hungry at *T*'.[5]

---

[4]By unqualified 'belief' hereafter I may mean either a type or a token (or both) according to context; similarly for 'desire', etc.

[5](i) *Pace* Davidson, although some causal contexts are opaque ('*K* is the *F* because *L* is the *G*'), (*C*) needn't be: (*C*) and '*K* is the *F*' can entail 'the *F* eats ...' and '... the *F* gets hungry at *T*'. Ditto for '*T* is the *U*' (see chapter 12). (ii) Causes and effects can't really be simultaneous

So this piece of causation is also a fact, corresponding to (*C*)'s truth, which can therefore also be caused, e.g. by the fact that, just before *T*, *K* comes to believe *IN*.

But how is *B*(*IN*)'s causal function from desires to actions linked to its content, i.e. to its semantic function $f_{IN}(X,Y)$? Well, consider Ramsey's (1927 p. 46) chicken, which believes 'a certain sort of caterpillar to be poisonous', a belief that Ramsey equates with the chicken's 'abstain[ing] from eating such caterpillars on account of unpleasant experiences connected with them'. These actions, Ramsey says, are 'such as to be useful if, and only if, the caterpillars were actually poisonous'. 'Thus', he continues, 'any set of actions for whose utility *P* is a necessary and sufficient condition might be called a belief that *P*, and so would be true if *P*, i.e. if they are useful.'

In other words, true beliefs make desires cause actions that succeed in achieving the desired end. And although a successful action can be caused by a false belief (as when I sweeten my coffee with saccharine, believing it's sugar), that's a fluke. Only truth will ensure success in every action that a belief would combine with some desire to cause.

Often of course beliefs must conjoin to make a desire cause an action, whose success will then depend on the truth of the conjunction. I won't go to a shop which I believe has something I want unless I also believe it's open, and even then I won't get what I want unless both beliefs are true. Neither truth is enough on its own. So 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.[6] Then its truth conditions are what I shall call its 'utility conditions': those in which all such actions would achieve the desired end.

Ramsey's pragmatist equation, of a belief's truth conditions with its utility conditions, is all we need. With it we can quickly show how simply subjective beliefs like *B*(*IN*) get their personally and temporally token-reflexive truth

---

(Mellor 1981 ch. 9): *K*'s hunger at *T* can't make him start to eat until some time *T'* after *T*. But since the denseness of time lets *T'*–*T* be less than any finite amount, the difference is immaterial.

[6]It doesn't matter that this restriction is rarely satisfied; nor that having any one belief *B*(*P*) entails having many others: *b*(*P*) may still be the only token that makes a given desire cause a given action. (Compare the many ineffective entailments of 'there was a short-circuit' in 'There was a fire because there was a short-circuit'.) For a full elaboration and defence of Ramsey's idea, see Whyte (1990, 1991).

conditions: i.e. how simply a token *b(IN)*, *X* believing *IN* at *Y*, refers to *X* and *Y*.[7]

5 How subjective beliefs refer Take any man *X* at any time *Y*, suppose him hungry then, and suppose some token belief *b(P)* makes this hunger of his cause him to eat what he faces. What belief must this be: i.e. what is *b(P)'s* truth condition? Ramsey's equation tells us: its utility condition, that in which the action it makes *X*'s hunger cause succeeds. But that action, *X* eating what he faces, will succeed (by assuaging *X*'s hunger) if and only if what *X* faces is food. That is *b(P)*'s utility and hence its truth condition.

But what token of *B(P)* is this: i.e. who is believing *P*, and when? *X*, of course, at *Y*. Beliefs don't affect the causal powers of desires at a distance. Only *my* beliefs affect how *my* desires make me act, and then only while I have them. So for *b(P)* to make *X*'s hunger at *Y* cause an action, it must be a belief that *X* himself has at *Y*. And since desires likewise don't cause actions at a distance, that action (the eating) must also be *X*'s and must start at *Y*:[8] an action which will therefore succeed in assuaging the hunger that caused it if and only if *X* faces food at *Y*.

So *b(P)*'s utility and hence truth condition is that whatever agent has this belief faces food when he, she or it has it. But that's the truth condition of a token of the subjective belief *B(IN)*. So that's what *b(P)* is. In other words, the belief agents need, to make their hunger cause them to eat what they face, is *B(IN)*. And that, in general, is why we need subjective beliefs: only they will make our basic desires (for food etc.) cause us to act to satisfy those desires – actions on whose success our lives frequently depend.

And the mechanism that links such a belief to the agent and the time it refers to is simplicity itself: the contiguity of cause and effect. A belief can be made subjective, by how it affects the effects of an agent's desires, only if the same agent has it, and at the same time. That's what gives it its personally and temporally token-reflexive truth conditions. Causal contiguity is how a subjective belief refers to whomever has it, and when.

This at once explains the otherwise puzzling features of subjective reference noted in **3**. For in order to believe *IN*, I need only be disposed

---

[7]Said of a sentence or belief, 'refers to *X*' means 'has *X* in its truth condition'; said of a referring term or its mental analogue, it – like 'represents *X*' – means 'makes *X* part of the truth condition of the sentence or belief containing it'.

[8]The denseness of causation prevents it ever being really immediate, but for present purposes we may ignore the bodily causal intermediaries between desires and actions, and the consequent delays caused by our size and the finite speed of causal transmission (see Dennett 1978).

(*inter alia*) to eat what I face if I feel hungry: a disposition which causal contiguity ensures that only *my* simultaneous hunger can provoke, and only into making *me* eat, and only *then*. That's what makes my belief refer to me and to when I have it. And that's why I need no idea who I am or what the time is, no concept of the self or of the present, no explicit or implicit grasp of any 'sense' of 'I' or 'now', to fix the referents of my subjective beliefs: causal contiguity fixes them for me.

And does so infallibly. My apparent belief in *KT* could always fail to refer to *K*: *K* might not exist, or might not be the man I took him for – my belief might be about someone else entirely. But subjective beliefs can't fail to refer in either of those ways. For no such belief can exist unless its referents do: *X* can't believe *IN* at *Y* unless *X* and *Y* exist. And causal contiguity means that this *b*(*IN*) can only make *X*'s desires at *Y* cause actions, and then only actions done by *X*, and at *Y*. The actions may still fail, if *X* doesn't in fact face food at *Y*: *b*(*IN*)'s utility and hence its truth condition may still not obtain. But the actions can't fail by being done by the wrong agent or at the wrong time. Causal contiguity sees to that: and thereby makes every subjective belief refer infallibly to whomever has it, and when.

6 Self-reference That's how causal contiguity enables agents without conscious or linguistic powers of self-knowledge or self-reference, or concepts of the self or of the present, to refer infallibly to themselves and to when they do so. But that's not all causal contiguity does for subjective reference. It also enables agents and times to be represented by themselves in the subjective beliefs that refer to them.

To see how, consider how *K* is referred to in beliefs like *B*(*KT*) and *B*(*KN*). What makes *K* part of their truth conditions? Suppose for example I want to write to *K*. This makes me write on the envelope what I believe is *K*'s address (*A*). That action will succeed if and only if *K* really does live at *A* (*KA*): which is what makes his living there the utility and hence the truth condition of my belief *b*(*KA*).

But how can what I write be affected by *K* living, miles away, at *A*? It can't, directly: there's no such action at a distance, and anyway *K* may not live there – my belief may be false. So what makes my desire to write to *K* make me write '*A*' on the envelope must be some causal surrogate for *K* living at *A*. I must, when I write (at time *Y*), have some intrinsic property V_1 – chemical, electrical or whatever – which makes my relevant desires make me act in ways that will succeed if *K* lives at *A*. This token property v_1 (the fact that I am V_1

at *Y*) embodies the belief *B*(*KA*) in me, much as '*KA*' embodies its content here.

v_1 refers to *K* as '*KA*' does: *K* is in its truth condition. But v_1 also has a causal structure, part of which refers to *K* as '*K*' does. For suppose that at *Y* I get a new token belief about *K*, say *b*(*KN*), embodied in another token intrinsic property v_2. My belief *b*(*KA*) will naturally make this give me the further belief, *b*(*AN*), that the man who lives at *A* now faces food. But for my *b*(*KA*) and *b*(*KN*) to interact like this because they share the referent *K*, v_1 and v_2 must share an intrinsic causal component, v_K (like the '*K*' in '*KA*' and '*KN*': see Fodor and Pylyshyn 1988). And that makes v_K, whatever it is, a causal surrogate for *K* in these beliefs: an internal representation of *K*, which refers to *K* just as tokens of '*K*' do.

This does not of course reduce mental representation to intrinsic properties like v_K: because intrinsic properties don't refer intrinsically to anything. v_K doesn't refer to *K* intrinsically: it refers to *K* only by linking certain beliefs of mine, which refer to *K* because of what they make my desires about *K* make me do. But the intrinsic properties on which those desires depend for their effects also don't refer to *K* intrinsically. They refer to *K* only because of what they make my beliefs about *K* make me do.

So the hard questions remain. How do impersonal and tenseless beliefs and desires refer to people like *K* and times like *T*? What links our internal representations to the people, times and other things to which our beliefs refer? Those are the real problems of mental reference, which theories of it try with varying success to solve.

But subjective beliefs like *B*(*IN*) pose no such problems, because their referents need no causal surrogates. Causal contiguity makes the fact that any *X* believes *IN* at any time *Y* affect only what effects *X*'s desires would have: which automatically puts *X* into that belief's utility and hence its truth condition. And it likewise makes the same belief affect what effects *X*'s desires would have only at *Y*, which automatically puts *Y* too into its truth condition.

Similarly for all *X*'s other subjective beliefs at *Y*: e.g. the belief, *b*(*IK*), that he is *K*. But this belief will naturally make *X*'s *b*(*IN*) give him the further belief *b*(*KN*): that *K* now faces food. So *X*'s *b*(*IK*) and *b*(*IN*) must share intrinsic causal components, like v_K in my *b*(*KA*) and *b*(*KN*). But unlike v_K, these components come free with these beliefs, i.e. with the facts that *X* at *Y* believes *IK* and *IN*. For causal contiguity allows a token *b*(*IN*) to make a token *b*(*IK*) cause a token *b*(*KN*) only when all three tokens occur in the same agent at the same time (i.e. when the inference is valid). So *X* and *Y* are themselves the shared intrinsic causal components that make both *b*(*IK*) and *b*(*IN*) refer to

X and *Y*: in other words, they represent themselves. And as in *b*(*IK*) and *b*(*IN*), so in all subjective beliefs. They need no causal surrogates, no internal representations of the agents and times they refer to. And that's why they pose no problems of reference: the relation of reference in subjective beliefs is simply that of identity.

7 Conscious self-reference We have seen that agents need no concepts of the self or the present to have subjective beliefs, and those beliefs contain no internal representations of the agents and times they refer to. But these beliefs aren't, or needn't be, conscious. It remains to be shown that *conscious* subjective beliefs need no such concepts or representations.

To believe something consciously I must believe that I now believe it: since to believe something consciously is at least to be conscious of believing it. So even an objective belief like *b*(*KT*), when conscious, involves a subjective belief: namely, the belief that I now believe *KT*.

This 'second order' belief, I argue in chapter 3, is all that conscious belief is. Certainly it's all that could affect its content, and hence any concepts or representations it might need. Other contentious aspects of conscious belief – sensations, Humean feelings of conviction – we may therefore set aside.

We get our second order beliefs, I maintain, from an inner sense that I call 'insight', which mostly makes us believe we believe things we do believe, and not things we don't. Just as eyesight is how we know about the things we see, so insight is how we know about our own present beliefs. Insight is the perceptual mechanism of our privileged, if fallible, access to what we believe.

And insight not only gives us our second order beliefs, it fixes their referents, just as eyesight does. When I see a star, for example, my belief (that it is a star) refers to the object which, *via* my eyesight, causes that belief. Similarly, my second order beliefs refer to the first order beliefs which, *via* my insight, cause them. But these second order beliefs are subjective: they refer to my own simultaneous first order beliefs. So those must be the beliefs that cause them. And so they will be, thanks to causal contiguity. Insight being an inner sense, unlike eyesight, causal contiguity will let it show me only what *I* believe; which in turn lets it, unlike eyesight, work almost instantly, thus showing me what I believe only while I believe it.

Suppose then that an agent *X* not only believes *IN* at *Y*, but does so consciously. That is, his belief makes him believe that he has it. Since both these beliefs are subjective, neither represents the agent or the time it refers to internally: in both beliefs, that agent and that time represent themselves. And causal contiguity makes the agent and the time the same in both beliefs: so the

second order belief needs no internal representation of the owner or the time of the first order belief that it's about.

In short, even conscious subjective beliefs contain no internal representation of the agents and times they refer to: a fact which incidentally explains Hume's (1739 p. 252) notorious inability to observe himself and Wittgenstein's (1922 5.632) subject being no part of the world. But above all, it explains why even these beliefs pose no problems of self-reference – nor, given insight, of self-knowledge. And why, to have them, we still need no concepts of the self or the present: i.e. no beliefs, conscious or otherwise, about what distinguishes selves and times from each other and from other things, or oneself and the present from other selves and times.

8 How 'I' and 'now' refer So far so good. It now only remains to show that even when put into words, our subjective beliefs still pose no problems of reference, and call for no concepts of the self or the present.

What does it take to put my subjective beliefs into words? First, of course, I must want to do it, and must have beliefs about the right words to use. But what beliefs? What beliefs about 'I' and 'now' will make me use those words to express a belief like *B(IN)*?

The right sentence to express a belief is one that shares its truth conditions and hence its referents. But every subjective belief refers to whomever has it and when – and causal contiguity will make that belief make only that person express it, and only at that time. So the right sentence to express it is one that refers to whomever produces it and when. And, I believe, the words which make sentences like *S(IN)* do that are 'I' and 'now'. That's the belief about those words which makes me use them to express my subjective beliefs.

And all I need to do so is a habit, and some more subjective beliefs. The habit is simply this: my desire to express a belief habitually causes me to use 'I' when the belief I want to express is first person, and 'now' when it's present tense. Given causal contiguity, that habit will automatically make any token 'I' and 'now' which it makes me produce, at any time *Y*, refer respectively to Hugh Mellor and to *Y*.

The habit needn't of course be conscious. Native speakers needn't consciously choose the words they use. But even when I do choose 'I' and 'now' consciously, I don't choose them for their alleged reference-fixing 'senses': for, as we saw in **3**, I would choose them even if I lacked all sense of the time and of who I was. I choose them simply because I believe them to be the right words to use for the subjective beliefs I wish to express.

And what makes this belief true is just that everyone shares my habit of using those words to do this. That's what makes all our 'I' and 'now' tokens refer, in sentences like *S(IN)*, to those who produce them and when. And we understand each other because we believe truly that that's what they do. That true subjective belief is what makes us rightly take each other's 'I' and 'now' tokens to refer to those who produce them and when: and that's what makes them the right words for me to use for that purpose (Lewis 1969).

A knowingly shared habit is therefore all it takes for me to use 'I' and 'now' correctly and successfully at any time *Y*. I still don't need a concept of the self or of the present; nor need I believe that my 'now' refers to a present now, or to *Y*, or that my 'I' refers to my *I*, or to Hugh Mellor. To refer as they do, those tokens need no such problematic senses. All they need is our habit of producing them when we want to express our subjective beliefs: beliefs whose linguistic expression therefore poses no more metaphysical, semantic or epistemic problems than they do themselves – namely, none at all.

3 *Consciousness and degrees of belief*

1 Many of our beliefs come by degrees. Do beliefs about beliefs, especially beliefs about our own present beliefs, do so? In other words, do we have 'second order' degrees of belief? Skyrms (1980b) defends the idea that we do 'against charges of inconsistency, illegitimacy and triviality' and goes on to show its 'theoretical usefulness in connection with the laws of motion for rational belief' (p. 109). I go further. The idea of second order beliefs is not only legitimate and useful, we positively need it in order to provide a theory of conscious belief. And with this theory we can ward off a recent attack on the theory that degrees of belief are subjective probabilities. Combining these two theories seems, however, to have some unattractive consequences; but none, I believe, that need force us to reject them.

In what follows I develop and defend these three claims in turn. In Part **I**, I develop a theory of conscious belief, and extend it to take in degrees of belief. In Part **II**, I use this theory as part of a general defence of subjective probability as a measure of degrees of belief. (I anticipate this result from time to time in Part **I** by taking degrees of belief to be subjective probabilities. But that is only to simplify the discussion; it is not essential, and does not beg the question.) In Part **III**, I show how the theories of Parts **I** and **II** together seem to entail that we consciously believe all the consequences of our conscious beliefs. This clearly false conclusion in fact follows, not from these theories, but from the easy but erroneous idea that we always believe definite propositions, i.e. that the contents of all our beliefs determine definite truth conditions for them. Not, however, wishing to pretend that I have worked out any better idea, I shall continue to use the term 'proposition' for the content of a belief.

I

2 Ramsey (1926) showed how to measure the degree of a belief by measuring how strongly its owner is disposed to act on it. Ramsey apropos of its degree, and Braithwaite (1933) in general, identify belief with this disposition to act rather than with any conscious feeling of conviction. Feelings of conviction

may accompany belief, but they need not: 'the beliefs we hold most strongly are often accompanied by practically no feeling at all' (Ramsey 1926 p. 65). For example, 'I believe quite thoroughly that the sun will shine tomorrow, but experience no particular feeling attached to the proposition believed' (Braithwaite 1933 p. 142). Feelings may no doubt be among the causes and effects of gaining or losing a belief, but it is the disposition which directly determines how the believer will behave. The disposition is what provides the proximate explanation of his actions.

This might be denied, on the grounds that, as Molière thought, it is vacuous to explain events by saying that things – or people – are disposed to make them happen. But this is not always so. The most reputable explanatory properties of people and things are just conjunctions of dispositions to bring about the events they explain (see chapter 6). Even if a person's belief were just a disposition, using it to explain his behaviour would be no worse than using his weight to explain his effect on the bathroom scales. If anything, it should seem better to those who suspect dispositional explanation of vacuity, since a particular belief is not a disposition to any specific kind of activity: what its owner does will depend also on his desires and his other beliefs. So the beliefs and desires which are used to explain an action cannot be defined by it in the simple way in which Molière complains that 'dormitive virtue' can be defined. This is indeed an important fact about beliefs, and one I shall have need of later on. Meanwhile it will be convenient to mark the fact immediately by calling beliefs 'quasi-dispositions', and Ramsey's and Braithwaite's theory of them an 'action' theory rather than a dispositional theory.

It is in fact neither vacuous nor easy to characterise beliefs by the actions which, along with desires, they generate. In particular, it is not easy to say in general how changes in belief affect behaviour. Fortunately, we shall be concerned only with variations in the strength of belief, rather than in its content; and their effects are relatively easy to state. What they affect is betting behaviour, broadly conceived. As Ramsey put it, 'all our lives we are in a sense betting. Whenever we go to the station we are betting that a train will really run, and if we had not a sufficient degree of belief in this we should decline the bet and stay at home' (Ramsey 1926 p. 79). And as the degree of our belief in the train's running gets less, we may add, so the journey has to matter more to us to get us to the station.

This is not meant to imply that we consciously calculate odds whenever our degrees of belief affect our activity. Degrees of belief can influence quite unselfconscious actions. That is indeed how degrees of belief are best revealed,

since *ex hypothesi* such actions will not be being done deliberately in order to mislead.

It is precisely because an action theory does not invoke conscious feelings of conviction that it can readily allow beliefs to explain actions done while the agent's conscious thought is miles away. And so it should: that is how most actions are done. Crossing a two-way street in Britain, for example, I nearly always look right first, that being the direction I most strongly believe traffic on my side of the road will be coming from; but this thought crosses my conscious mind much less often than I cross British streets. Animals likewise, we suppose, can act on beliefs without having to be conscious of them. 'Often enough, my cat's behaviour makes it clear to me that he believes he is about to be fed', says Jeffrey (1983 p. 70), and I believe him. At least, I believe him more strongly than I believe his cat has any conscious thought about the matter. What is more, the animal could just as easily make it clear to me that he is none too sure of being fed, less sure for instance than he is of the prospects to be opened up by stalking a nearby thrush. In short, we readily grant some animals degrees of belief; and we grant them these degrees more readily than we grant them conscious convictions. It is a virtue therefore in Ramsey's theory to let them have the one without the other.

3 Virtues, however, have their price. The price here is that an action theory does not as it stands make sense of conscious belief. It is not, as Ramsey (1926 p. 65) claims, 'observably false ... to suppose that the degree of a belief is something perceptible by its owner'. On the contrary, their being perceptible to us is what makes immediately obvious the truth of Ramsey's other claim that 'the beliefs we hold most strongly are often accompanied by practically no feeling at all'.

If belief were a feeling of conviction, the phenomenon of conscious belief would pose no problem, since kinds of consciousness are just what kinds of feelings are. But dispositions to act are not kinds of consciousness, and nor are the quasi-dispositional states of belief and desire which, according to the action theory, cause them. Such states of mind as these will not automatically intimate their presence to their owners as feelings do. Nor can we credibly claim a general ability to become conscious of our own properties, whether they be physical or mental. Fortunately for the medical industry, for example, I cannot detect by introspection either what my blood group is or whether I am colour blind. My beliefs, on the other hand, I can call into my consciousness almost whenever I like (doubtless with Freudian exceptions). The Humean idea of belief as a kind of self-intimating feeling still appeals precisely because it

accounts immediately for this fact. And until the action theory can produce an alternative account of what conscious belief is and how it occurs, that Humean idea, for all its defects, will never be properly scotched.

Where should an alternative to the Humean idea of conscious belief be looked for? Armstrong, who espouses an action theory of belief, maintains that 'an account of having a belief before the mind, as a current content of consciousness, does not demand development of the theory of belief but rather of the quite general notion of consciousness' (Armstrong 1973 p. 22). I disagree: the contents of consciousness are too diverse. The action theory is after all forced on us in the first place because even conscious belief is so obviously unlike a feeling. It seems evident to me that conscious belief should be explained by relating it to its 'unconscious' action-producing counterpart, not by lumping it in a general theory of consciousness with items as unlike it as pains, visual sensations and emotions are. That at least is how I shall set about explaining it; it will be time to consider Armstrong's preferred alternative if and when it ever appears.

4 I need however to say a little more about conscious belief before offering a theory of it. I take it that a man's mind changes in some way when he acquires a conscious belief, even if he acquires it only by becoming conscious of a belief he already has. Suppose I walk towards the back of a room in order to get out of it. I walk that way because that is where I most strongly believe the exit to be. But this need not be a conscious degree of belief: consciously, I may be completely preoccupied with my reason for leaving the room. If I do start to think consciously about where the exit is, I shall come thereby into a new state of mind, different from the one which would in any case propel me towards the door.

For this new state, of conscious belief in the exit's whereabouts, I need a new name, in order to discourage the idea that it is just the quasi-dispositional belief state plus consciousness. This conscious state is actually what Hume (1739 p. 629) called simply '*belief*, which is a term that every one sufficiently understands in common life'; and no doubt the conscious state is still what most people think of as belief. But since our account of belief is going to start from the basic concept of the action theory, 'belief' is best reserved for that 'unconscious' quasi-disposition. As in my earlier paper on conscious belief (Mellor 1978), therefore, I follow Price (1969 p. 189) and others in using Hume's other term, 'assent', to refer to conscious belief. The term is not ideal, as I remarked on first adopting it: in particular, its connotations of a public display of acquiescence need to be discarded. Still, as Hume said, 'provided we

agree about the thing, 'tis needless to dispute about the terms'; and the mental state I have in mind is as familiar in common life today as it was to Hume, despite having proved so elusive to analysis. Throughout what follows, therefore, it must be kept in mind that this familiar inner state is what I mean by 'assent', not any of its outward manifestations in behaviour; and by 'dissent' I shall likewise mean the inner state of conscious *dis*belief.

Assent, thus understood, evidently comes by degrees as much as plain belief does. Assenting to something is at least consciously believing it more strongly than its negation, and is quite often not much more than that. While between assent and dissent, as between belief and disbelief, lie the intermediate degrees of doubt, in this case of conscious doubt. In order to talk about all these conscious states together I shall use 'assent' as 'belief' is also used: namely, not only for one extreme of this family of mental states, but also as a name for the family itself, ranging in degree from what we may call full assent, through increasingly sceptical shades of doubt, to outright dissent.

Not only can we assent in varying degrees to propositions of all sorts, we incessantly do. Almost whenever we are conscious, that consciousness includes assenting to something, or dissenting from it or consciously doubting it. This aspect of our consciousness moreover is what gives us an almost instant access to nearly all our own beliefs, which we have to no one else's. I may see some of your beliefs reflected in your behaviour (as Jeffrey sees his cat's), but I need not watch my own actions to get my own beliefs and doubts into my consciousness. Suppose for example I hear you put to me almost any question, '*h*?' which I understand. I will at once be made conscious, as you will not, of whether I believe *h*, disbelieve it or doubt it: i.e. I will at once assent to *h* in some greater or lesser degree. Now I may of course be wrong, not only about *h* itself but about the strength of my own belief in *h*. That is, not only may I assent to *h* when *h* is false, I may assent to *h* when I do not believe it. The degree of my assent, in other words, may differ markedly from the degree of plain belief which my unselfconscious actions would reveal to others. And it is an important fact that when this happens, we take it to be a case of self-deception: what we conclude is that I do not really believe what I consciously believe. This shows both how assent is thought to involve a fallible perception of belief, and how right the action theory is to take the unconscious state, and not assent, to be the paradigm of really believing something.

Self-deception about belief is, I suppose, almost always possible. But it is not common; and, right or wrong, my assent does at least tell me something about my belief, which it does not tell me about anyone else's. This is the familiar fact, referred to in **3**, of our privileged access to our own beliefs.

This is the fact that an action theory needs to account for by construing assent also as a disposition or quasi-disposition to action.

5 The cue for an action theory of assent is given by the fact that some behaviour needs assent, not just plain belief. In particular, linguistic behaviour needs assent. To talk or write to someone I have to be conscious of what I am talking or writing about. My plain belief, that traffic here keeps left in two-way streets, may steer me safely through a town without it becoming a conscious belief, but it will not steer me through a conversation on the subject. For that I need assent. I do not mean that I need to assent fully to everything I say, since I can of course choose to lie. But what I lack when I lie is not consciousness of my state of belief. On the contrary, I have to be as conscious of what I believe in order to lie about it, as I do in order to tell what I believe to be the truth. What I lack when I lie is the degree of assent which is implicit in the remarks I choose to make. To lie, I must dissent from what I assert, or at least consciously doubt it more than my speech conveys. Plain disbelief is not enough.

Some degree of assent therefore is needed for every assertion. But assent is not a simple disposition to assert, as the case of lying shows. What I say depends not only on what I assent to, but on how I desire to affect my audience. I shall only assert what I assent to if I want to tell my audience the truth; otherwise I shall deny it, equivocate or say nothing at all. So no degree of assent, however high, can be defined as a disposition to make any specific assertion, any more than plain beliefs, however strong, can be defined as dispositions to do any other specific thing.

Assent, therefore, like belief, at least includes a quasi-disposition to act – but only to act in special ways, such as conversing, which call for it. Now if we actually identify assent with such a quasi-disposition, we can spare ourselves the unhopeful search for a special kind of consciousness to accompany, or partly constitute, assent. We already know that no feeling or sensation does so: the whole point of the quotations from Ramsey and Braithwaite in **2** was to invite us to make some strong belief conscious, i.e. to generate a high degree of assent, and then consciously to note the absence of any particular feeling.

Our problem then will be to relate these two quasi-dispositions, assent and plain belief. The problem is especially acute for those of us who wish to follow Ramsey in measuring the strength of a belief by the believer's choice of odds for a bet on its truth. For literally choosing odds, like talking, calls for consciousness of the proposition in question, i.e. it calls for assent, and not just

for belief. We have to give some reason for assuming that a conscious choice between explicitly risky alternatives will reveal the strength of beliefs of which we need not, even in action, be conscious at all. Nor even perhaps, in the case of animals, be capable of bringing to consciousness. We have credited animals with beliefs, and with beliefs of variable strength, but not with what it takes to choose odds, not with degrees of assent. Why should the strength of an assent of which they are incapable be supposed to measure the strength of the essentially unconscious belief that they actually have?

6 I propose to solve these problems with the simplest possible action theory of assent: namely, that *assenting to a proposition is believing one believes it.* I put this theory forward first in Mellor (1978), where I deliberately confined it to the qualitative case of full belief and full assent. But since belief and assent do in fact come by degrees, the theory must be shown also to accommodate that fact; and this is my present task. First, however, I should summarise the qualitative theory, correcting some errors in the original version and dealing with some objections that have been made to it.

I write '*Btah*' for a person or animal *a* believing a proposition *h* at a time *t*. Where the identity of *a*, *t* or *h* is immaterial, I abbreviate this as convenient to '*Bth*', '*Bah*', '*Bh*', or just '*B*'. Disbelief I take to be belief in the negation, which I write '*B*~*h*': '~*Bh*' means merely that *h* is not believed, which might mean disbelief, or doubt or merely a lack of any attitude at all to *h*. Originally I took *a* believing at *t* that he believes *h* to be *BtaBtah* (Mellor 1978 pp. 90–1), but this is wrong. Believing one believes something is an essentially indexical state: it is *a* believing at *t* that *he* believes *h now*. This is not the same as his believing that *a* believes *h* at *t*, as Perry (1979) has shown, even though he is *a* and *t* is the present time. For instance, *a* may have forgotten the time, or even who he is, and not believe *Btah* under any true non-indexical description of *t* and *a*. But he can still believe that he himself, whoever he is, believes *h* now, whatever the time now is; and this is the state that concerns me.

Since this state of believing one believes is indexed in this way to its owner and the present time, it will suffice for my purposes to write it simply as '*BtaB***h*', or '*BB**' for short, the '*' serving to indicate its indexicality in these two respects.

Then if *A* is the state of assent, as *B* is of plain belief, my thesis is that

(1) $A = BB^*$.

7 (1) involves two claims: (1a), that *A* entails *BB**; and (1b), that *BB** entails *A*. My argument for (1a) may be summarised by the following excerpts from Mellor (1978), duly modified to correct the error noticed above.

> There is *prima facie* such a thing as believing one believes something. I have beliefs about all sorts of things: why not about my own beliefs? [Moreover] we recognise a difference [between *BB*h* and *Bh*], as our concept of a state of self-deception shows ... A husband, we suppose, can (subconsciously) believe his wife to be unfaithful, while (consciously) believing that he believes nothing of the sort. (p. 91)
>
> If we have such states [as *BB*h*] at all, they surely occur when we assent to propositions ... Assent is the conscious belief in *h* that is required for the sincere affirmation of *h*'s truth. In coming to assent to *h* I have perceived (or in cases of self-deception, misperceived) my belief in *h*: if that does not involve believing one believes *h*, what does? (p. 92)

In the original discussion, the following counter-example was proposed to this rhetorical question. Suppose I am reluctantly persuaded, e.g. by a psychoanalyst, that I subconsciously believe some nasty proposition about my parents. Since it makes no odds to the argument, I shall for ease of exposition represent this as subconsciously disbelieving its negation, the nice proposition *h*. Then, despite the analyst, I still *assent* to *h*; but, it was alleged, I no longer believe I believe it, which shows that (1a), and therefore (1), is false.

Not so: the case has been misstated. What we really have here is my assenting to a proposition *h* and also *assenting* to the indexical proposition *~B*h* (= I don't now believe *h*). That is not *Ah* and *~BB*h* as was alleged; it is *Ah* and *A~B*h*, which according to (1) is *BB*h* and *BB*~B*h*, two perfectly compatible states of mind. Now I had indeed conjectured (Mellor 1978 p. 92) 'that we have in fact no more distinct states [*BB*B*h*, *BB*B*B*h*] etc.' beyond *BB*h*; and this case perhaps shows that we must occasionally allow *BB*B*h* to be a state distinct from *BB*h*. But my chief contention, (1a), survives unscathed: assenting entails believing one believes.

The converse contention, (1b), is less obvious. If *Bh* can exist without my being conscious of it, as on an action theory it clearly can, one might ask why *BB*h* cannot do so. And if it can, it can presumably exist without assent, in which case (1b) is false. But this suggestion misses the whole point of (1). The thesis is, after all, that for *Bh* to be conscious just *is* for the state *BB*h* to

occur. For that state in turn to be conscious would therefore be for *BB*B*h* to occur. But with the possible exception just noticed, I see little reason to suppose that there is any such state distinct from *BB*h* itself. And of course this account does not invoke consciousness in saying what assent is. That is a virtue, not a defect, since its very object is to say in other terms what this elusive kind of consciousness is.

What a defender of (1) does need to do is, positively, to use it to explain why the kinds of behaviour that need assent do so, i.e. why they need *BB** and not just *B*; and, negatively, to explain away putative counter-examples. In my (1978) I undertook both of these tasks; here I briefly summarise the results.

Positively, I used a couple of Gricean truisms about language to show that sincerely communicating any *h* calls for *BB*h*, not just *Bh*. (What I originally claimed in Mellor (1978 VII) was that communication needs *BBh*, but in fact the token-reflexive state *BB*h* is what the argument really shows to be needed.) In being told that *h*, the hearer is intended by the speaker to be convinced that the speaker himself believes *h* as he speaks. To have this intention the speaker needs, if he is sincere, to believe that he himself now believes *h*, i.e. he needs to be in the state *BB*h*. (If he is being insincere, the state he needs to be in is *BB*~h*.) And if (1) is true, that fact explains why he needs to be in the state *Ah* (*A~h*), because it is the very same state. The thesis does therefore explain why conversation needs assent, not just belief.

To dispose of most proposed counter-examples to (1b), it suffices to insist on the essentially indexical, present-tense nature of assent. Tenseless beliefs about beliefs, beliefs of tenses other than the present tense, beliefs about the beliefs of other people: none of these is assent, and any of them could exist without assent, for all (1b) says. The suspicion that *BB** could do so stems largely from confounding it with other sorts of belief about beliefs.

But one apparent counter-example to (1b) is properly indexed: namely, self-deception itself. A man deceives himself if he believes he believes *h* when he does not believe *h*: i.e. *BB*h&~Bh*. But we should not normally say that he was self-deceived about his belief in *h* only when that belief was conscious. So it seems that he could be in the state *BB*h* without being in any degree of the state *Ah*; so (1b) is false.

In fact this is not so. The state of self-deception that can exist even when the man is not assenting to *h* is not in fact *BB*h&~Bh*. Now in saying this I am not just trying to rescue (1b) by stipulation. Recall that our objective is to extend the action theory to cope with assent, and the basic concept of belief to be used must therefore be that of the action theory itself. We have already seen that it differs from everyman's concept in not implying consciousness; it

differs also in not applying to this state of self-deception. This is because what the action theory calls belief is a quasi-disposition (see **2** above): i.e. how a belief displays itself in action depends on what else its owner believes and what he desires. But this is not true of self-deception and self-knowledge. They are straight dispositions. Like fragility and Molière's 'dormitive virtue', these states always display themselves in the same way, by which therefore they can be defined. Regardless of what else he believes and desires, anyone who assents to (dissents from) *h* only when he believes it is, by definition, a man who knows (deceives) himself in this respect.

An action theory of belief that can cope with assent is thus automatically equipped to cope with self-knowledge and self-deceit. The man who believes *h*, and knows himself in this respect, is a man *disposed* to assent to *h*, whether he is actually assenting to it or not. This indeed is a state distinct both from just believing *h* and from actually assenting to it, even though it is definable in terms of those two states. So there are three states to be accounted for altogether, but with the addition of our thesis (1), the action theory can account for all of them: the basic quasi-disposition, *Bh*; assent, *BB*h*; and the disposition, to *BB*h* (if $BB^*h \vee B\sim B^*h$) only if *Bh*. This disposition is not a belief as action theory construes belief. *A fortiori* it is not the belief $BB^*h\&\sim Bh$; so the fact that it can exist without *Ah* is no objection to the thesis (1).

8 Although self-knowledge is definable in terms of belief and assent, it is nevertheless a distinct, and a real, property of those that have it. It is a property distinct from belief and assent in just the way the fragility of a glass is a property distinct from its being dropped and its breaking. And, like fragility, self-knowledge is a real property in the sense, for example, that changes in it are real events with causes and effects (see chapter 6.1). It has in fact a causal mechanism, whereby whenever one's belief about *h* comes to consciousness, the state *Bh* will cause the state *BB*h*, and the state $B\sim h$ will cause the state $BB^*\sim h$. It is the means whereby, as Ramsey (1926 p. 67) put it, 'what determines how we should act determines us also directly or indirectly to have a correct opinion as to how we should act'; only not, as Ramsey thought, 'without its ever coming into consciousness'.

The mechanism of self-knowledge is like the mechanism whereby, when something we are looking at is red, it causes us to believe it's red, and when it isn't, to believe it isn't. In short, self-knowledge is a species of perceptual ability, which enables facts to cause us to believe in them. That is, its causal mechanism is a sense, just as eyesight is a sense. Only, since what it informs us

of is one of our own inner states of mind, it is an inner sense. It will be embodied entirely within the brain, because that is where both its objects, *B*, and the perceptions it delivers, *BB**, are embodied. Its organs are therefore naturally unfamiliar to us, because they are less visible than those of our outer senses; its very existence, indeed, is for this reason quite easily overlooked. This inner sense is nevertheless crucial to my account of the relation between belief and assent, and in particular of the relation between first and second order degrees of belief. I must therefore beg leave to expatiate on it to some extent.

I mean the claim, that we have an inner sense, literally, not just metaphorically; and I defended it at some length in Mellor (1978 IX), where I took the liberty of calling it 'insight'. Insight is the mechanism of the privileged access that we have to our own beliefs, and as such it completes in principle the account, with which I said in **3** the basic action theory must be supplemented, of the phenomenon of assent. Details of the working of our insight, like the working of our other senses, are for physiological psychology to supply; what matters here is that we recognise its existence, and thus that there is nothing either infallible or scientifically ineffable about introspecting our own beliefs. When we introspect beliefs we are simply perceiving them with a special sense, just as with other senses we perceive visible and audible aspects of the outside world.

Now the outside world of course contains more visual detail than meets the naked eye. In the same way, our minds no doubt contain more cognitive detail than meets our unassisted insight. The extreme case, corresponding to literal blindness, would be that of animals who have beliefs but have no insight at all. Since they lack the means of perceiving their own beliefs, they cannot assent in any degree to what their actions show them to believe. This in turn explains why they have no language, being incapable of Gricean intentions to communicate their beliefs, and it does so without depriving them at all of sensations, emotions and other important aspects of consciousness which languageless animals may well possess.

People, however, are not internally blind. We do have insight; but it is not perfect, any more than our eyesight is. At any time, most of our beliefs are not being perceived by us; and those that are will not be perceived in every detail. Thus, to anticipate an observation that is going to cause trouble in Part **III**, we are never aware of all the consequences of our conscious beliefs, not even of all those consequences which we do in fact believe. I believe, for example, that to drive quickly from Cambridge to London I should start off along Trumpington Road; and given my other geographical beliefs, it follows

that I should turn left from the end of Lensfield Road. But I can easily be conscious of the first of these beliefs without being conscious of the second.

Moreover, as we have seen, insight, like eyesight, is fallible. People can and do deceive themselves. And once we recognise that insight is just another sense, there is no more mystery about recognising self-deception than there is about recognising colour-blindness. Naturally neither state is directly perceptible to its owner. But both can be seen by others, and in essentially the same way, namely by comparing the patient's faulty perception with the thing perceived. And (*pace* Davidson 1970b and others) there is no special mystery about perceiving people's beliefs and assentings. Mostly we infer them from what we know they have seen and heard. For the rest, we infer people's beliefs from their unselfconscious actions, and what they assent to from their unstudied speech. That is why, when someone's seemingly sincere speech makes too little sense of the rest of his behaviour, we infer that he is self-deceived – as when we listen sceptically to the subconsciously suspicious husband's eager rationalisations of his incessant phone calls home.

Admitting the fallibility of insight is thus by no means admitting wholesale scepticism about its deliverances. Insight we know to be no less reliable than our outer senses: for one thing, they themselves rely on it in conscious observation, since that gives us assent as well as plain belief. And we know that our senses are reliable, that people mostly see things as they are and believe what they assent to, because we can see that they do. The reliability of our senses is a straightforwardly observable fact about the world of which we, and they, are perceptible parts. It is important to realise this, and moreover to realise that, as Grandy (1980) remarks, there is no epistemologically vicious circle or regress involved in it. It is because our senses are reliable that our seeing that they are amounts to our knowing it: i.e. it is a true belief got by a reliable process (Ramsey 1929b p. 121). We need not prove *a priori* that our senses are largely free of error before using them to confirm the fact (see chapter 15.5).

It follows in particular that we need no special or *a priori* justification for assuming that insight generally reveals to us the degree of our plain beliefs. When Ramsey assumed that a conscious choice of betting odds generally reveals the degree of our unconscious belief, he assumed nothing that calls for *a priori* defence. He was simply, and rightly, taking for granted the observable reliability of one of our senses.

9 If (1) is true, and belief comes by degrees, assent should come by degrees in two distinct ways. Suppose my plain or 'first order' belief in h is neither full

belief nor full disbelief, but is of some intermediate degree of doubt. For that to be the conscious degree of my belief in *h* is, I take it, for me to believe fully that I have this degree of belief in *h*. That is the natural way to construe degrees of assent. Insight, like our other senses, usually delivers full belief about what we perceive by means of it. But, again like our other senses, it need not. Just as poor eyesight may leave me unsure of the colour of an object, so poor insight may leave me unsure of the strength of my first order beliefs. In that case, if my degrees of belief are subjective probabilities, I should have a subjective probability distribution over possible values of my first order degrees of belief in *h*. And this is a second way, according to the theory, in which assent should be capable of coming by degrees.

Is this really so? How, it will be asked, do we distinguish second order subjective probabilities from first order ones? What is the difference between being sure of doubting something and doubting that one is sure of it? It is natural to suspect second order probabilities of collapsing onto first order ones; and if they do so collapse, so does my theory. So the suspicion needs to be dispelled. Let us therefore consider some *prima facie* cases of second order probabilities.

Suppose for example I know a coin to be fairly tossed but fear it may be double-headed. That is, I know the chance of heads is either 0.5 or 1, but am not sure which. Suppose I think each value equally likely. Here we have second order probabilities: degrees of belief in alternative values of the objective chance. (Actually, we have degrees of assent, in the first and straightforward sense; but therefore also degrees of belief if I am not self-deceived – and as the extra element of consciousness is immaterial, I shall discuss the case in the simpler terms of plain belief.) Now as far as my expectation of heads goes, these second order probabilities do indeed collapse onto first order ones. Their net effect on me is that I should bet on heads at 3:1 on, i.e. behave as if I thought the chance 0.75. So how do these supposed second order probabilities show up in my behaviour?

They show up, for one thing, in how I should react to seeing the results of successive tosses: in particular, in how my degree of belief in heads resulting from the next toss should be affected by them. It affects, as Skyrms says, 'the laws of motion for rational belief'. Tails just once, for example, should instantly reduce the degree of my belief in heads to 0.5, by proving that the coin is not double-headed; whereas it should have no such immediate effect if I had merely thought the chance 0.75.

So the second order probabilities do not in this case collapse completely onto first order ones. But in this case the first order probabilities are

objective. There is an objective difference between a coin's being double-headed and its merely being biased. Doubt about the one possibility is bound in the end to be distinguishable from certain belief in the other by how one reacts to evidence as to which is actually the case. When both first and second order probabilities are subjective, however, it is not so obvious how they are to be distinguished. If I doubt a proposition *h*, why should I react differently to fresh evidence for *h* or against it just because my doubt is second order rather than first? How, for example, will the alleged difference show up between (i) believing to degree 1 that I believe *h* to degree 0.75 and (ii) believing to degree 0.5 that I believe *h* to degree 1, and to degree 0.5 that I believe it to degree 0.5?

On the theory I am propounding, the difference will be a difference in consciousness and hence, in particular, in linguistic ability. In state (i), I am conscious of a precise degree of doubt in *h*, which I can if I wish sincerely report. I can predict exactly how I should react to options whose desirability depends on the truth of *h*. I would have no hesitation now in picking precise odds for a bet on *h*'s truth.

State (ii) is a very different, and a very odd state. What is odd about it is the discontinuous two-peaked second order probability distribution over the first order degrees of belief. Our senses, inner and outer, do not usually deliver such bizarre distributions. My eyes, for example, would rarely if ever make me see the colour of something to be possibly either pure white or pitch black but certainly nothing in between. They normally give me a roughly bell-shaped probability density distribution over a continuous range of colours similar in hue, brightness and saturation. That is, as the colour becomes more similar the degree of my assurance, that it isn't the colour I saw, diminishes more rapidly. And similarly with insight. My consciousness of my doubts usually consists in a roughly bell-shaped probability density distribution over my first order degrees of belief. That is, the closer two degrees of my first order belief become, the more rapidly does the degree of my assurance, as to which one I really have, diminish.

State (ii) seems bizarre simply because it is so rare. It is, nevertheless, a perfectly intelligible state, and there is no difficulty in saying how it differs from state (i). In state (ii) I should be less aware of what I believe about *h*. I could sincerely say that I do not totally disbelieve *h*. But all I could say positively is that, in unselfconscious actions to whose outcome *h* mattered, I should either act as if I were certain of *h* or dither in a way equivalent to betting on *h* at evens; and that, if asked now to bet on which of these two I would then do, evens are again the odds I should propose.

The oddity of (ii) does not therefore count against our theory of assent. On the contrary, that theory is what tells us what it would be like to be in state (ii), and what people in it would be apt to do and say; and this is how we know that state (ii) is so rare. We know we ourselves are almost never in states like that, and that other people are not apt to do and say such things. Our theory of assent, in short, makes plenty of behaviour and respectable introspection available to distinguish state (ii) from state (i) – and to distinguish both from the unconscious and inarticulate 0.75 degree of merely first order belief in h, which my unselfconscious action might reveal, with no degree (not even zero) of second order belief in h at all.

So second order subjective probabilities do not collapse onto first order ones. They are introspectibly distinct states of mind and show up in quite different kinds of behaviour. But there are of course causal connections between them. They are connected one way by insight: with good insight, a degree of first order belief will produce a very high degree of second order belief in its true value. There are also connections the other way. Conscious thought is, amongst other things, a means of affecting our habits of action, i.e. our desires and the degree of our first order beliefs. After a shock from an electric kettle, considerable conscious reflection on the safety of the rewiring may be needed to make me pick it up again without hesitation, i.e. to drive the degree of my first order belief in its safety back towards 1. But it can be done: calculation can affect our beliefs just as observation can. I may not, by taking conscious thought, be able to add cubits to my stature, but I can add degrees to my first order beliefs. In other words, we are not only disposed to assent to what we believe: we are also disposed, on the whole, to believe what we assent to.

II

10 If the thesis (1) is true, we will have second order subjective probabilities if we have first order ones. Higher orders are more doubtful. I conjectured in **6** that we generally lack third and higher order beliefs: insight does not show us all our inner states. But even first order subjective probabilities are controversial. One may admit that beliefs vary roughly in strength and still resist the full panoply of subjective probability. First order beliefs do not obviously come in such a finely graded range of strengths, and the second order range is still less obvious. The whole apparatus may reasonably be suspected of fantasy, of the heroic but unrealistic model building that seems, among some Scandinavian philosophers, to have superseded the saga-making of old.

Kyburg (1978) recently marshalled a notable attack on subjective probability, and one can readily imagine his reaction to iterating it as I have done. But his attack seems to me mostly misdirected, and the rest of it can be met with the aid of the present theory of assent. With it we can in particular remove from the concept of subjective probability a pernicious equivocation of which Kyburg rightly complains.

Subjective probability is mostly presented as measuring how strong people's beliefs actually are. But there is evidence, which Kyburg adduces, purporting to show that the strength of people's beliefs does not in fact satisfy this measure. Faced with this evidence, subjectivists have been apt to claim only to be prescribing the measure on grounds of rationality, i.e. to say not that people do have probabilistic degrees of belief, but that they are irrational if they don't. But subjectivists who say that cannot also claim to be doing psychology. In particular they cannot claim to explain away people's actual agreement on objective probabilities as being merely the result of their conditionalising their subjective probabilities in response to shared evidence. Moreover, reference to rationality at this stage seriously confuses the question of what people's degrees of belief actually are with the question of what they ought to be, given knowledge of chances or of other inductive evidence for or against their truth. So it is important on several counts to decide whether subjective probability is supposed to be a theory of real or merely of rational degrees of belief. The theory of assent will enable us to do just that.

I must emphasise that in what follows I shall not be arguing for the full-blooded subjective view of probability. I do not believe for a moment that one man's degree of belief is as good as another's, nor that conditionalisation does in fact explain away the phenomenon of agreement on values of objective chance. It is true that my view fits Kyburg's definition of 'subjective probability', since I do think 'the assignment of a numerical probability ... [need] not reflect any known or hypothetical frequencies'. But Kyburg's is an unreasonably broad definition: one can quite well be an objectivist without being a frequentist. On the other hand, I do wish to characterise objective probability by the constraints it prescribes for degrees of belief, so I have some interest in defending subjective probability: namely, that the statement of my views will be somewhat simplified if degrees of belief have a probability measure.

11 Kyburg does not deny that we think of belief as coming in rough and comparative degrees. He would not dispute the examples I have given, and I need not therefore multiply or defend them. But it is still worth emphasising how

deeply the concept is rooted in our everyday judgments. Subjective probability is not a radical innovation of theoretical psychology such as Freudian theories, for example, have proposed. It is no more than a quantitative development of a perfectly familiar state of mind which we freely and uncontentiously attribute to ourselves and other people all the time. The only question is why one should jib at such a development.

Let us compare it, as Ramsey (1926 p. 69) did, with Newtonian mechanics. That is a quantitative development of rough and comparative ideas of force and mass, i.e. of what sets things moving and of how hard things are to move. No one jibs at this development and, though I admit it is often easier to measure forces than beliefs, the case for each is much the same. In each case we have a theory that postulates a continuum of degrees of a state whose rough gradation is a commonplace. Each postulation is supported by quantitative applications which link the quantity proposed with its causes and effects. Perception produces beliefs of various strengths which in turn combine with desires *ditto* to produce definite action even in conditions of uncertainty. Similarly, gravity and other causes produce forces of various strengths, which in turn combine with masses *ditto* to produce definite accelerations. Force and mass are quantitative quasi-dispositions to yield the behaviour they explain, just as belief and desire are. And the range of behaviour explained by the theory that invokes each of these pairs of states is what gives us reason to believe in their existence.

The reason may look more impressive in the case of force and mass, but the difference is only one of degree, and the degree should not be exaggerated. One could easily level at Newtonian mechanics most of the criticisms brought against subjective probability. One could object especially to the factual underdetermination of Newton's theory. Its apparatus can easily be made to appear grossly disproportionate to the data which it is based on and which it is used to explain. A thing's mass, for instance, is an infinite conjunction of the dispositions it has to accelerate under forces of different strength. Only one of these dispositions is ever displayed at any one time, and an infinite number will never be displayed at all. One might well ask what the point can be of attributing to a thing so extravagant a profusion of dispositions, and how one could possibly claim to know the truth of such attributions.

The second, epistemological, question I shall not try to answer. The answer to the first question can only be that these Newtonian dispositions are supposed to be present, whether they are displayed or not, in order to be available to explain any of an infinite number of possible interactions with other things. That is, we suppose it to be true of the thing now that it would accelerate at

definite rates under all these definite forces, whether or not it will ever be exposed to them. Well, it may in the same way be true of me now that I would behave in definite ways in an infinite range of definite circumstances, whether I am ever in them or not. I may have a quite precise degree of belief, even though it will only show up in actions which, given my desires and other beliefs, a slightly weaker (or slightly stronger) belief would make me forgo. The degrees of most of my beliefs never will show up in action. But that will be no reason to deny their existence unless I am also to be denied Newtonian mass.

Now I am not in fact claiming precise degrees for my beliefs, any more than I would claim to have a precise mass. Nor does the theory of subjective probability entail so rash a claim. Its mathematics indeed makes sense of indefinitely precise degrees of belief, just as Newtonian mechanics does of precise degrees of force and mass. That is because it is not for the mathematics to limit *a priori* the precision of these states. Their precision of course is limited, as that of all quantitative states is (see Mellor 1967), but only *a posteriori*, by the several natures of the kinds of things involved. It is nonsense, for example, to give the mass of Everest to the nearest gram; but that does not prevent Newtonian mechanics applying to the Himalayas. Similarly, subjective probability is not seriously impugned by the obvious fact that most of our beliefs have imprecise degrees, and are more properly represented by intervals than by precise values of subjective probability (cf. Levi 1974).

Imprecision is not a feature peculiar to degrees of belief, and its significance should not be exaggerated. Nor should its extent. We may quite easily have more precise degrees of belief than we think we have. Assent, we have seen, is a product of insight, and insight is no more precise a sense than eyesight is. My unselfconscious choices might well display more definiteness in my first order doubts that I can in advance be consciously aware of. I may be conscious only of thinking something probable, i.e. that my degree of belief in it lies in the interval (0.5, 1), when my first order degree of belief actually only spans the narrower interval (0.7, 0.8). No one may know this, just as no one knows the mass of most things; yet, it may be so.

12 All this, however, is still only a preface to the main dispute. Kyburg's chief complaint is not that subjective probabilities are too inactive, too numerous or too precise for the behavioural data they purport to explain. His complaint is that the Dutch book argument he considers fails to show that degrees of belief are subjective probabilities at all; and that, on the contrary, real betting behaviour shows them not to be. And to answer this complaint, the

relation of betting to belief needs to be examined more closely than it usually is.

We do not normally learn the strength of people's beliefs by making them choose odds for bets, any more than we normally measure their masses by their accelerations under unit forces. We infer beliefs and masses alike from behaviour which is also affected by many other factors. Such inferences of course need knowledge of what the other factors are, and of how they interact with belief and mass respectively to produce the behaviour we observe. Moreover, we need this knowledge even to discuss how in principle these quantities could be directly and unequivocally measured: because, to measure them, we have to try and specify a situation in which the influence of the other factors we know of can be either eliminated or allowed for. Then we can take behaviour in that situation to be an explicit measure of belief, mass, or whatever else concerns us. That our specification will succeed, that the measure will be of the very quantity we want, cannot be guaranteed *a priori* or independently of theories in which the quantity already figures, and which tell us what other factors interact with it. It is not a matter of an arbitrary specification stipulating what we are to mean by degree of belief or mass: we already know most of what we mean, and our specification must accommodate itself to what we know. It is not, for example, by stipulation that IQ tests measure intelligence. If they do, it is because we already know that intelligence is one of the factors which affect their results, and that the other mental factors which might do so have, as a matter of fact, had their effects eliminated or allowed for. What these other factors are, and how their effects may be eliminated, is for our psychology to say; the information is not given us either by pure reason or by raw uninterpreted experience.

So it is with the betting measure of belief. We know a belief's strength affects its owner's choice of odds, or quotient, for a bet on its truth. But we also know that this is not the only mental factor which affects the choice. So we must constrain the betting situation in order to eliminate or allow for the effects of these other factors. In Mellor (1971 ch. 2) I proposed constraints to that end: the bet was to be compulsory, with the opponent choosing its direction and the stake after the quotient is fixed. The point of these constraints is to prevent the quotient being directly affected by a preference for particular stakes, by desire for a particular outcome, or by a like or dislike of the process of betting. The constraints moreover do this in a way that was explicitly designed to meet objections which Kyburg has now resurrected against the inept Dutch book argument that is unfortunately the only one he considers. That argument tries to use the least odds a man would accept for a

bet on h's truth in order to measure his degree of belief in h – and this, as Kyburg reminds us, is nonsense if the man is to be made to bet, as he must be for a Dutch book to be made against him. 'No odds can be unacceptable to a man who is compelled to bet in any case' (Mellor 1971 p. 36). But it makes perfect sense for a man to say what odds he would choose if he had to bet on h with his opponent deciding subsequently which of them will win if h turns out to be true. That is a very plausible measure of his degree of belief in h, if he has one, whether he is willing to bet or not.

It must be emphasised that the justification for placing these constraints on the betting situation has nothing whatever to do with making degrees of belief satisfy the probability calculus. The constraints are there simply to prevent the quotient being affected by factors other than the degree of belief it is supposed to measure. They may not, as a matter of fact, suffice to make the quotient measure nothing but degree of belief, but they are clearly at least necessary to that end. And they do suffice to set up a valid Dutch book argument. The only way a man who is compelled to bet, at stakes and in directions subsequently determined by his opponent, can prevent certain loss is by choosing 'coherent' quotients, i.e. quotients which are probabilities. That the loss would otherwise be certain follows from the fact that his opponent is also trying to win the bet: if he weren't, it wouldn't be a bet at all. But nor would it be a bet if one party were certain to lose; it would only be a pointlessly complex method of giving goods away.

It is irrelevant to remark as Kyburg (1978 p. 162) does that professional gamblers make money quite rationally by offering non-probabilistic betting quotients. So they do, but they would not do so in situations constrained to reflect only their degrees of belief and not their greed. They would be coherent all right in the situation I have specified.

The experiments Kyburg cites (p. 165), in which actual quotients close to 1 and 0 were not coherent, are equally inconclusive. His description of the experiments makes it clear that they did not satisfy the constraints needed for the quotients to measure only degrees of belief. The choice of quotients was clearly open to influence by some of the other factors I have mentioned, factors which would be especially likely to affect values close to 1 and 0, since these values greatly increase the possible gain and loss. Those experiments no more falsify subjective probability theory than a car's needing an engine to keep going on the flat falsifies Newton's first law of motion.

13 Still, it may be urged, even in the betting situation I have specified, a man could choose incoherent betting quotients. He will lose money, but he need not

have had that intention; and if he did not, that intention cannot have been the cause of his behaviour. Lacking any other cause, then, we can only attribute it to his having non-probabilistic degrees of belief. We may of course suspect other unknown causes – but, after all, there may not be any. To insist *a priori* that there must be would simply beg the question in favour of subjective probability.

I remarked in **10** that many subjectivists admit that people may have non-probabilistic degrees of belief, and use the Dutch book argument only to show that such people are irrational. And that much the argument certainly shows: a man who has no desire to give away his goods should certainly not choose betting quotients which he knows will inevitably have that effect. But we can do better than this. With the aid of the theory of assent developed in Part **I**, we can show that degrees of belief really are probabilities, not just that they ought to be.

First we should observe that the argument just given for the possibility of incoherent quotients only shows that degrees of belief might not be probabilities, not that they are not. It would only defeat an *a priori* argument for a probability measure, and although I hope my argument will prove compelling, I do not mean it to be *a priori.* The facts it appeals to, about assent and the psychology of betting, are entirely contingent. It could be defeated by an experimental proof that degrees of belief are not probabilities; but Kyburg, we saw in **12**, has provided no such thing.

Next, I have to say that the significance of betting for the measurement of belief has been uniformly misconceived in the literature, a misconception with which I have so far gone along. It is actually not at all like measuring masses by accelerating them, or even by weighing them. We cannot directly discover degrees of belief by observing quotients chosen for actual bets. The most we could discover directly in this way are degrees of assent, not degrees of first order belief. Choosing odds is a selfconscious activity, as I remarked in **5**. What it reveals is what the chooser believes the degree of his first order beliefs to be. Usually he is right, for the reasons discussed in **8**: insight is a generally reliable sense. But right or not, a gambler's choice of odds does no more than report the result of his internal observation of the degree of his first order beliefs, an observation he can make without going on to gamble at all.

This is why there is no point in making people bet in order to determine the degree of their beliefs. Thinking of odds or quotients is just a way of providing a uniform scale for reporting the deliverances of insight. It is like training people to report their feelings of warmth in degrees Celsius, or their

visual sensations of colour in Angstrom units. Such a training has the virtue of sharpening our senses by providing an indefinitely precise and extendable vocabulary for stating our observations in. That such a vocabulary does have this virtue may be only a contingent fact, but it is a very familiar and important one. My ability to discriminate different degrees in one of my beliefs is undoubtedly improved by my habit of thinking what odds I should choose if I had no other reason for my choice. And not only different degrees of the same belief: the vocabulary of odds enables me to compare the strength of beliefs with widely different contents. Without this vocabulary, our perception of these inner states would certainly be much less precise. And because of the causal links between first and second order beliefs (see **9** above), the degrees of our first order beliefs might then well be less precise themselves.

But what is the relevance of betting, if there is no point in actually doing it? How does the vocabulary of betting quotients acquire its virtue? Suppose I am thinking what quotients I should choose if I had no reason for my choice other than my degree of belief. I must still conceive myself to be choosing for a bet, albeit one constrained as specified in **12**. If there were no bet, then even my degree of belief would give me no occasion for choosing one quotient rather than any other. And the occasion the bet supplies is simply one of possible but uncertain gain or loss. If either gain or loss were certain, I should again have no occasion to choose a particular quotient: my insight would be given no basis on which to answer my hypothetical question. So what I must suppose is that neither gain nor loss is certain in the circumstances; which means that the quotients must be supposed to be coherent, and therefore probabilities.

I might of course sincerely produce incoherent betting quotients in reporting the degrees of my beliefs. But if I did, I should not be rightly reporting an irrational state of mind. Irrationality only lies in actually betting at such quotients in situations so constrained as to measure belief, and I am not actually doing that. But it is just because all such quotients would be equally irrational that they make no psychological sense as measures of my particular degree of belief. So the fact that I produce them shows, not that I am being irrational, but that, in introspecting my degrees of belief, I have made some error of measurement. It is as if I had insight also into my mass, and felt myself to be more or less massive than my acceleration under Newtonian forces showed me to be. Newtonian mechanics tells me my introspection must have deceived me; and it will not do to retort that Newtonian mechanics might be wrong. So it is, but it was not that lightly overthrown; nor is subjective probability overthrown by the exactly analogous experiments that Kyburg cites.

Even if the inept design of these experiments had not made them inconclusive, therefore, even as introspections of degrees of belief, they would still have been completely insignificant. But what then would overthrow subjective probability, if not experiments of some such kind? I am not advocating subjective probability *a priori*, so something should be able to show if it is wrong. Well, consider what would show Newtonian mechanics to be wrong. Suppose certain bodies did not always accelerate as Newton requires under all forces. As a matter of fact, they don't exactly, and not only for relativistic reasons. Different motions strip off different numbers of their surface molecules, and carry different amounts of air along with them. So the accelerations produced by different Newtonian forces are not, even at low speeds, related to each other by any one precise value of inertial mass. The utility, and the truth, of Newton's theory consists in the accelerations being almost always related by values lying within some narrow interval (cf. Mellor 1967). If that interval became too wide for useful prediction, the concept of mass would eventually be discarded: Newtonian mechanics would have been falsified. But whilst we retain the concept at all, we retain the laws of motion by which its values, however imprecise, are determined.

And so it is with degrees of belief. Our unselfconscious behaviour no doubt exhibits some quantitative inconsistency. I dare say I should not exhibit precisely the same degree of first order belief in imminent rain (say) in all the diverse unselfconscious actions which some degree of that belief might serve to explain. The best that can truly be said of me is that the degrees of this first order belief which would be exhibited in most of these actions lie within a certain interval of subjective probability. The wider the interval, the less useful the concept, and it is an entirely contingent matter that it has any use at all. My unselfconscious behaviour could be so erratic as to falsify the claim that I had any definite degree whatever of some particular belief. This is likely to be true, for example, of beliefs expressible only in languages I do not understand, such as the mathematical languages of most microphysics. I know I have no determinate degree of belief on those matters, i.e. that my unselfconscious actions are in no way affected by any such state of mind. Consequently, although I could indeed be forced to choose precise odds for a compulsory bet on the truth of some such microphysical proposition, I know it would manifest no insight into the degree of any first order belief I have.

But on matters I understand, my insight does tell me to avoid quotients outside certain intervals. There is a significantly narrow interval of subjective probability values within which I believe, with high second order subjective probability, my first order degree of belief to be confined. On that basis I can

predict, and generally claim to know, what my behaviour would be under quite a wide range of circumstances. Where that is the case, we have the same justification for applying the theory by which degrees of belief may, however imprecisely, be measured as we have in the case of mass for employing Newton's laws of motion.

As a matter of fact, therefore, many of our beliefs have usefully precise degrees. Of those that do, the Dutch book argument shows there to be a probabilistic measure. It is not a matter of rationality, but a matter of a theoretically based scale of measurement. Rationality comes in only later, in considering what degrees people should have of various beliefs in various circumstances. (And even that is mostly a matter not of rationality, but of chance.)

III

14 Parts **I** and **II** have given reasons for thinking that assent is second order belief, that first and second order beliefs come more or less precisely by degrees, and that these degrees are subjective probabilities. I have not proved any of these propositions: I only incline to believe them because they enable us to explain sundry psychological phenomena. But that of course is not enough. I need also to check that their other consequences are equally acceptable.

Some of the other consequences are both familiar and attractive. One is the explanation they provide of how we can fail to believe consequences of our beliefs. The explanation runs briefly as follows. Clearly no one believes any contingent proposition h to degree 1, i.e. no one behaves as if he would risk unlimited loss should h prove false for a penny gain should h prove true. So the qualitative state of full belief can only call for some high degree of belief, short of 1. The degree will no doubt be determined by the context: roughly, the least degree such that, in the context, no higher degree would affect anything the believer would be at all likely to do. The details are both tricky and important, but all that matters here is that the degree of a full belief can be less than 1. For then I can believe two things without believing their conjunction. Let $p(k)$ be the probabilistic degree of my belief in any proposition k, and suppose there are two particular propositions h and i such that

$$(2) \quad p(h) = p(i) = 0.95.$$

Suppose I also believe h and i to be independent, i.e.

$$(3) \quad p(h\&i) = p(h)p(i) < 0.91.$$

Suppose further that for some range of propositions k including h and i,

(4) $Bk \vdash p(k)>n_k$

where $0.91<n_k<0.95$. Then I believe h and believe i, but I do not believe $h\&i$.

As for belief, so for assent, if

(1) $A = BB^*$

is true. Suppose I have virtually perfect insight into my beliefs about h and i, i.e. I believe (2), (3) and (4) to a degree very close to 1. Then I assent to h and to i, but not to $h\&i$.

Given enough more or less independent propositions, I can both assent to each of them individually and dissent from their conjunction, so low can my degree of conscious belief in it be. This is a most attractive result, because it so naturally resolves the paradoxes of the lottery and the preface. I can consciously believe that some ticket will win a lottery while also consciously believing of each ticket that it will not win. I can sincerely apologise in my preface for mistakes which I am sure my book contains, while assenting to every single sentence in it.

So far so good. However, our theories have other, less agreeable consequences. If I believe h to a probabilistic degree x, I must believe its negation, $\sim h$, to degree $1\text{-}x$, and therefore the disjunction, $h\vee\sim h$, to degree 1. That is,

(5) $(\exists x)(p(h)=x) \vdash (p(h\vee\sim h)=1)$.

In other words, however high the degree required for full belief, $p(h\vee\sim h)$ will exceed it. So, whatever the context, I believe $h\vee\sim h$. This is a rather disconcerting result. Recall that I am taking subjective probability to measure my actual degrees of belief, not just those I ought to have. Perhaps I ought to believe $h\vee\sim h$ to degree 1, since as a tautology it cannot be false, so that a bet on its truth, whatever the stakes, runs no risk of being lost. But (5) says that I actually *do* believe $h\vee\sim h$ to this degree, not just that I ought to. Yet it seems possible *prima facie* to doubt even a tautology.

One might on reflection be willing to accept (5), since $h\vee\sim h$ is a very simple and obvious tautology, and yet not to accept that one believed all tautologies, however complex, to degree 1. Unfortunately, it can be proved (see Field 1977) that all tautologies, indeed all theorems of the predicate calculus, have probability 1, and *a fortiori* have subjective probability 1. The same result seems to follow also from the betting considerations of **13**. Quotients that would be irrational in the betting situation specified in **12** cannot be measures of one's actual degree of belief. But in that situation the

only rational betting quotient on any knowably necessary proposition would seem to be 1, since otherwise the opponent can arrange to win as much as he likes on it with no risk of loss. So 1 is indeed the probabilistic degree of my belief in all knowably necessary propositions, however complex.

This conclusion, in (5) at least, is derived from the assumption that I do have some degree of belief in h itself. Now I suggested in **13** that there might be some contingent propositions I have no degree of first order belief in at all, and perhaps I do not believe to degree 1 tautologies whose contingent constituents are all of this kind. But this, even if true, would be small comfort. In particular, it would not mean that imprecision in my degrees of belief could be relied on to prevent me believing tautologies to degree 1. However erratic my behaviour, and hence imprecise my degree of belief in h and therefore in $\sim h$, there will be nothing imprecise about their sum. Uncertainty about the right quotient for a bet on h is not going to make me think I could win a bet on the falsity of $h \vee \sim h$.

Believing at least most necessary truths to degree 1 seems to be an inevitable consequence of our theories. But this seems at second sight a defensible, if not an attractive, result. For it only concerns first order belief, not assent. It does not mean we consciously believe most necessary truths. All it really means is that none of our unselfconscious actions can display, or be explained by, our failing to believe them. Only our speech would do that, and what that displays is lack of assent, not lack of first order belief. First order belief in necessary truth need not affect anything we do; so why not accept it as the innocuous extreme end of our range of quasi-dispositions to action? It does not commit us to being able consciously to recognise every complex instance of necessity that our actions trivially show us to believe.

15 Unfortunately, giving necessary truths subjective probability 1 has further and far less digestible consequences. Suppose I have some degree of belief in contingent propositions h and i such that i follows from h, i.e.

(6) $\quad h \vdash i.$

That is, $h \supset i$ is necessary, so I believe it to degree 1. It follows from the probability calculus that I believe i at least as strongly as I believe h, i.e.

(7) $\quad p(h) \leq p(i).$[1]

[1]For all propositions h and i,
$p(h\&i) \leq p(i)$,
$p(h) = p(h\&i) + p(h\&\sim i)$,

Now suppose p_1 is the subjective probability I need in the circumstances for full belief. Then, whatever p_1 is,

(8) $(p(h)>p_1) \vdash (p(i)>p_1)$.

That is, if I believe h, I must believe i:

(9) $Bh \vdash Bi$.

In other words, I must believe every consequence of each proposition I believe.

This again, although not an attractive result, could be defended in the same way as its predecessor. It is not a matter of assent, only of first order belief construed as a quasi-disposition to unselfconscious action. My actions could of course show that I have changed my mind from believing h at one time to not believing i at another, or display the sort of inconsistency which makes my degrees of belief in h and i imprecise. Granting that, one could argue that no single action could reveal both that I definitely believe something and at the same time that I definitely fail to believe its logical consequences. So (9) might perhaps be accepted, albeit reluctantly and with foreboding.

The foreboding is well justified. Consider our second order beliefs in h and i. 'Bh' and 'Bi' in (9) are of course abbreviations, as explained in **6**, the reference to the common time and believer being left out. The abbreviation is all right because (9) is a necessary truth: it holds of all believers at all times. So it holds in particular of myself at the present time, whatever time that is and whoever I am. That is, it holds of the indexical states B^*h and B^*i:

(10) $B^*h \vdash B^*i$.

That is, $B^*h \supset B^*i$ is necessary, so I believe it to degree 1. It follows that I believe B^*i at least as strongly as I believe B^*h, i.e.

(11) $p(B^*h) \leq p(B^*i)$.

Now suppose p_2 is the subjective probability I need in the circumstances for full second order beliefs. Then, whatever p_2 is,

(12) $(p(B^*h)>p_2) \vdash (p(B^*i)>p_2)$.

---

and $h \supset i \vdash \sim(h\&\sim i)$.

So $p(h \supset i)=1 \vdash p(h\&\sim i)=0$.

So if $h \vdash i$, $p(h) = p(h\&i) \leq p(i)$.

That is, if at any time I believe I now believe h, I must at the same time believe I now believe i:

(13) $BB^*h \vdash BB^*i$.

But that, given (1), entails

(14) $Ah \vdash Ai$.

In other words, if I assent to h, I must assent to i. We are being required to assent simultaneously to every logical consequence of any single proposition we assent to.

This result cannot possibly be accepted. It is a result about conscious belief, not just about unconscious quasi-dispositions to action. No amount of theory could defend it against our consciousness of almost never consciously believing all the consequences of one of our conscious beliefs. So unless there is a flaw in the argument leading to (14), one or other of its premises will have to go. Either assent is not believing one believes, or belief does not come by degrees of subjective probability. Let us survey the options.

16 There are two reasons for trying to keep the theory (1), that assent is believing one believes. One is that without it we shall have no theory of assent; nor, I suggest, of second order belief either. If believing one believes is not assenting, I cannot imagine what it is. But whatever it is, (13) seems to me barely more believable than (14). Believing one believes a proposition does not seem to entail believing one believes all its consequences, since something might be a consequence without one believing one believed it was. And that is the other reason for keeping (1): giving it up generates other problems without solving this one. The real trouble with (14) is (13), not (1).

Perhaps the trouble with (13) is (12), i.e. lies in the theory of subjective probability. If so, it is not that our argument has credited beliefs with implausibly precise degrees. It nowhere depends on two contingent beliefs being of precisely the same strength. So long as my beliefs in h, i and in my first order beliefs in h and i are strong enough to be full beliefs, (12) follows. And imprecision in belief will not generally stop it being full belief. It might for propositions like those referred to in 13, expressible only in languages I barely understand. Perhaps there are some propositions of physics I understand so little I can neither fully believe nor fully disbelieve them. But (14) is no better when h and i are propositions so plain I must be able to believe them: e.g. h, that John was born in 1948; i, that he was born in a leap year. I can easily assent to h without thinking about i at all.

And that is the clue: the trouble with (12) is in fact (11), or rather a condition on which (11) depends, namely that I have some degree of belief in B^*i. But on our theory of assent I have no such degree of belief unless I am (to some extent) consciously believing i to some degree – which usually I am not. Moreover, quite independently of considerations of assent, we must be able to avoid having higher order degrees of belief, however hard it may be to avoid having first order ones. Otherwise there would be a psychologically quite incredible regress of beliefs about one's beliefs about one's beliefs ...

So the condition on which (11) follows from (10) is contingent. (11) follows only if read hypothetically, as relating $p(B^*h)$ and $p(B^*i)$ provided they exist; and the same goes for (12). (13) and (14), on the other hand, are clearly not hypothetical. If a high enough $p(B^*h)$ can exist without any $p(B^*i)$, (13) is false. But a high $p(B^*h)$ can easily exist without any $p(B^*i)$: nothing prevents me thinking of h and not of i. So (13) and (14) do not follow from (11) and (12). They only follow if (11) and (12) are construed categorically, and then (11) and (12) do not follow from (10). So either way (13) and (14) can be rejected without impugning either the theory of assent or the theory of subjective probability.

17 This result is some relief, but not enough. Provided I *am* thinking of h and i together, I do still have to assent to i if I assent to h – basically because if $h \supset i$ is necessary, so will $B^*h \supset B^*i$ be, and I must therefore believe that also to degree 1. Yet I could surely fail to assent to i, even while thinking about it and assenting to h, by not assenting to $h \supset i$, i.e. by failing to assent to a truth of which I am conscious and which is in fact necessary.

The problem therefore boils down to this. The subjective probability of necessary propositions has to be 1. So if subjective probability measures degree of belief, and assent is believing one believes, necessary truths cannot be consciously doubted. Yet obviously they can. I can consciously doubt that $h \supset i$, or that $10^{23}+1$ is a prime number, even if these propositions are necessary. Given this evident possibility, is there any defensible alternative to giving up the theory of assent or the theory of subjective probability?

One reason for seeking an alternative is that, as I remarked in **16**, giving up these theories may generate other problems without really solving this one. Giving up the theory of assent, for example, still leaves us with second order subjective probabilities of 1 in necessary propositions. And I can make no sense of my having to be believing I believe that $10^{23}+1$ is prime even when I am consciously doubting, or even dissenting from, that proposition. So some of subjective probability theory will have to go. But we cannot hope to escape

just by tinkering with the theory, nor by taking it after all to measure rational rather than actual degrees of belief. For instance, it will not help to let the subjective probability of necessary propositions fall a little short of 1: if I can doubt them at all, I can disbelieve them entirely. And if that is possible, I see nothing irrational in believing that $10^{23}+1$ is prime to degree 0.3 – on the grounds, say, that 0.3 is the proportion of primes so far found among other numbers of the form 10^N+1. If subjective probability has to go, I reckon it will all have to go, whether as a theory of actual or merely of rational degrees of belief. That is a high price to pay for solving this problem. Before paying it, therefore, we should make sure both that it would deliver the goods and that we can't get them more cheaply in some other way.

The problem stems from our thinking that there is no way of losing a bet on the truth of a belief which is in fact necessarily true. This is because to think the content of a belief to be a necessary truth is, I take it, to have a conception of truth conditions – e.g. possible worlds – on which a belief with that content would come out true under all such conditions. On that conception there will similarly be no way of both winning a bet on the truth of a belief whose content is h and losing one on the truth of a belief whose content is any logical consequence of h, such as i. If $h \supset i$ *is* necessary, there are no conditions under which it would be false, i.e. all conditions under which h would be true are conditions under which i would be true. This is why the probabilistic degree of belief in necessary truths has to be 1, and that of my belief in any proposition has to be no greater than that of my beliefs in its logical consequences. These incredible subjective probabilities result, therefore, from assuming that the content of a belief in a necessary truth is sufficient to make its truth conditions differ in this way from those of contingent and false beliefs. But is it?

It looks a very innocuous assumption: it is, after all, only a special case of the almost invariable assumption that the contents of beliefs suffice to fix their truth conditions. In other words, if two beliefs have different truth conditions, they have different contents. And this assumption is indeed innocuous enough for contingent beliefs. In particular, the content of a contingent belief can fix its truth conditions and still be the same content whether the belief is true or false, since the same truth conditions will allow it to be either. But this is not so for beliefs that are necessarily true if they are true at all. Consider the belief that $10^{23}+1$ is prime. Suppose I am convinced that this belief is, if true, necessary. That means I am convinced that its truth conditions depend on its truth value in a way in which those of contingent beliefs do not. For, if it is true, its truth conditions are, I believe, such as to forbid its falsity; so if it *is*

false, I must suppose them to be something different. If therefore I thought the content of this belief fixed its truth conditions, I should have to think the content also depended on its truth value. But of course I think no such thing. I do not for a moment suppose that my being uncertain of its truth makes me uncertain of its content. I take the content of the belief that $10^{23}+1$ is prime to be the same whether it turns out true or false. Indeed, what I want to find out, when I am uncertain of it, is just which truth value it, namely the belief state with that definite content, has.

The content of a belief which is, if true, necessary does not, therefore, suffice to make its truth conditions such that it cannot be false. If we use the philosopher's term of art, 'proposition', for a belief content which does determine its truth conditions, we can put this by saying that these beliefs at least are in something less definite than propositions; and perhaps all beliefs are. But however we put it, it follows that these beliefs need not have subjective probability 1. I can have any degree of belief I like in the prospect of winning a bet on $10^{23}+1$ being prime, even while believing that, if I do win, I could not have lost. And the same goes for my degree of belief in $h \supset i$; so my degree of belief in i can be less than my degree of belief in h, even though, if $h \supset i$, $h \vdash i$. Therefore I need not, after all, when I assent to something, assent to all the consequences of it that come to mind.

It is, unfortunately, not easy to say what the content of a belief is, if not a definite proposition, i.e. something that fixes the belief's truth conditions. So the assumption that we believe definite propositions continues to be made *faute de mieux*. Mostly it gives no trouble, especially not with contingent beliefs. But occasionally it delivers spectacular falsehoods, in this case through the medium of our theories of assent and of subjective probability. When that happens, one should take care not to blame the medium for the message. Perhaps a more realistic account of the content of beliefs will produce further objections to our theories; but until it does, I believe we can continue to assent to them with a clear conscience.

4 *How much of the mind is a computer?*

1 How much of the mind is a computer? Computational psychologists, who 'see psychology as the study of the various computational processes whereby mental representations are constructed, organised and transformed' (Boden 1984), say that most if not all of it is. I think they are wrong. Most of the mind is not a computer: most mental processes are not computations.

Before arguing for this conclusion I must distinguish my question from others that may be confused with it, which I will not discuss. One is its converse: how much of a mind do computers have? So far not much, but for different reasons. Maybe in future they will have more, but I have nothing to say about that. Nor about the extent to which psychology should be computable, so that psychologists can use computer models of the mind. That question is not about the mind at all, but about theories of it. A computable theory has no doubt the virtue of being easier to study. But that virtue is not peculiar to theories of the mind, and it doesn't make theories that have it theories of computers: computer models of the mind no more imply that the mind is a computer than computer models of the economy imply that the economy is.

Setting these red herrings aside, what an answer to *my* question tells us about the mind depends on what computers are. On some conceptions of them, it tells us almost nothing. So to give the question a useful sense, I must first distinguish and discard those conceptions. And this will take some time, because it takes time to shake the conviction that we know what computers are because we are familiar with the general purpose programmable electronic digital computers that are their modern paradigms. They are what give our question its interest and its real sense: 'How much of the mind is to be explained in terms of such machines?' And this sense seems clear because we have a largely agreed vocabulary for describing the working and uses of these machines: such terms as 'computation', 'representation', 'information', 'data', 'processing', 'program', 'syntax', 'semantics' and 'algorithm'. This all generates a spurious consensus about what computers are – spurious because different people read key terms in this list, notably 'information' and

'representation', very differently. Those differences in turn affect the meanings of other terms and hence the sense of our question. But before showing how that happens, and identifying the sense the question needs, I must sketch some agreed relations between the relevant terms.

2 First, all parties agree that a computer is defined by what it does, namely compute. It doesn't matter how it computes. When, for instance, one computation needs the results of others, it is immaterial whether the others are done in series (one after the other) or in parallel (together). Nor does it matter whether a computer computes with silicon chips or with brain cells. It need not even compute with matter: a spiritual computer is not a contradiction in terms. Whether mental processes are computations is independent of whether they are material – so fortunately we can set that traditionally vexed question aside.

Next, however computing is done, it is agreed to be the processing of information, as opposed to the processing of matter. A food processor making pâté from its ingredients is processing matter: a word processor making a description of the pâté from a description of its ingredients is processing information. What is the difference?

The main difference is that information is true or false, whereas matter just is. Pieces of information are what I shall call 'propositions', whether they are expressed or embodied in sentences, pictures, computers, beliefs or any other way. A proposition, e.g. *the earth is round*, corresponds to a state of affairs (the earth's being round), which may or may not obtain. If it does (if the earth *is* round), it is a fact, which makes the proposition true; if not, the proposition is false.

To say this is not to assume specific theories of propositions, states of affairs, facts or truth (e.g. a correspondence theory of truth). The only assumptions I make here are those I shall make explicitly. The main one is that propositions exist, or at least that we have attitudes (belief, desire, hope, fear, etc.) of which they are the *prima facie* contents: e.g., my belief that the earth is round. The assumption has been disputed, e.g. by Stich (1983), but like Fodor (1981) and others, I stand by it. I cannot argue for it here, except to say that for the present we clearly need such propositional attitudes to make sense of the mind. Maybe psychology will eventually displace them, but I doubt it. (I doubt especially that computational psychology will: for, as we shall see, computation gets its own sense from the concept of belief.)

Another assumption I make is that besides being true or false, the contents of our beliefs also differ from matter by being abstract, with no location in

space and time, and no causal powers. The earth must be somewhere, and facts about it will have causes and effects (see chapter 12). But propositions about it, whatever they are, are nowhere in particular, and neither affect nor are affected by anything. Yet computers process propositions causally, in definite places and at definite times. They must therefore process them indirectly, by processing *tokens* of them, which do have causes, effects and spatio-temporal location. A proposition therefore is a *type*: again unlike a piece of matter, the very same piece of information may be processed many times, and in any number of places at once. The information output by a single computation will therefore not be a new proposition, but a new token of a proposition, generated by a causal process from input tokens of other propositions.

This causal processing of information we may take to be deterministic. (Exceptions, e.g. outputting a random choice among many solutions to an equation, raise no special issues and only complicate the discussion.) There may be 'noise' in a deterministic computation, but there shouldn't be. The information output by it should be fixed by the information input. That is, it should be a function of it in the sense in which (e.g.) birthdays are a function of people but not *vice versa*: namely, that people – the birthday function's 'arguments' – have only one birthday each – its 'values' for those arguments – whereas many people share the same birthday. (This sense of 'function', since it figures largely in what follows, I must say at once has almost no connection with the concept of a function in biology, anthropology or the philosophy of mind.)

But a computation's 'information function' cannot be given just by its causal processing. Causal processes work only on intrinsic properties of the tokens involved. The properties may be chemical, or electrical, or even mental – e.g. being some kind of pain or visual sensation. The range of intrinsic properties is disputed; but no one thinks it includes being a token of anything like a proposition, say that the earth is round. Causal processing can produce a token with the intrinsic shape of the sentence '*The earth is round*', but not one which intrinsically corresponds to that state of affairs.

What causal processing supplies is a function from intrinsic properties of input tokens to intrinsic properties of output tokens. For this causal function to yield an information function, those properties must be correlated with propositions. The correlation need not be one-to-one: tokens with different intrinsic properties – e.g. upper-case instead of lower-case tokens of 'THE EARTH IS ROUND' – may well be processed in the same way. So the relevant intrinsic properties of tokens need not be a function of the correlated propositions. But the propositions must be a function – a 'semantic function' – of

them if the causal function is to embody an information function. A semantic function, in short, is what makes a causal process a computation. Any computer, i.e. any causal system for processing information, must impose or exploit some suitable semantic function.

There are many semantic functions: the same proposition can be processed by many different systems. Think of the range of tokens of the proposition that the earth is round: all the different sentences to that effect in the world's many languages; all the kinds of globes, photographs and other depictions of the earth that are used *inter alia* to show that it is round. But in a single computation this proposition will be some definite semantic function of one or more of its tokens' intrinsic properties: e.g. of any visually recognisable variant of the printed English sentence '*The earth is round*'. And those properties in turn define a type, any token of which would, given the same semantic function, be *ipso facto* a token of the same proposition. In any computation using that function, any token of that type will represent that state of affairs, the earth's being round.

Representing a state of affairs by intrinsic properties of a token proposition, by which it is causally processed, is the core of computing. But computers represent more than states of affairs. They also represent their constituents: e.g. the earth (or the concept of it), which is a constituent of real or supposed facts about it. They do this by using semantic functions that are compositional, i.e. which represent a state of affairs by a token's structural properties: properties that are spatial or temporal functions of properties of its parts, which in turn represent constituents of the state of affairs. Thus the shape of a printed token of '*The earth is round*' is a spatial function of the shapes of the words in it that represent the earth (or the concept of it) and the property or concept of being round; just as the sound of a spoken token of that sentence is a temporal function of the sounds of its spoken words. These further functions we may call 'syntactic', because collectively they constitute a computing system's grammatical rules of composition, its *syntax*; just as the arguments of those functions – e.g. English words – are its syntactic elements, its vocabulary.

The arguments of syntactic functions naturally include phrases, clauses and whole sentences as well as words: grammar covers sentential as well as subsentential combinations. '*Mars and the earth are round*' is as much a syntactic function of '*Mars is round*' and '*The earth is round*' as they are of their subsentential components. So in particular, a causal function that is a computation is automatically a syntactic function, from the causally relevant types of the input tokens to those of the output token. And the total correlation of a

system's syntactic types (the arguments and values of all its syntactic functions) with what those types represent (states of affairs and their constituents) is the system's *semantics.*

The point of characterising syntax and semantics in this way is of course to stop confining them to natural languages. They become features of any system, physical or mental, digital or analogue, linguistic or pictorial, that processes information. But even so, syntax and semantics remain interdependent. In particular, *pace* Stich (1983), a syntax presupposes a semantics. What makes some of a token's intrinsic properties syntactic is that the semantics of some processing system, of which the token is a causal part, makes some state of affairs a function of those properties. Semantics does not exploit syntactic properties: it creates them. Until a semantics uses it, no intrinsic property, or causal process, is syntactic.

3 So far, I hope, so trite. But what does it take to process information, i.e. to turn a causal process into a computation? That will determine what a computer is, and hence the sense of our question. But this sense must be constrained to make the answer to the question enlightening about the mind. So on the one hand 'information' and related terms must not be so restricted that only minds could process information. Our paradigm, the general purpose programmable electronic digital computer, must not need a mind in order to be a computer. And on the other hand, we must not let too many causal processes count as computations. Otherwise our question will tend toward the trivial: how many mental processes are causal? Probably most of them, whether they are material or not. But that is no news, and does nothing to mark off mental from other causal processes.

This needs saying because the question is easily trivialised in this way, *via* a sense of 'information' which makes some causal functions seem semantic when they aren't; which then makes other causal functions seem syntactic when they aren't. To see how that can happen, consider first a causal function which really is semantic (as of course most are not, just as most semantic functions are not causal): veridical perception, where the very fact that makes a proposition true is what causes us to believe it.

Suppose for instance I see that one object is longer than another ($a>b$) and the same length as a third ($a=c$), as depicted on the left hand side of Figure 1. In the circumstances (the objects are clearly visible, my eyesight is good, etc.) I acquire these beliefs only because they are true: the facts $a>b$ and $a=c$ are necessary causes of my believing them. These perceptual causal processes embody semantic functions: the effects are token beliefs '$a>b$' and '$a=c$' which

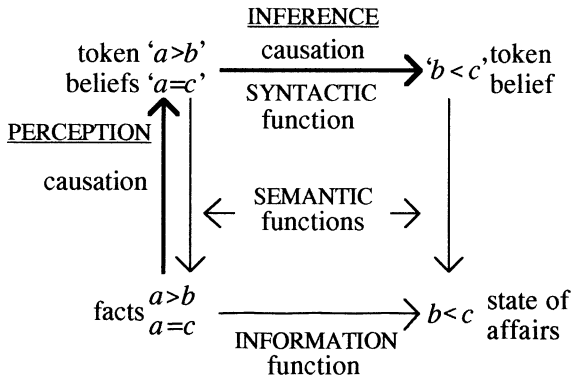


Figure 1: Perception and inference

represent the very facts that cause them. But for precisely this reason, the processes do not embody syntactic functions: they are not computations. In these perceptions, the effects represent their causes, not some function of what their causes represent – if only because their causes ($a>b$ and $a=c$) represent nothing.

But now suppose I infer from these beliefs that $b<c$, as depicted along the top of Figure 1. This too is a causal process, my token beliefs '$a>b$' and '$a=c$' combining to cause a new token belief '$b<c$'. But this process does not embody a semantic function: '$b<c$' does not represent the token beliefs that caused it. What it represents is a function – an information function – of the states of affairs they represent. So the inference, unlike the perceptions, is a computation: its causation does embody a syntactic function – which is made syntactic by the semantic function embodied in its perceptual precursors.

Now compare the situation shown in Figure 1 with one in Newtonian mechanics that has the same causal structure, but where none of the causal functions is either semantic or syntactic. Suppose two forces f_1 and f_2 act in different directions on a mass M, causing accelerations a_1 and a_2 that combine vectorially into a net acceleration a. Any force F acting on M is a well-known function ($F=MA$) of the acceleration A which it causes, and the notional net force f on M is defined to be the same function of M's net acceleration a. (For brevity I am here using 'f' etc. to refer both to token forces and to their strength; and similarly for 'M', 'a', etc.) (See Figure 2.)

Now pretend that $F=MA$ is a semantic function, i.e. that the accelerations a_1 and a_2 represent the forces f_1 and f_2 which cause them, and that the net accel-

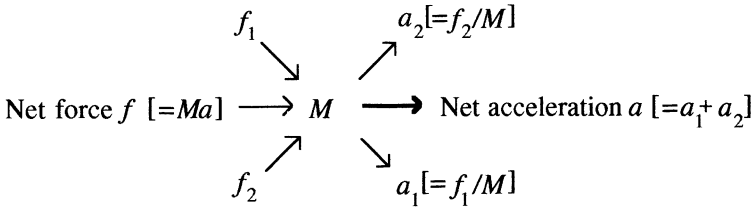


Figure 2: Compositional mechanics

eration a represents the net force f. (Strictly speaking: pretend the magnitudes a_1, a_2 and a of these accelerations define types 'f_1', 'f_2' and 'f' of events representing states of affairs in which forces of magnitude f_1, f_2 and f are acting on M.) This makes the vectorial combination of M's accelerations a syntactic function, representing an information function from applied to net forces. In effect, we are pretending that the mass M 'perceives' the forces f_1 and f_2 and then 'infers' a net force f by using vectorial addition to compute an output token a from the input tokens a_1 and a_2 – as shown in Figure 3.

Figure 3 is of course a fantasy. The accelerations of objects do not represent the forces that cause them, so their vectorial combination is not a computation. But there is a sense of 'information', in which effects embody information about their causes, that is apt to induce this fantasy: the sense used in communication theory (see e.g. Dretske 1981). In it, M's acceleration a_1, in direction d (say) embodies the information that the force acting on M in that direction is f_1, rather than anything else. The information reduces to 1 a greater number of initially possible forces, and the greater the reduction, the more the information. In this sense information is a quantity, increasing with the number of possibilities eliminated. If n equally likely forces might have acted on M in direction d, the acceleration a_1, showing that only f_1 did, embodies $\log_2 n$ 'bits' of information. ('Bits' = 'binary digits', the number of digits in the binary numeral 'n'.)

But 'bits' of information are not pieces of it. The propositions P_1 and P_2, that forces f_1 and f_2 act on M, are pieces of information, however few or many bits they are. But however much information this is, a_1 and a_2 are not tokens of it. They embody it only in a dispositional sense: since they are a causal function of the facts that make P_1 and P_2 true, those facts are what they *would* represent if that function were semantic.

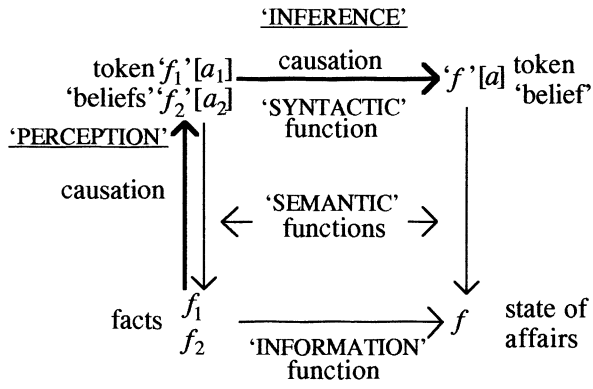


Figure 3: Computational mechanics

That sense is good enough for communication theory, which is concerned with how much information causal links *could* process, whether they actually process it or not. But it is not good enough for us. Our question is whether *M* actually does process information about forces acting on it, not whether it could. And it would only do so if its accelerations represented the forces that caused them, which clearly they do not. Effects do not automatically represent their causes, and nothing else here makes them do so. That is why Figure 3 is fantastic: not because it makes no sense, but because it is plainly false. 'Computational mechanics' *might* be true: it just isn't. a_1 and a_2 do not actually embody any information, and their vectorial combination does not actually process it.

4 We must take computers to embody the information they process in an actual and not merely a dispositional sense. The concept of computation is after all meant to explain how inference and other mental processes differ from the reactions of masses to forces. We must therefore set aside concepts of information, and hence of computation, that are too weak to mark the difference. (And if we think there really is no difference, it would be better not to call minds computers at all than to do so in a sense that puts the priciest product of IBM on a par with a simple mass.)

So when is a causal process a computation? I said in **2** that it is when a semantic function makes the causally relevant properties of its stages syntactic. This implies that the semantics of computation is what generates its syntax. But it is not so obvious what generates its semantics: what makes a stage of a causal

process actually embody information. So it is worth asking if I have got the interdependence of syntax and semantics the right way round. Perhaps syntax comes first after all. Perhaps information is embodied *because* it is processed, rather than *vice versa.* But if so, there should be a criterion for a causal process being syntactic that is independent of semantics. But is there?

There are two *prima facie* candidates. (1) A process is syntactic when it is governed not just by laws but by *rules* (as for instance our use of language is). (2) A process is syntactic when, like a Turing machine, it embodies an algorithm that can take input, and yield output, of unlimited complexity. In fact neither candidate will do, and seeing why not will help to show why computation really must be defined semantically.

Take the rule-governed criterion first. Here of course we must not take it psychologically: the rules must not need to be laid down by people. Nor may we appeal to representations in saying what rules are. The only non-question-begging sense of 'rule-governed' seems to be this: a rule-governed system is one that satisfies a generalisation because a sub-system corrects incipient deviations from it. This is certainly a sense in which our use of language is rule-governed: we remain disposed to use words properly by being disposed to correct each other's misuses of them.

(1) does then rule out the accelerating mass M of **3**, which has no sub-system to correct deviations from the proper vectorial combination of its accelerations. But it still admits far too much. Consider how a floor keeps you level as you walk across it. As you move, gravity makes you start to fall. That in turn makes the floor bend beneath you, until its elastic reaction produces an upward force to match your weight and stop your fall. So according to (1) you and the floor are a rule-governed system: you satisfy a generalisation (that you stay on a level) because a sub-system (the floor) corrects incipient deviations from it.

This is nonsense. It makes the corrections real-time computations, which need to be remarkably fast and accurate: too little, too late or in the wrong place, and you fall onto or through the floor; too soon or too much, and you levitate. But you don't: the floor is a remarkable computer! And so, according to (1), is any rigid object, the elasticity of whose surface will preserve its shape against any random bombardment by air molecules. So indeed is any system with equilibrium states made stable by causal mechanisms. Of course we and our computers are such systems, and would never survive if we were not. But then so are all macroscopic objects. (1) is a hopelessly weak criterion for syntactic processes.

Can (2) do better? Not unless 'algorithm' also has a non-question-begging sense, weaker than its dictionary definition ('method or procedure of computation'). But the only suitable sense seems to be 'function' as I am using the term; and then (2) does not even rule out our mass *M*. For in this sense *M* does embody algorithms, such as the vectorial addition function of its accelerations illustrated in Figure 2. And the function is not limited to the two arguments shown there: it can take unlimitedly complex input – e.g. the impact of any number of air molecules. And *M* also embodies functions with unlimitedly complex output – e.g. functions from the momenta of any number of molecules hitting *M* to their momenta immediately afterwards.

These functions are admittedly not embodied in *M* in the way they would be in Turing machines or other computers. Specifically, they are not programmed into *M*. But we cannot save (2) by limiting algorithms to functions that have been programmed. 'Program' is a semantic term: computer programs *represent* functions, i.e. their tokens represent states of affairs in which the functions' arguments yield their prescribed values. To program a computer is to input a token of the program in order to make the computer embody the very function which the program represents; and to run the program is then to input tokens of the function's arguments, thereby outputting tokens of its corresponding values. In short, to program a causal function is just to generate it from a representation of it. But what it is to do that depends on what it is to represent something: this criterion for a causal function's being syntactic depends on semantics.

And anyway, the criterion doesn't work. Programming a causal function does not make it syntactic, because it does not make its arguments and values represent anything. Suppose for instance we program *M* to combine accelerations vectorially by giving it an engine that accelerates it at a_0 when tokens of a sentence 'Let $a = a_0+a_1+a_2$' are input. a is now whatever function of a_1 and a_2 those token sentences represent: *M's* vectorial combination function has been programmed into it. In particular, the function $a = a_1+a_2$ is programmed into it, by making a representation of that very function turn its engine off. But that does not make *M* a computer: it does not make the combination of a_1 and a_2 syntactic, because it does not make them represent anything, and in particular not the forces that caused them.

In short, although computers can be programmed, that is not what makes them computers. Of course programs are used to induce in computers causal functions that will match their intended semantic functions, and thereby become syntactic counterparts of those functions. And of course every program has a semantics, because it represents the function it induces. But this

is not what makes that function syntactic. What makes it syntactic is that its arguments and values represent something, and the program which induces it is not what makes them do that.

The fact is that (2) is credible only if an algorithm's input and output are already taken to be representations; just as (1) is credible only if the subsystem that makes a system rule-governed is taken to react, not to deviations from a generalisation, but to representations of them. But then what it is to be an algorithm, or to be rule-governed, depends on what it is to represent something, not the other way round. Thus the syntax of computers is after all a product of their semantics. Only when we know where their semantics come from will we know what it takes to be a computer, and how much of the mind may be one.

5 It is in fact obvious, once we think of it, where the semantics of computers come from: they come from us. The computers that have prompted and given sense to our question compute because we compute with them. Computers may in future compute for themselves, but at present they compute for us, and represent what we use them to represent.

When do we compute? First, when we infer, i.e. derive new beliefs from old ones. Whatever computing is, it includes inference; and whatever representations are, they include token beliefs. How beliefs represent states of affairs is another question, which I fortunately need not answer. All I need is the fact that they do. My believing any proposition *P* embodies the information *P*; and any inference made from that belief processes that information. The inference of Figure 1, where my token beliefs '$a>b$' and '$a=c$' cause a token belief '$b<c$', is a paradigm of computation.

To see how this paradigm can bestow its semantics on a computer, let me use one to make the inference of Figure 1, as shown in Figure 4. I give my computer language a compositional semantics, with symbols '*X*', '*Y*' ... representing objects, and '>>', '<>' and '<<' representing the >, = and < relations. That is, to make inferences from what I take to be facts of the form $x>y$, $x=y$ or $x<y$, I input symbols of the form '$X>>Y$', '$X<>Y$' or '$X<<Y$'; and I let output of those forms cause me to acquire the corresponding beliefs.

So now my token beliefs '$a>b$' and '$a=c$' make me input tokens '$A>>B$' and '$A<>C$'. Suppose they are processed in two steps. First one function (F_1) makes them cause an internal token '$C>>B$'; and then another (F_2) makes that cause an output token '$B<<C$'. Finally my seeing the output completes the inference by inducing in me the token belief '$b<c$'.

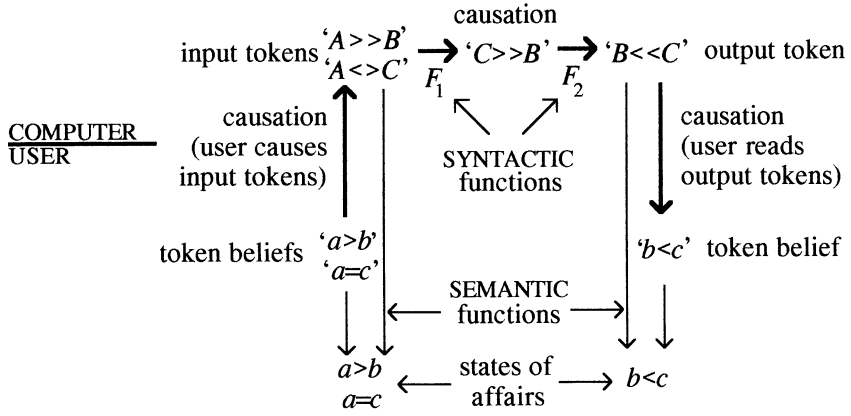


Figure 4: Computer-aided inference

The input and output tokens here are clearly representations, and would be even if I weren't computing with them. My computer language is a language like any other: its tokens are of types that regularly cause, or are caused by, beliefs in propositions they thereby embody, as they do here. '*A*<<*B*', '*A*<>*C*' and '*B*<<*C*' represent the states of affairs *a*>*b*, *a*=*c* and *b*<*c* just as tokens of any other language would do.

And the internal token '*C*>>*B*' represents the state of affairs *c*>*b* (which may or may not be the same as *b*<*c*). Of course, being internal, its intrinsic properties differ from those I directly cause or perceive. But that is immaterial: computers work with electronic rather than with printed tokens only to speed things up. Their syntactic functions are the same whatever intrinsic properties they work with. And though the token '*C*>>*B*' is neither an effect nor a cause of a corresponding belief in me, it is related by syntactic functions to tokens that are. Take F_2, and suppose it turns '*C*>>*B*' into the output '*B*<<*C*' by replacing the token '>>' with a token '<<' and swapping the tokens '*B*' and '*C*'. But these tokens represent *b* and *c* both in the output and in the input that causes '*C*>>*B*', just as '>>' represents > and '<<' represents <. But '*C*>>*B*' might have been the input: I might have used the computer to infer *b*<*c* from *c*>*b*. Then '*C*>>*B*' would certainly represent *c*>*b*. But F_2 has the same semantics whether it is the whole or (as here) only part of a computation. So '*C*>>*B*' represents *c*>*b* anyway.

It also represents it when I use the computer to infer *b*<*c* conditionally, i.e. without believing in it, because I don't believe in *a*>*b* and *a*=*c*. That inference

too is a computation, though the mental token '$b<c$' it causes is not a belief. But it does make me believe that *if* $a>b$ and $a=c$, then $b<c$: i.e. believing those two propositions would now make me believe that one. In short, a conditional inference just disposes me to make the unconditional one: my mental tokens are processed *as if* they were beliefs. The semantics of conditional computation is that of its unconditional counterpart. In particular, F_1 and F_2 will have the same semantics in a conditional inference as in an unconditional one, and so therefore will '$C>>B$'.

Computers are widely used to make conditional inferences (e.g. to predict consequences of alternative economic policies), whose input and output are thus not to be unconditionally believed. This tends to obscure their semantic dependence on unconditional inference and hence on belief. It also explains why their input and output seem sometimes not to represent states of affairs at all. Suppose for instance I make my computer work out squares by programming the x^2 function into it. Suppose I then input a token '5': it outputs '25'. These input and output tokens are numerals, not sentences, representing respectively the numbers 5 and 25. And numbers, whatever they are, are not states of affairs: they correspond to no propositions, true or false. One cannot believe or disbelieve 5 or 25 or any other number.

Yet I do process information when I compute my squares. I am in fact using my computer to make a general conditional inference: if the number of anything is 5, its square will be 25. That inference disposes me, if (e.g.) I believe that the side of a square is 5 cm, to believe also that its area will be 25 cm^2. And if I were not using the computer to make such an inference, its input and output tokens would not even be numerals. On their own they can no more represent numbers than the token '$C>>B$' can represent the state of affairs $c>b$. They, like it, represent only what the inferences they are part of make them represent.

6 All this, however obvious, may well seem to give our question little useful sense. That computers compute because we infer with them tells us more about them than it does about the mind. Calling inference 'computation' makes that aspect of the mind a computer only by definition. Still, there can be informative computational accounts of *how* we infer.

Suppose for instance I make the inference of Figure 1 without using my computer. The process is still causal, only now it is inside me. Suppose again it is a two-step process, as shown in Figure 5, and an unconditional inference, i.e. its 'input' and 'output' tokens are beliefs.

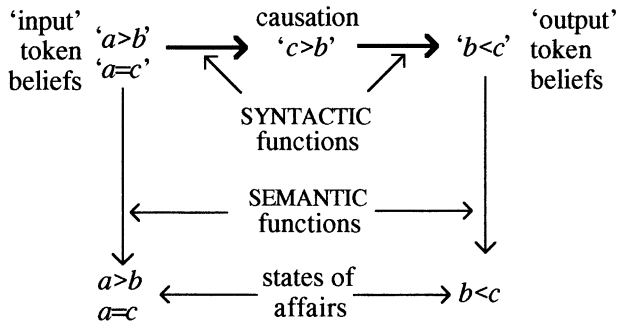


Figure 5: Internal computation

The two steps of the process may be inferences: the intermediate token '*c*>*b*' may be a belief. But it may not be conscious: I may not know how I infer '*b*<*c*'. That gives scope for a computational theory, which may be tested (e.g.) by how fast various inferences are made: roughly, the fewer steps the faster. Thus if this one is as shown, it should be slower than a one-step inference to '*c*>*b*'; and that is a testable prediction.

And such tests do support computational theories, like Johnson-Laird's (1983) theory of syllogistic inference. Johnson-Laird says we do syllogisms using 'mental models' of premises like 'All the *A*s are *B*s', i.e. with mental arrays like

a = *b*
a = *b*
(*b*)
(*b*)

in which token '*a*'s and '*b*'s represent arbitrary *A*s and *B*s, and token '(*b*)'s represent possible *B*s that are not *A*s. Since according to the theory some syllogisms need more models than others, they should take longer; and Johnson-Laird cites evidence that they do.

So there would be scope for computational psychology even if mental computation were restricted to inference. And it isn't, not even within inferences. The two steps of the inference in Figure 5 may be computations without themselves being inferences, because the token '*c*>*b*' may not be a belief: not because these steps are only conditional inferences (they aren't), but because '*c*>*b*' may not combine with desires to cause action in the way beliefs do. If

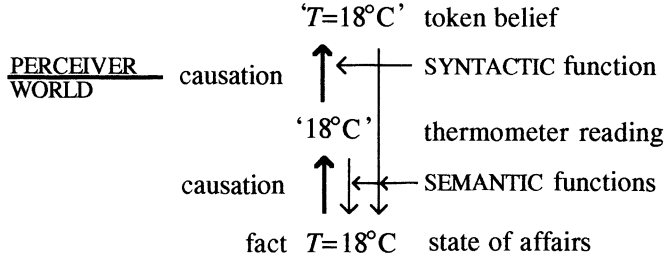


Figure 6: Indirect perception

'*c*>*b*' *is* a belief and (for instance) I want something longer than *b* when only *c* is to hand, it will make me take *c*. But it may not: whatever I want, it may only make me believe *b*<*c*. But then it is not a belief, and wouldn't on its own represent *c*>*b* at all. (Why not is a question I cannot tackle here.) Like my computer token '*C*>>*B*', it represents *c*>*b* only *via* the token beliefs that are its causes and effects. But it does then represent it, as '*C*>>*B*' does. The steps of my inference may not be inferences, but they are still computations. Computation can outstrip inference even when it gets its sense from it: the computational psychology of inference is not a trivial subject.

7 Nor, I admit, is the computational psychology of perception, in which non-inferential computing may also occur, both inside and outside the perceiver. That it occurs outside, in indirect perception, is undeniable. Suppose for instance I perceive a temperature *T* indirectly by using a thermometer, e.g. one with a visible volume *V* of mercury that varies with *T*. (*V* represents *T* by being marked in degrees Celsius, so that it induces beliefs about *T* rather than about *V* itself.) Now suppose in particular it tells me that *T* is 18°C. It does so in two steps (analogous to the steps of the inference in Figure 4). First, *T* makes it read '18°C'; and secondly, my seeing that token causes in me a token belief '*T*=18°C'. (See Figure 6.)

In this process of indirect perception, the thermometer reading '18°C' represents what the belief it causes represents. But that is the very fact, *T*=18°C, which caused it. The cause here is a semantic function of its effect, not a syntactic one. '18°C' represents the fact that *T*=18°C, not a function of anything that fact represents (it represents nothing). The first step of the process is not a computation.

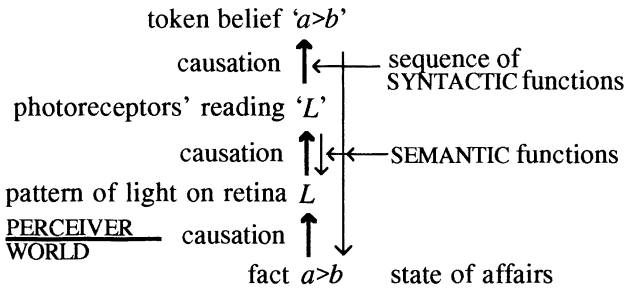


Figure 7: Perceptual computation

The second step is. My belief '*T*=18°C' represents, not the token '18°C' that causes it, but a function (the identity function!) of what that token represents, namely *T*=18°C. My belief is the output of a computation, albeit one embodying a trivial information function (identity).

But perceptual computing need not be trivial. Suppose a photograph tells me that a man has been running (*R*). What the picture, being still, actually represents is his being in a running posture (*U*). But though I react to the depiction of the state of affairs *U*, I may not notice it. I may only notice, i.e. be caused to believe, *R*. I cannot then have inferred *R* from *U*, since I never believed *U*. But I still believe *R* because the picture represents *U*: my token belief '*R*' is computed from the pictorial representation '*U*'.

In both these cases the computing starts outside me. But in perception, as in inference, it may also start inside; and that makes it the province of computational psychology. Thus Marr (1982) for instance says that vision starts with 'arrays of image intensity values, as detected by photoreceptors in the retina' and proceeds 'by mapping from one representation to another' (p. 31) until it causes a belief about 'what is present in the world and where it is' (p. 3).

So according to Marr, when I see that *a*>*b*, light reflected from *a* and *b* causes an array *L* of light intensities to fall on photoreceptors in my retina. Their reaction to *L* he says represents it, and a sequence of computations then transforms this 'viewer-centred' representation (*via* e.g. representations of *a*'s and *b*'s edge-orientations) into the 'object-centred' belief '*a*>*b*' as shown in Figure 7.

Marr's representations start in the retina because that is where, in direct visual perception, systematic links with belief start. The optic nerve embodies definite causal functions, which transmit the semantics of visual beliefs back to some of their causes. (In indirect perception we extend the links outward to

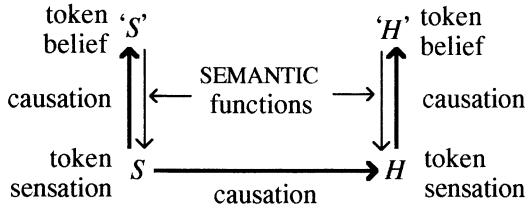


Figure 8: Processing sensations

thermometers, photographs, etc., thus letting them represent states of affairs also. But that is not psychology: the mind starts – if only by definition – at the body's boundary.)

Marr's theory certainly makes sense. If thermometers and photographs can embody propositions by systematically making us believe related ones, so can our sense organs. If doing that can make *T*=18°C's causation of '18°C' semantic, why not *L*'s causation of '*L*'? If it can make the causal function from '*U*' to '*R*' syntactic, why not the function from '*L*' to '*a*>*b*'? After all, the function need not embody an inference: it need not make perception deliver even unconscious beliefs about our retinas. '*L*' need be no more like a belief than a thermometer reading or a photograph is.

And the theory is testable. It postulates specific sequences of computations, from tokens like '*L*' to the beliefs they cause, whose syntactic functions must be embodied causally in the optic nerve. If corresponding causal functions are found, then so far so good; if not, the theory must be altered or abandoned. It is an empirical matter whether, how much and how our eyes – and other sense organs – compute.

How much computing there is in perception is therefore not for me to say: maybe a little, but maybe a lot. And inferring is computing by definition. But there is much more to mental life than perceiving and inferring; and none of the rest of it, I shall now argue, is computation.

8 Many mental processes are not computations because their stages are pains or other sensations, which represent nothing. If for instance a loud aural sensation *S* gives me a headache *H*, that is a mental process. But only the tokens *S* and *H* are processed: no information is, because nothing is represented. True, *S* and *H* are self-intimating (when they occur, they make me believe they occur), so the process produces a parallel sequence of beliefs, which are representations. But that sequence is not a computation either. My

token belief '*H*' (that I have *H*) is not inferred from my belief '*S*' (that I have *S*): no causal process leads from '*S*' to '*H*'. '*S*' is a mere side-effect of *S*. The causal order is not '*S*'→*S*→*H*→'*H*', but that shown in Figure 8.

No causal function in Figure 8 is syntactic. The functions from *S* to '*S*' and *H* to '*H*' are semantic, not syntactic; and the function from *S* to *H* is neither, since *S* and *H* represent neither each other nor anything else. There is no computing in the processing of sensations.

But the mind processes more than sensations on the one hand and beliefs on the other. It processes other propositional attitudes, and they too are representations in their own right. If believing in *P* represents *P*, so do wanting, hoping for and fearing that *P*. And all these attitudes figure in causal processes that depend on the propositions they embody. Suppose for instance I close a door, i.e. make a state of affairs *C* (the door's being closed) obtain, because I want less of a draught (*D*) and believe that if *C* then *D* (*C*→*D*). My action and my desire are tokens respectively of the propositions '*C*' and '*D*', as my belief is of '*C*→*D*'. My action embodies a 'proposition function' of that desire and that belief, just as an inference, such as the inference of Figure 1, does of the beliefs it starts from.

And as in inference, this function must have a syntactic surrogate. It cannot be given just by the causation involved. Causal processes, I noted in **2**, work on intrinsic properties, and embodying a proposition is never an intrinsic property. So my token desire, like my token belief, must have some causally relevant intrinsic property, of which *D* is a semantic function: i.e., a syntactic property. And as for desire, so for every propositional attitude. The causal processing of propositions embodied in tokens of all attitudes must be syntactic. But why then call it 'computation' only when the attitude is belief (or, as mentioned in **5**, is being processed, in a conditional computation, as if it were a belief – a qualification I hereafter take as read)? What makes the processing of belief so special?

The short answer is that processing tokens of other attitudes is not processing information. Pieces of information are all propositions; but not all propositional attitudes embody information. Only beliefs do. Believing in *P* does embody the information that *P*; wanting, hoping for and fearing *P* do not. That is why inferring is processing information, i.e. computing, and the processing of desires, hopes and fears is not.

But why do propositional attitudes only embody information when they are (or entail) beliefs? To see why, consider first why questions and commands do not embody information. The spoken question 'Is the door closed?' ('*C*?') is a token proposition: it represents the state of affairs *C*; but no one would say it

embodied information. Nor would one say that of the command 'Close the door!' ('*C*!'), which also represents *C*. And the reason is obvious: these tokens don't represent *C* as obtaining, i.e. as a fact: they don't embody the proposition '*C*' as true. '*C*?' obviously doesn't, since it is asking whether '*C*' *is* true. And nor does '*C*!', since when '*C*' is true (when the door is closed) the command is redundant.

Information, in short, takes states of affairs as facts: tokens of information embody propositions as true. They may of course not be true: information can be false. But it is meant to be true, and is processed accordingly, even in conditional computing. Information functions are always meant to preserve truth: the reliability of all computing is measured by how rarely it turns true input into false output.

Desires, hopes and fears do not embody propositions as true. A desire for *P* does not take *P* for a fact; and nor does a hope or fear that *P*. That is why those attitudes do not embody the information that *P*. A belief in *P* does take it for a fact, i.e. embody '*P*' as true. To take *P* for a fact just *is* to believe it, whether or not one has any desire, hope, fear or any other attitude towards it. That is the defining characteristic of belief (and why there is the so-called 'paradox of belief' (Moore 1942) – the essential absurdity of saying of any *P*, '*P* [is true] but I don't believe it' – a paradox that has no analogue for any other attitude).

The reason inference and perception aim for truth is that what they deliver is belief. That is why inference is by definition computation, and how inference and perception can include computations that are not inferences. My reaction to light falling on my retina embodies information (if it does) because it is then processed by functions that generally preserve truth in order in the end to yield a true belief.

Other attitudes often do not aim at truth at all, and then there is no need to preserve it. Sometimes falsity is preserved: e.g., in processing desires, which take states of affairs *not* to be facts. (We don't want what we already have – we may want to keep it, but that's another matter.) Take my closing a door because I want less of a draught. The causing of that embodies a general proposition function from '*Q*' and '*P*→*Q*' to '*P*', where '*Q*' is a token desire, '*P*→*Q*' a token belief and *P* the state of affairs brought about by the action they cause. The function is a common one because it always transmits the unwanted falsity of '*Q*' to '*P*', which the action then remedies. It is not (one hopes) so common when '*Q*', '*P*→*Q*' and '*P*' are all beliefs, since it notoriously does not always transmit truth: it is not a good information function.

Conversely, my wanting two states of affairs together, *P&Q*, will often not make me want *P*, because I want only *Q*, or only *P&Q*, because I have *P* already. That is, the function from '*P&Q*' to '*P*' is applied much less often to desires than it is to beliefs partly because, although it always transmits truth, it doesn't always transmit falsity. It is a good information function, but not a good function for desire.

As for desire, so for other attitudes. Proposition functions vary with the attitudes embodied in their arguments and values: functions common for some attitudes will be rare for others. Now to say of what causal functions any attitude's tokens are arguments or values is to say how they are caused, how they interact and how they affect our behaviour. And that, on a functionalist view of mind (see e.g. Block 1980), is enough to define them. But even if it isn't, it still tells us a lot about them, and hence about the mind. So discovering these functions must be a large part of psychology, if it is to explain our propositional attitudes (as opposed to explaining them away). To say what they are, and why we have them, it must say how they affect our processing of propositions, i.e. what their typical proposition functions are, and why we embody those functions.

It should be obvious then that computational theories, restricted to truth-preserving information functions typical only of belief, cannot cope with other propositional attitudes. But there are two ways of making them appear to cope, and I should therefore show explicitly why they do not work.

One way is exemplified by a decision theory model of my action '*C*' above, in which I compute expected utilities (*EU*s) of *C* and not-*C* from my beliefs and desires, and then act to make whichever has the higher *EU* a fact (see Jeffrey 1983). The details don't matter: what matters is the idea that I form token beliefs like 'I want *D*' and 'I believe $C \rightarrow D$' and apply decision theory to them to compute a token 'I should do *C*'. Like the computational mechanics of **3**, the idea makes perfectly good sense: only it's obviously false. I *could* act by using decision theory to work out how to act and then acting that way. But I don't: any more than the mass *M* of Figure 2 uses Newton's laws to work out how to react to applied forces. It simply reacts as the laws say it will. Similarly, my beliefs and desires may well cause me to act as Jeffrey's theory says they do. That doesn't mean I compute my action from them, and I don't.

The objection to such computational theories is not that they are nonsense, or vacuous, but that they are false. Unfortunately in psychology they are also tempting. No one thinks the truth of a computable theory of mechanics makes *M* a computer, because we know *M* can't compute. But we can compute: hence the tempting idea that what makes a computable theory of the mind true is that

we compute with it. And precisely because this might be true, we must not stretch the meaning of 'computation' to make it true by definition. That will only make *M* a computer too, and evacuate real computational theories of the mind of their substantive content.

The other way of making our processing of other propositional attitudes look like computations is to identify them with beliefs in related states of affairs. Thus we might say that wanting *P* is really believing *P* to be desirable; that fearing it is believing it to be fearful; and so on. That is, for each attitude *A* we find – or define – a propositional operator O_A such that $A(P)=B(O_A(P))$ for all *P*, where *B* is belief (and O_B is the identity operator). This transforms all *A*'s proposition functions into corresponding ones typical of belief, i.e. into truth-preserving information functions. We make all our processing of propositions computational in appearance.

But not in reality. Calling a desire for *P* a belief, namely the belief that *P* is desirable, may (by definition) make processing preserve the truth of '*P* is desirable'. But not the truth of '*P*'. The proposition functions in which '*P*' figures as the content of a token of an attitude *A* remain what they always were. And we shall not discover what those functions are by transforming them into the corresponding information functions of '$O_A(P)$'. On the contrary, until we know what the original functions are, we shan't know how to transform them: i.e., we shan't know what O_A is. This way of making the mind look like a computer is a trick: it tells us nothing about the mind.

Both these ways of making theories of all propositional attitudes computational fail, and I see no other way. I conclude that the psychology of attitudes other than belief will never be computational in any serious sense. Computational psychology will remain restricted to the aetiology of belief, especially in perception and inference. That part of the mind may well be a computer; the rest of it, which is most of it, is not.

5 *There is no question of physicalism*

with TIM CRANE

1 What is physicalism? Many philosophers are impressed by the progress achieved by physical sciences. This has had an especially deep effect on their ontological views: it has made many of them physicalists. Physicalists believe that everything is physical: more precisely, that all entities, properties, relations, and facts are those which are studied by physics or other physical sciences. They may not all agree with the spirit of Rutherford's quoted remark that 'there is physics; and there is stamp-collecting' (Longuet-Higgins 1972 p. 16), but they all grant physical science a unique ontological authority: the authority to tell us what there is.

Physicalism is now almost orthodox in much philosophy, notably in much recent philosophy of mind. But although often invoked, it is rarely explicitly defined. It should be. The claim that everything is physical is not as clear as it seems. In this paper, we examine a number of proposed definitions of physicalism and reasons for being a physicalist. We will argue both that physicalism lacks a clear and credible definition, and that in no non-vacuous interpretation is it true.

We are concerned here only with physicalism as a doctrine about the *empirical* world. In particular, it should not be confused with nominalism, the doctrine that there are no universals (*pace* Field 1980 and Stroud 1987 p. 264). Nominalism and physicalism are quite independent doctrines. Believers in universals may as consistently assert as deny that the only properties and relations are those studied by physical science. And nominalists may with equal consistency assert or deny that physical science could provide enough predicates to describe the world. That is the question which concerns physicalists, not whether physical predicates name real universals. (We will for brevity write as if they do, but we do not need that assumption.)

As we will understand it, then, physicalism is not a doctrine about universals or other abstract objects, but about the empirical world, and specifically about minds. It says that mental entities, properties, relations and facts are all really physical. The mental is physicalism's chief target; but one we think it does not hit.

Physicalism is a kind of monism, opposing the dualist's distinction between two kinds of substance: matter and mind. As such, it is descended from materialism: the view that everything is matter – for instance, the view that nothing exists but collections of atoms in the void – as opposed say to Cartesian dualism which held that as well as matter (extended substance) there is also mind (thinking substance). Many physicalists take their doctrine to be a modern version of materialism: defending the hegemony of modern matter against the mysteries of mental substance and of mind/matter interaction.

But physicalism differs significantly from its materialist ancestors. In its seventeenth-century form of mechanism, for instance, materialism was a metaphysical doctrine: it attempted to limit physics *a priori* by requiring matter to be solid, inert, impenetrable and conserved, and to interact deterministically and only on contact. But as it has subsequently developed, physics has shown this conception of matter to be wrong in almost every respect: the 'matter' of modern physics is not all solid, or inert, or impenetrable, or conserved; and it interacts indeterministically and arguably sometimes at a distance. Faced with these discoveries, materialism's modern descendants have – understandably – lost their metaphysical nerve. No longer trying to limit the matter of physics *a priori*, they now take a more subservient attitude: the empirical world, they claim, contains just what a true complete physical science would say it contains.

But this raises two questions. What is physical science: that is, what sciences does it comprise? And what gives it this ontological authority? In other words, what entitles certain sciences to tell us *in their own terms* what the world contains – thereby entitling them to the physicalist's honorific title 'physical'?

'Physical science' so construed certainly includes physics proper. Physics is the paradigm (hence 'physical'). And chemistry, molecular biology and neurophysiology are also indisputably physical sciences. But not psychology, sociology, and economics. One may debate the exact boundary of physical science: but unless some human sciences, of which psychology will be our exemplar, lie beyond its pale, physicalism, as a doctrine about the mind, will be vacuous.

What puts psychology beyond the pale of the physical? Not the *a priori* metaphysics of seventeenth-century materialism, since that has been refuted by physics itself. Nor the materialist's denial of mental substance. Psychology can – and mostly does – deny that too: but it still does not count as a physical science. The question of whether there is 'mental substance' as well as 'physical substance' is an irrelevant one. For that contrast of substances is really a contrast between their characteristic properties: between thinking, say,

and being extended. A merely thinking substance is not a physically respectable entity because thinking is not a physically respectable property. But why not? What, if not the metaphysics of materialism, prevents the empirical psychology of thought, and of other mental phenomena, adding in its own terms, as physics does, to our inventory of what there is?

It is often said that the human sciences have produced fewer results than the admittedly physical sciences. Their laws are said to be few and ill established, and their theories to proliferate, and to predict far less than those of gravity, say, or of molecular biology. Perhaps this paucity of results provides an epistemic basis for denying that psychology and the rest are physical – that is, entitled to tell us what there is. Perhaps they are just not good enough.

But that cannot be why psychology lacks the ontological authority of physics, chemistry and the rest. There are, as we shall see, many well-established psychological laws. And anyway, this epistemic argument is the wrong way round. Those who think psychology is epistemically suspect do so *because* its subject matter is not physical, not *vice versa.* In other words, they have some other basis for taking physical science to exclude psychology, a basis from which psychology's epistemic inferiority is supposed to follow. And it is really quite obvious that this basis is not epistemic. For accepting the results of psychology does not entail accepting them as *physical*: on the contrary, the more such results physicalists accept, the more they reckon they have to explain (or explain away) in non-psychological terms.

The bounds of the physical are in fact set from the outside. Something about the mental is supposed to deprive psychology of the ontological authority of physics and chemistry. But what? What prevents psychology from telling us in its own terms what kinds of mental things and events there are? There are a number of answers to that question: but none, we shall argue, justifies the *prima facie* exclusion of psychology from the realm of the physical which is needed to make physicalism a non-vacuous doctrine about the mind.

2 Reduction to physics To assess physicalists' reasons for dismissing psychology as non-physical, and thus ontologically inconsequential, we must ask what makes them classify their favoured sciences as physical. What makes them count as physical not only the many diverse branches of physics itself (mechanics, electromagnetism, thermodynamics, gravity, and particle physics), but also sciences like chemistry and molecular biology?

One common answer is that these other sciences are physical because they *reduce* to physics, which for present purposes we may take to mean that a

physics enhanced with suitable bridge principles (to link its vocabulary to theirs) would entail credible approximations of all their established laws (see e.g. Hempel 1966 ch. 8).

Some theories in other physical sciences have indeed been reduced to physics in this sense (see Friedman 1981), but by no means all. But those for whom reduction to physics is the touchstone of the physical do not propose to do it in practice. They simply insist that it can be done 'in principle'. But what is the principle? It cannot be physicalism. These sciences cannot be reducible in principle because they are physical if reducibility in principle (RIP) is supposed to tell us where the bounds of the physical lie. So what other principle will tell us which sciences could 'in principle' be reduced to physics?

To answer this, we must first ask to *what* physics the RIP principle is supposed to be applied: to present physics, or to some hypothetical future physics? This question poses a dilemma. For applying the principle to present physics entails that any future extensions of it would not be physical: that physics, the paradigm physical science, is already complete. But no one believes this. And if we apply the principle to an otherwise unspecified future physics, we shall not be able to say which sciences are physical until we know which of them *that* physics must cover – which is just what the principle was supposed to tell us. To use RIP to future physics to say what that physics must cover if it is to cover everything physical is obviously viciously circular. So the physical cannot be defined as what is reducible in principle to physics, either present or future.

We think the RIP principle's specious appeal actually rests on two other prejudices. One is the old dream of the 'unity of science', of being able to derive all scientific laws from one 'ever more adequate grand scheme' (Feigl 1963 p. 315). But we see no reason either to believe in or to aim for such a scheme. The world even of the admittedly physical sciences contains a vast number of very different kinds of entities, properties and facts. That is why so many different sciences, using widely different methods, are needed to study them. No one could think astrophysics and genetics unified even in their methods, except under the most abstract descriptions of scientific methodology. And in their contents, they display no more unity than that of a conjunction. Nothing wrong with that – but then why cannot psychology supply another conjunct?

But even if some 'unity of science' thesis were credible, it would not enable the RIP principle to define the physical. For even physics proper is not unified. Maybe it will be some day; but even if it is not, physicalists will still accept gravity, quantum and electromagnetic phenomena as physical, to be

identified and described in their own terms by independent physical sciences. Similarly for the sciences of chemical, biological, and neurophysiological phenomena. So why not for psychology, the science of mental phenomena?

The other source of the RIP principle's appeal is the idea that there is really no more to things than the smallest particles they are made up of. Let us call this thesis 'microreduction' (Schlesinger 1963 ch. 2), or 'MR' for short. The idea is very persistent. Take Eddington's (1929 pp. xi–xiv) two tables: his commonplace one, with extension, colour, and permanence, versus his 'scientific' one, nothing but myriad minute particles in empty space, the table which 'modern physics has by delicate test and remorseless logic assured me ... is the only one which is really there'. Or more recently, McGinn's (1983 ch. 7) claim that science tells us that the way things are is very different from the way they look. The table that looks and feels so solid is, he thinks physics tells us, really full of holes.

Now the study of the smallest entities is indeed traditionally called 'physics': departments of physics have by long-established custom cornered that particular market. And this makes MR say that the empirical world is physical, since it consists only of its smallest particles. We are back with the doctrine of atoms in the void – or at least, in the field – which count as physical simply because they are microscopic.

The fact that physics by mere convention includes the study of the very small does indeed trivially entail that everything extended in space either is physical or has some physical parts; and for some, this trivial truth is all that physicalism means (see Hellman and Thompson 1975). But for physicalism so defined to be non-vacuous, one must also take these smallest things to be all there is. But what reason is there to think this? Why should we suppose the existence of subatomic particles to require the non-existence of atoms, molecules, tables, trees, or tennis rackets, figs or fast food restaurants – or animals or people with minds?

Proponents of MR can of course distinguish our non-existence from that of, say, unicorns. There are undoubted facts which at least appear to be about us, whereas there are no such facts apparently about unicorns. And of course, since physics itself also studies very large things – galaxies, quasars, etc. – MR is also obliged to say why facts about even these admittedly physical things are different from facts about unicorns on the one hand and facts about subatomic particles on the other. What MR actually does say is that all these facts – about galaxies, minds, and the rest – reduce to facts about their subatomic parts. So those parts are all there is, perhaps because we need not quantify over

anything else in order to state all the facts – and we think with Quine (1953) that we should not multiply entities beyond quantificational necessity.

But this appeal to reduction shows that MR itself needs a strong form of the very RIP principle it is supposed to support. And it cannot have it. For unless the sciences of the relatively large, including psychology, reduce to microphysics, we shall still need to quantify over entities described in those sciences' terms. But in fact, as we shall now show, even the *physics* of the relatively large does not reduce to microphysics. So even if all sciences were reducible in principle to physics, this would not entail that the smallest particles are all there is: MR would be false, even if the RIP principle were true. So the RIP principle cannot be used to support MR.

What is true is that facts about parts often explain facts about wholes. As a thesis about explanation, MR is often a good working hypothesis. But it is not always verified, even in microphysics. If for example we take the quantum mechanical description of a quantum ensemble to be complete (as orthodox interpretations do), the superposition principle entails that its properties will not be a function only of those of its isolated constituents plus relations between them. Orthodox quantum physics is not microreductive. And some physics is positively *macro*reductive: Mach's principle, for example, which makes the inertial mass even of microparticles depend on how matter is distributed throughout the universe. We realise of course that Mach's principle and orthodox quantum theory are controversial, and that a future physics might well abandon them. But they cannot be abandoned because they conflict with an MR entailed by modern microphysics: since, as they show, it entails no such thing.

And even in the most ordinary physics, MR does not always hold. It is indeed usually true that where the parts of something go, the whole thing must go too: that a gas sample must go where its molecules go. But equally, its molecules must go where it goes: since any that do not will thereby cease to be its molecules. And that is not the only way in which a gas's molecules are as much governed by it as it is by them. Suppose for instance that our sample's volume is suddenly halved at a constant temperature. If the gas is ideal, Boyle's law entails that when its pressure settles down again it will be twice what it was. That law does not dictate all the interim behaviour of the sample's molecules – except that it must be such as will eventually double the sample's pressure. That much of their behaviour is determined – and thereby explained – *macro*reductively by a law governing the sample as a whole.

So even as a principle of explanation, MR does not always hold, even in physics. Its explanatory value cannot therefore support it as an ontological

thesis. *A fortiori*, it cannot support physicalism. But it could not even do that if it were true. For no true reading of MR could entail that macroscopic entities, and their properties and relations, are *impugned* by being linked by laws to properties and relations of their smaller parts. They cannot be. For if they were, there would have to be some smallest entities, without parts: that is, a limit to the small-scale structure of matter. But there clearly need be no such limit. So the existence of the currently smallest known particles could not be refuted by the discovery of even smaller ones inside them. But then atoms too must be able to co-exist with their subatomic parts, molecules with their atoms, and so on, up to tables, trees – and us. The existence of animals and people, with their psychological and social properties and relations, cannot be denied just by crediting them with parts small enough to matter to microphysics.

The fact that physics takes in the very small has fostered the myth that it is a universal science in a sense in which others – like psychology – are not universal but merely 'special sciences' (Fodor 1974). It has fostered this myth because it makes everything bigger than a point have (or have parts small enough to have) properties that are physical by mere definition. In a similar way, everything that moves has physical properties, such as inertial mass, just because physics by definition includes the science of mechanics. But that does not make these sciences *universal*, in the sense of encompassing all the properties and relations of things; nor *basic*, in the sense that other sciences must reduce to them. In any sense that would support a non-vacuous definition of the physical, which is what physicalism needs, mechanics and microphysics are no more universal or basic than psychology is. They are merely the special sciences of motion and of the very small.

In short, if the phenomena of psychology are less ontologically acceptable than those of physics and chemistry, it cannot be because psychology is irreducible to present or future physics. Reducibility to physics, or to microphysics, is a hopeless test of the ontological authority of a science: a test which not even a physicalist can apply consistently. For as we have seen, reducibility in practice is neither feasible nor to the point; while those who claim reducibility 'in principle' either beg the question or appeal to principles, of the unity of science or of microreduction, which modern physics itself denies.

3 Mental causation and intentionality How else might physical (that is, ontologically authoritative) science be defined so as to exclude psychology? Perhaps by causation, which many think is essentially physical. Perhaps the physical just is the causal, and what physicalism really means is that the

empirical world comprises all and only those entities, properties, relations, and facts which have causes or effects. This definition clearly underlies one familiar formulation of the mind–body problem: how can mental states have effects in a physical world? This question would not pose such a problem if it were not assumed that causation is essentially non-mental.

But why should we assume this? It is surely obvious that there is plenty of mental causation. Suppose you see a friend, and this causes you to wave to him: how? Something like this: light is reflected from him onto your retina; impulses travel up your optic nerve; your striate cortex processes the information carried by them; you form (somehow) the belief that your friend is there; this makes you form the intention to greet your friend; that makes certain things happen in your motor systems; they cause your arm to rise ... Both physical and mental facts seem equally involved in this chain of causation. How then can a physicalism defined by causation exclude these apparently mental causes and effects?

It is indeed an old thought that mental causation is hard to make sense of, and especially causation linking the mental to the non-mental, because they seem to be so different. But why should that impress anyone who has learned from Hume that causation never 'makes sense': that it is always a matter of fact, not of reason? Nothing in either Humean or other modern analyses of causation forces causes to be like their effects; nor does anything in them stop causes and effects being mental.

Take the requirement that token causes and effects be localised in space and time, so that they can be contiguous (or, if need be, dense or continuous) and so that one can precede the other. Token sensations and even token thoughts can certainly be localised enough for that (since localising a token thought no more localises its unlocalisable abstract content than localising a red object localises the abstract colour red). Nothing about the mental prevents people's token thoughts, feelings, and sensations being wherever and whenever those people are, in order to be where they can have the immediate and therefore contiguous mental and non-mental causes and effects which they clearly appear to have.

Other common demands on causation are also just as easily met by mental as by non-mental causes and effects: for instance, the demand that causes be in the circumstances sufficient for their effects, or necessary (or both); or that they make their effects more probable than they would have been without them; or that causal relations instantiate laws. It is hard to see why any such condition should present any obstacle to the existence of mental causes and effects.

If there is a problem with mental causation, it lies in intentionality, the mind's capacity to represent aspects of the world. And intentionality is indeed often supposed to prevent mental phenomena from being, as such, physical. Thus Fodor (1987 p. 97):

> I suppose that sooner or later the physicists will complete the catalogue they've been compiling of the ultimate and irreducible properties of things. When they do, the likes of *spin*, *charm*, and *charge* will perhaps appear on their list. But *aboutness* surely won't; intentionality simply doesn't go that deep.

But in the previous section, we have already disputed the pretensions of physics to provide all 'the ultimate and irreducible properties of things'. And no one impressed by our arguments will think it matters that intentionality goes less 'deep' in this sense than spin and charge. For many non-mental (e.g. chemical and biological) properties will also not figure on the physicists' list; and if that does not impugn them or the entities they characterise, why should it impugn intentionality or the entities it characterises?

But many philosophers would still agree with Fodor's subsequent comment that 'the deepest motivation for intentional irrealism derives from a certain ontological intuition: that there is no place for intentional categories in a physicalist view of the world'. Thus Field (1978 p. 78) writes: 'Any materialist who takes beliefs and desires at face value ... must show that the relations in question are not irreducibly mental.' If this intuition were correct, and there were independent reasons for accepting his 'physicalist view of the world', then we would indeed have reason to deny the reality of intentionality, and thus much, if not all, of the mental. But as we shall see, the intuition is wrong.

What is the problem of intentionality supposed to be? Intentional states typically have three distinctive features: (i) they seem to be affected by, and to cause actions involving, distant objects or events; (ii) their ascription creates non-extensional contexts – sentences whose truth-value may alter when names or descriptions in them are replaced by others that apply to the same things; and (iii) they can be about objects or events which do not exist. Suppose for example that (i) you read something about Santa Fe that makes you want to go there, which causes you to get on a plane and do so. But (ii) you do not want to go to the most beautiful city in New Mexico, which Santa Fe is, because you do not know that it is. And (iii) you could have wanted to go to Santa Fe even if, like Eldorado, it did not exist.

The challenge which (i) and (iii) present is to explain how Santa Fe can cause you to act as you do when it is so far away, and need not even exist. No one believes that a city can have such effects directly at such distances – especially when it need not exist. Your action must be directly caused by some intrinsic property you actually have, not by your relations to distant and possibly non-existent objects, like Santa Fe, or to abstract ones like the possibly false proposition that it is in New Mexico.

But this does not mean that the causal powers of token thoughts and other mental states cannot depend on their contents: they can. All it means is that they must do so indirectly, *via* a mental representation, i.e. *via* some intrinsic non-relational property of the mental state (or of its owner). A token thought must have some such intrinsic property, correlated somehow with its content, to give it its right causes and effects. An instance of this property is, we may say, the local *causal surrogate* for that content.

But these intrinsic properties could still be mental. They could be sensations, or visual or other mental images or models – which need not, incidentally, be conscious (see Shepard and Cooper 1982, Johnson-Laird 1983). And even if the compositional structure of thought requires these tokens to form a correspondingly complex ('syntactic') structure, they could still be images – like Shepard's shapes composed of images of cubes (Shepard and Cooper 1982 ch. 3). So the problem which thoughts pose for causation is *not* that they are mental and causation is not. It is that causation depends directly only on intrinsic properties, whereas the causal powers of token thoughts depend on their contents, which are not intrinsic. This indeed shows that these contents need causal surrogates: but not because they are mental, since the surrogates could be mental too.

Moreover, the need for causal surrogates is by no means confined to psychology. They are needed throughout physical science. It is, for example, a standard function of physical fields to provide local causal surrogates for what would otherwise be unmediated action at a distance. But no one thinks that accepting Newtonian gravitational fields means denying the physical status, or the existence, of the Newtonian gravity they mediate.

In other parts of physics and chemistry, causal surrogates are needed also to bring about what would otherwise have to be backward causation. We noted earlier that Boyle's law makes the eventual pressure of an ideal gas sample double after its volume is suddenly halved at constant temperature. But that token equilibrium pressure, P, cannot directly affect the non-equilibrium processes which lead to it, since that would need backward causation. Moreover, P, like Santa Fe, need not even exist. The sample's volume may be

altered again before it reaches equilibrium: but this cannot affect its behaviour before that. So the future *P* needs a causal surrogate *in the present* to make the sample head for *P*, just as Santa Fe needs one *in you* to make you head for it.

And as for this case, so for all systems that tend to stable equilibria: from simple pendulums to chemical and biological reactions of all kinds. The Gibbs' potentials of chemical thermodynamics (Denbigh 1955 p. 76), for example, are causal surrogates for the equilibria to which chemical systems tend: equilibria whose existence and physical status they certainly do not impugn, any more than the field mechanisms of Newtonian gravity impugn it, or than the kinetics of gases refutes Boyle's law or shows that halving a gas sample's volume does not really cause the doubling of its pressure. Why therefore should physiological or psychological accounts of how the contents of token thoughts produce their effects contradict them, or the causal explanations they give of our actions?

So much for the alleged problems posed by (i) and (iii) for the ontological authority of intentional psychology. What about (ii), the non-extensionality of ascriptions of intentional mental states? This does not exclude the mental from the physical either, since non-extensionality occurs in physics too (see e.g. Enç 1982). This is because laws entail non-extensional conditionals. Suppose for example that *H* and *K* are the genes that give us hearts and kidneys. The fact that we all have both does not make 'anyone who had gene *H* would have a heart' entail either 'anyone who had gene *K* would have a heart' or 'anyone who had gene *H* would have a kidney'.

The probabilistic laws of modern microphysics cannot be extensional for another reason too, because '$p(\ldots)=n$' is not extensional: for if it were, '*a* is the *F*' and the necessary truth 'p(*a* is *a*)=1' would entail 'p(*a* is the *F*)=1', which it clearly does not, on any view of probability (take for example '*F*'='next Prime Minister').

The non-extensionality of probability incidentally explains that of many singular causal instances of the contexts '... because ...', even in physics. This is because causation gives effects probabilities, if only subjective ones. Probabilistic accounts of causation make that explicit, and it is implicit even in deterministic accounts (see chapter 13). Effects of sufficient causes, for example, have probability 1; and effects of necessary ones would in their absence have probability 0. So '*E* because ...' must be non-extensional, since though *a*'s being the *F* might give '*E*' a contingent probability, *a*'s being *a* cannot. And '... because *C*' cannot be extensional either, because of its counterfactual implications: the probability of a necessary truth like '*a* is *a*' cannot depend on *C*, even if that of a true '*a* is the *F*' does.

These and other reasons convince us, *pace* Davidson and others, that even in physics singular causation never depends on, and mostly is not, an extensional causal relation between particulars (see chapter 12). But if causal contexts can be non-extensional anyway, they can perfectly well contain non-extensional contexts like 'believes ...', 'wants ...', 'fears ...', etc.: as in '*b* fears that *a* is the *F* because *a* told her so' or '*b* does *D* because she wants *a* to be the *F* and believes he will be only if she does *D*'. So we see no reason either to deny the causation which such sentences obviously report, or to suppose that it must be based on, or reduced to, any extensional causal relation – let alone a non-mental one that relates non-mental particulars.

In short, all the supposedly problematic features of intentional states are as endemic to physics, and in particular to non-mental causation, as they are to psychology. The notion of causation will thus not serve to define the physical (and hence ontologically authoritative) sciences in such a way as to exclude psychology. Defining the physical as the causal will not make physicalism a non-vacuous doctrine about the mind.

4 Psychological and psychophysical laws In order for the issue of physicalism to be a serious one, there has to be a principled distinction between the mental and the physical which explains why non-mental sciences have an ontological authority which psychology lacks. We have seen that neither causation nor reducibility to physics can provide such a distinction. But perhaps laws can. The ontological authority of science arguably rests on the laws it discovers, which tell us what kinds of things there are, and what properties and relations distinguish them. But many agree with Davidson (1970b) that the mental is 'anomalous': that strictly speaking there are no psychological or psychophysical laws. If that were so, psychology would add nothing to our ontology of non-mental kinds, with their distinctive non-mental properties and relations.

But why should we deny that, for example, 'All men are mortal' (a true psychophysical generalisation limiting the sentience of members of our species) is a law? There are some bad reasons for denying it, which we shall not consider in detail. One is the idea that laws are necessarily true, which no generalisations about the mental ever are. Thus for McGinn, for mental terms to feature in laws is for 'universal generalisations containing mental terms [to be] metaphysically necessary' (McGinn 1980 p. 187). But, he argues – influenced by the well-known argument of Kripke (1972) against the identity theory – no non-analytic necessarily true generalisations link mental terms either to non-mental or to other mental terms.

Nor they do: but then none links the terms of physics to each other either. The laws of physics are not metaphysically necessary. We agree with Davidson (1970b p. 217) that laws must be 'supported by their instances' and 'support counterfactual and subjunctive claims' ('if x were F it would be G'). But 'All men are mortal' can clearly meet these conditions without being a necessary truth: the fact that something *would not* live for ever if it were human does not mean it *could not*, any more than 'if we went we would go by bus' means we could not go by train. Nor therefore does the fact that the laws of physics meet these conditions show them to be necessary truths; and one of us has argued elsewhere that they are not (see chapter 8). So if terms had to feature in non-analytic necessary laws in order to count as physical, the terms of *physics* would not count, never mind those of psychology.

The law that all Fs are G entails only that anything *would* be G if it were F, not that it *must* be. (And the probabilistic law that all Fs have a chance p of being G, where $0<p<1$, does not even entail that – not even if it is a necessary truth.) The mere possibility of exceptions to psychological and psychophysical generalisations cannot therefore stop them being laws. And even if it could, even if laws did have to be both necessary and deterministic, how would one show without begging the question that a true generalisation really could have exceptions? Not just by imagining them. We can all imagine light going faster in a vacuum than its actual speed c. This does not show that it really could go faster, still less that the constancy of c is not a law. And similarly in psychology. Anyone can imagine brine tasting like port: it can still be a law that to no one with our taste-buds would it ever taste anything like that.

Another bad reason for denying the existence of psychophysical laws is the so-called 'variable realisation' of mental states: the fact that 'the range of physical states fit to realise a given mental state can be indefinitely various' (McGinn 1978 p. 197). That cannot stop psychophysical generalisations being laws. For if it did, there would be hardly any laws in physics either. States like masses, volumes, and temperatures are even more variously realised than mental states: one can have a gram or a litre of almost anything, at any one of an indenumerable infinity of temperatures (see Wilson 1985 p. 235). So if variable realisation does not rule out laws in mechanics and thermodynamics, it can hardly rule them out in psychology.

Nor should we be impressed by the inability of armchair reflection to excogitate psychological or psychophysical generalisations. Physics and chemistry are not excogitatable *a priori*, and we see no reason why psychology should be. It can take as much unobvious theory and experiment to discover the psychophysics of taste, or of vision (Marr 1982), or the unconscious

psychology of inference (Johnson-Laird 1983), as to uncover (say) the biochemistry of reproduction.

So in particular, Stich's (1983 ch. 7) failure to excogitate laws featuring intentional mental states does not mean there are not any. The obvious explanation of our inability to state such laws in simple and exceptionless forms is that our intentional psychology is too complex and (probably) probabilistic. But so is the meteorology of hurricanes, and the quantum mechanics of large molecules. Their laws, for those very reasons, are not statable by us in simple and exceptionless forms. No one infers from this that there are no such laws; and the inference is no better in psychology.

Davidson himself does not use these arguments against the existence of psychological laws. His own argument (1970b pp. 223–5) goes as follows:

(1) There are no strict psycho*physical* laws.
(2) Singular causes and effects must instantiate strict laws.
(3) The mental is not 'a comprehensive closed' system, being affected by the non-mental, which does form such a system.

But by (1) these mental effects cannot instantiate strict psychophysical laws. So

(4) 'there are no strict laws at all on the basis of which we can predict and explain mental phenomena'.

The argument fails at every step. (1) is false, and not only because 'All men are mortal' is a law. There are many more such laws, linking sensations – like pains, smells, tastes, and visual, aural and tactile sensations – to non-mental features of those who have them. There must be, because whole industries depend on them. Think of the laws which must underlie the reliable production and use of anaesthetics, scents, narcotics, sweeteners, coloured paints and lights, loudspeakers, and soft cushions. And if Newton's laws of motion suffice to add masses and forces to our physical ontology, these laws must suffice to add to it the kinds of sensations that feature in them.

But even if there were no such psychophysical laws, this would not undermine the ontological authority of psychology. Even if no laws linked the mental to the non-mental, psychology could still have its own laws, defining its own mental ontology, on a par with that of chemistry (say). For as we saw in **2**, chemistry's ontological authority does not depend on its being reducible to physics via physicochemical laws. Nor therefore can psychology's ontological authority depend on there being psychophysical laws (see Lycan 1981).

But as we have seen, our sensations are in fact subject to psychophysical laws, which themselves suffice to refute Davidson's (1970b p. 212) denial that

'there can be strict laws linking the mental and the [non-mental]', since these laws may very well be 'strict' (i.e. deterministic). Davidson is admittedly more interested in intentional states than in sensations; but the refutation still holds, since sensations are indisputably mental – as Davidson himself admits (p. 211, 1985a p. 246).

So (1) is false. And so is (2): causes and effects need only instantiate probabilistic laws (see chapter 13). But can we not therefore make (2) true – and strengthen (4) – by deleting 'strict' throughout? Indeed we can, and we should: but that will not help Davidson, since it only makes (1) even more incredible.

Nor does admitting probabilistic laws do anything to rescue (3). For whether causation needs strict or merely probabilistic laws, the non-mental no more forms 'a comprehensive closed system' than the mental does.

For what does 'a comprehensive closed system' mean? For Davidson (1970b p. 219), it means a system of 'homonomic' laws, which 'can hope to be precise, explicit and as exceptionless as possible' only because they draw their concepts 'from a comprehensive closed theory'. The non-mental sciences can provide such a theory, Davidson claims; but psychology cannot. Its generalisations are hopelessly 'heteronomic': that is, they 'may give us reason to believe there is a precise law at work, but one that can be stated only by shifting to a different vocabulary'.

But this distinction will not do, since physics itself is full of heteronomic laws. Take Newtonian mechanics, which defines Newtonian concepts of force and mass by saying how they combine to cause acceleration. But the laws of motion that do this do not form a closed theory. Indeed, without some further law relating force to other concepts, they form no testable theory at all. In the theory of Newton's *Principia*, the further law is the inverse square law of gravity. But that theory is not closed either. There are many other kinds of force: electrical, magnetic, viscous, etc. So as a law of net force, Newton's law of gravity is as hopelessly heteronomic as the laws of psychology: it can be made exceptionless only by provisos invoking alien concepts of electricity, etc. (see Cartwright 1983 ch. 3). And similarly for all the other laws of force. All are true only as laws of *kinds* of forces: gravitational, electrical, etc., which combine into net forces by vectorial addition. The theory of Newtonian mechanics is just the conjunction of all such laws, however diverse their other concepts, with Newton's laws of motion.

Newtonian mechanics has of course been superseded, but not because it was only a conjunction: for a conjunction, as we saw in **2**, has all the unity a science needs. So our 'comprehensive closed theory' can also be a simple

conjunction: the conjunction of all true scientific theories and laws. But then to say, as Davidson does, that the non-mental sciences can supply this conjunction on their own is simply to deny the existence of psychological laws: which both begs the question and is refuted by the laws that we know link sensations to their non-mental causes.

So Davidson's argument (1)–(4) quite fails to show that there are no psychophysical or psychological laws. But this does not refute his claim that there are no laws linking *intentional* mental states. And for that claim Davidson gives a special argument, which rests mainly on two connected ideas: the 'holism' of the intentional, and the 'constitutive ideal of rationality'.

The holism of intentional mental states amounts to their being conceptually interdependent, which sensations are not. The belief that *P*, for example, must inhibit the belief that not-*P*, and also the desire that *P* (people do not want what they think they already have). Again, neither belief nor desire can cause action on its own. To do that they must combine, and different combinations can cause the same action: I can say '*P*', for example, either because I believe it and want to speak truly or because I disbelieve it and want to lie. And there is no doubt that such familiar relations between beliefs, desires and actions do partly define them, and thus stop any laws involving them being wholly independent.

But these facts cannot stop there being such laws, because they too have Newtonian parallels. Newtonian force (*f*) and mass (*m*) are also conceptually interdependent, being partly defined by the relation $f=ma$, which stops laws involving them being independent of each other. And this relation too requires forces and masses to combine to produce their effects (accelerations) – and lets many combinations cause the same effect. So we can no more infer a force f or a mass m from the acceleration a they cause than we can infer a belief or a desire from the action they cause. In short, holism alone will not suffice to distinguish the intentional from the non-mental in a way that will show it to be anomalous – as Davidson (1970b p. 221, 1985a p. 248) again admits.

What about Davidson's 'constitutive ideal of rationality'? This is the idea that the relations between beliefs, desires and actions mentioned above partly define or constitute (hence 'constitutive') what it is to be rational. For instance, the fact that the belief that *P* will generally inhibit the belief that not-*P* is one of the holistic truisms that help to define rationality: it is rational not to have obviously contradictory pairs of beliefs. Rationality is an ideal because thinkers can be more or less rational: they can fail to have the totality of their intentional states standing in all these 'rational' relations.

This may all be true: but again it cannot rule out psychological laws since it too has a Newtonian parallel. Indeed everything that Davidson (1970b p. 223) says is peculiar to 'our use of the concepts of belief, desire and the rest' has a Newtonian parallel. Here it is:

> We must stand prepared, as the evidence [of accelerations induced by gravity, electricity, etc.] accumulates, to adjust our theory [of the forces and masses involved] in the light of considerations of cogency [satisfying Newton's laws]: the constitutive ideal of *rationality* [Newton's laws] partly controls each phase in the evolution of what must be an evolving theory. An arbitrary choice of translation scheme [from accelerations to forces] would preclude such opportunistic tempering of theory: put differently, *a right arbitrary choice of a translation manual* would be of a manual acceptable in the light of all possible evidence, and this is *a choice we cannot make.*

We have italicised the two debatable analogies. First, *rationality*, which many think is a normative notion, constraining, for example, what one *ought* to believe. Well, maybe it is, but a belief's rationality may still be a fact about it: for example, something that makes it probably true; with the constitutive ideal simply requiring beliefs to be so related to each other, and to their perceptual causes, that under normal conditions most of them *are* true. And that, far from preventing laws linking the contents of our beliefs to our surroundings and to the non-mental operation of our senses, positively requires there to be some such laws (if only probabilistic ones).

Secondly, the claim that no evidence can enable us to choose a right translation scheme: that is, one which correctly infers beliefs, desires, etc. from their perceptual causes and behavioural effects. But if this is to provide a disanalogy with mechanics, it cannot just mean that no evidence could entail the right theory. That is true in spades in Newtonian mechanics, even if forces are observable, since every ascription of a mass at any time t entails an indenumerable infinity of net accelerations under different net forces at t, none of which entails any other, and only one of which can be actual. How could intentional mental states be more underdetermined by the evidence for them than that?

Davidson (1970b p. 222), however, thinks that Quine's (1960 ch. 2) 'indeterminacy of translation' shows that they must be. He says that the anomalism of the mental 'traces back' to the 'central role of translation' and its indeterminacy. For if there is no determinate translation of sentences, there is

no right statement of what they mean. So, since their meanings are the contents of the beliefs they would express, there is no right statement of those either: that is, beliefs (and *a fortiori* other intentional mental states) have no determinate contents. Contents, like sentence meanings, are not just underdetermined by evidence: they simply do not exist. And if they do not exist, they certainly cannot have instances which feature in laws.

But we deny the indeterminacy of translation, for familiar reasons (see e.g. Kirk 1986 pt III). As we have already seen, and many others have pointed out (e.g. Chomsky 1969), it cannot be entailed by the underdetermination of theories by evidence, or even Newtonian mechanics would have no laws. Nor, without begging the present question, can it be entailed by 'Quine's claim that theories of translation are ... underdetermined even by the totality of truths expressible in terms of physics' (Kirk 1986 p. 136), since psychological and psychophysical laws are *ex hypothesi* not so expressible. To base the indeterminacy of translation on that claim is to base it on what we saw in **2** is an untenable version – the reducibility-to-physics version – of the very physicalism it is being used to support.

We know of no other reason to deny *a priori* the existence of laws involving intentional mental states. But might not the totality of all true non-mental theories be *in fact* so comprehensive and closed as to preclude psychological and psychophysical laws? We do not see how. No amount of physics, for example, can stop mental states instantiating other laws as well. The other laws must of course be consistent with physics – but only because all truths must be consistent with each other. That truism gives no priority to physics, whose laws must equally be consistent with those of psychology.

But perhaps this reading of (3) may look more plausible as a thesis about *causation*, rather than about laws: and Davidson himself (1970b p. 244) suggests that this is how he understands (3) when he says that 'too much happens to affect the mental that is not itself a systematic part of the mental'.

For suppose physics did form a comprehensive causal system: so that laws of physics made each brain state or bodily movement b_2 of yours at any time t_2 be determined by your brain states at an earlier time t_1 (plus non-mental input between t_1 and t_2). How could your mental states between t_1 and t_2 also affect b_2 without violating these deterministic laws?

But now consider a parallel case. Suppose Kepler's laws made the Earth's orbital position p_2 at t_2 be determined by its position p_1 at t_1 (plus its velocity then, and input from space between t_1 and t_2). How, we might equally ask, could the Earth's positions between t_1 and t_2 also affect p_2 without violating Kepler's laws? Yet they must affect p_2 if p_1 does, for p_1 itself comes between

p_2 and still earlier positions p_0 which, given Kepler's laws, also determine p_2. There's nothing special about t_1.

The solution to this puzzle lies in the counterfactual conditional (*C*) which this causal claim entails: if p_1 had been different, so would p_2 – but p_0 would not. In other words, what violates Kepler's laws is only (*C*)'s counterfactual antecedent. (*C*) itself does not violate them, and nor therefore does the causal claim which entails it: indeed Kepler's laws are what make (*C*), and hence the causal claim, true.

Similarly in our original case. 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.

5 Supervenience We have seen that neither laws nor causation deprive psychology of the ontological authority of non-mental sciences. But that still leaves one non-vacuous interpretation of physicalism. The last refuge of the modern physicalist is *supervenience*: the thesis that there is no change or difference without a non-mental change or difference. Two things will never change or differ in any way without also changing or differing in some non-mental way. The physical excludes the mental by being that on which everything else, including the mental, supervenes.

Supervenience is stronger than the trivial claim that everything extended in space has physical parts, but weaker than reductionism, since it says nothing about *which* non-mental difference will accompany any mental one: it does not entail the existence of any psychophysical laws. But it must be stronger than we have so far indicated. For given the multitude of changeable non-mental properties which any thing has (including its spatio-temporal location), all things that change or differ mentally are bound to change or differ in fact in some non-mental respect. So supervenience, to be serious, must mean more than that. The relevant range of non-mental respects must be restricted (at least by excluding spatio-temporal location), and the claim must be at least subjunctive – 'Two things *would* never differ ...' – and arguably even stronger – 'Two things *could* never differ ...'.

However, to give supervenience a run for its money, we will take it as weakly as we can: in its subjunctive form, and with the relevant non-mental respects restricted as little as possible. Even so, we see no reason to believe it. The evidence for it cannot be empirical, since the prospect of ever finding two

things, complex enough to have psychological properties, type-identical in every reasonable non-mental respect, is extremely slight, to say the least. The only remotely plausible argument for supervenience is one which appeals to the causal principle mentioned in **3**, that there is no unmediated action at a distance. This means, as we saw there, that tokens of beliefs and other intentional mental states need intrinsic properties to act as causal surrogates for their contents. And if these properties are all non-mental, and sensations are likewise determined by their non-mental causes, then supervenience may well seem to follow.

But it does not. First, as we have already observed, the intrinsic properties which act as causal surrogates for the contents of token beliefs and other intentional states may very well be mental. And secondly, whether they are mental or not, beliefs will still not supervene on them. For two thinkers could easily have all the same intrinsic properties and still have different beliefs. This is an obvious moral of Putnam's 'Twin Earth' stories (Putnam 1975, Burge 1986): the content of your Twin Earth duplicate's belief that water is wet differs from yours, because his (or her) water is XYZ and yours is H_2O. And similarly for indexical beliefs. If the content of 'That's an elm' includes the tree you look at as you think it, it will differ for two people looking at different trees, even if they have all the same intrinsic properties.

The defender of supervenience might respond that this only shows that thoughts do not supervene on their thinkers' intrinsic properties. They might still supervene on those plus thinkers' non-mental (e.g. spatio-temporal) relations to other things, and those things' non-mental properties (being an elm, or H_2O). But that is not true either, as we can see by considering how thinkers make mistakes. Suppose for example that you and your intrinsically identical twin now look at the same elm, but that although this makes you think it is an elm, it makes him or her think it is an oak. Same intrinsic properties, same relations, same properties of the thing thought about: but different thoughts.

Again, the defender of supervenience might respond that in such a case there would always be some relevant non-mental difference: if not in your eyes, then in how the tree looks from your different viewpoints. But we doubt this. You and your twin might well differ in only mental respects: for example, in your beliefs about what elms look like – beliefs which need supervene on nothing present or non-mental, merely on the different mental effects trees have had on you in the past. And we see no non-question-begging reason to think that those effects must supervene on past non-mental differences.

On the other hand, your and your twin's past experiences do have present effects: they make you think 'That's an elm', and your twin think 'That's an oak'. And being at a temporal distance, they cannot have those effects immediately: their effects must be mediated by some present intrinsic properties of you and of your twin. So perhaps your thoughts must supervene on your intrinsic properties after all? Not so. For not only, as we have seen, may these mediating properties themselves be mental, but even if they are not, they need not differ just because their mental effects do. Causation need not, after all, be deterministic, and modern physics tells us that it often is not. So we have every reason to expect some indeterminism in the causal processes of our perception, our reasoning, and our action: this being one way in which these processes can go wrong and make us make mistakes. But when causation is indeterministic, causes and effects will not supervene on each other. In short, modern physics gives us reason to deny the supervenience of the contents of our token thoughts on even the most extensive list of our other intrinsic and extrinsic properties and relations.

And as for thoughts, so for sensations. Their having non-mental causes or effects will not make them supervenient. On the contrary, if the relevant causation is somewhat indeterministic, sensations *cannot* supervene on their non-mental causes.

Yet again, however, the defenders of supervenience may reply that causation, unlike supervenience, takes time – and we agree (see Mellor 1981 chs 9–10). Causes always precede their effects, whereas token thoughts and sensations are only supposed to supervene on *simultaneous* tokens of non-mental properties. So showing that they do not supervene on their earlier non-mental causes does not directly refute that claim.

But it does refute it indirectly. For suppose an intrinsic non-mental property P causes a mental property M indeterministically. (Say for example that one's chance of being M at t_2 is 0.9 if one has just been P (at t_1), and 0.1 if one has not.) Now suppose that at t_1 many people share *all* their intrinsic non-mental properties, including P. At t_2, therefore, most but not all of them will be M: that is, some pairs of people, atom-for-atom alike at t_1, will differ at t_2 in this mental respect.

Now let a and b be any such pair: at t_2, a is M and b is not. What about a's and b's intrinsic non-mental properties at t_2? Well, these may all be determined by a's and b's shared non-mental state at t_1. But if so, then they too will all be shared, and M will not supervene on them either. But M will not supervene on them anyway. For even if some relevant laws of physics are indeterministic, so that a's and b's state at t_1 does not *make* them share all their

intrinsic non-mental properties at t_2, it still will not *stop* them doing so. On the contrary: given enough such *a*s and *b*s, some will certainly differ mentally at t_2 without differing in any other way.

In other words, modern indeterministic physics must predict that some pairs of people, atom-for-atom alike in all non-mental respects, will differ in some simultaneous mental respects: and will do so precisely because the properties involved are causally related. In short, modern physics suggests that even the weakest serious form of supervenience, which is itself the weakest non-vacuous form of physicalism, is false. And physicalists can surely not expect a physicalism that is falsified by physics to be verified by anything else.

6 The end We have argued that no defensible definition of physicalism will deprive psychology of the ontological status of the non-mental sciences. In no non-vacuous sense is physicalism true. But this does not mean that we want to encourage a revival of Cartesian dualism. On the contrary, our arguments entail that there is no divide between the mental and the non-mental sufficient even to set physicalism up as a serious question, let alone as a serious answer to it. Physicalism is the wrong answer to an essentially trivial question. So it cannot begin to help philosophers of mind answer the serious questions about the mind and, above all, about intentionality: what enables some parts of the world (us) to think about other parts, including other people (and of course ourselves). And to those questions it is quite obvious that neither dualism nor physicalism has anything to contribute. The dualist does not even try to explain intentionality: he just takes it for granted, stipulating it into existence. And saying that minds are all physical no more helps to explain how some physical things can think than saying that all flesh is grass helps to explain the difference between carnivores and vegetarians. This, therefore, should really be the last paper on the subject of physicalism. But we fear it will not be.

6 *In defence of dispositions*

1 Dispositions are as shameful in many eyes as pregnant spinsters used to be – ideally to be explained away, or entitled by a shotgun wedding to take the name of some decently real categorical property. It is time to remove this lingering Victorian prejudice. Dispositions, like unmarried mothers, can manage on their own. They have been traduced, and my object here is to restore their good name.

Raising the status of dispositions matters more than might at first appear. It matters to the propensity theory of chance, which takes the half-life of a radium atom and the increased physiological age of a heavy smoker to be dispositions (Mellor 1971 ch. 5). It matters to the theory that beliefs and other mental states are dispositions (e.g Price 1969 pt II). But it is more politic to argue from the usual paradigms than from such complex and controversial examples.

My strategy will be to show the offending features of dispositions to be either mythical or common to other properties of things; just as loose living is no prerogative of the unmarried and so is no proper basis for discriminating against them. I shall put my case in terms of dispositions being 'real' properties. Nominalists will have to rephrase it (in terms of the intelligible application of dispositional predicates not requiring them to be coextensive with non-dispositional predicates). Nominalism is not the issue here, as Goodman (1965 p. 42) observes; my realist terminology begs no relevant questions.

What then is the force of saying that dispositions are real properties? Take its shape as an archetypally real (and *prima facie* non-dispositional) property of a thing. To call a piece of cardboard 'triangular' seems clearly to say how it is in itself, not merely how it is disposed to behave in this or that situation. (We shall see in the end that this distinction has no force; but let it stand for now.) No one doubts that a triangular thing thereby differs substantially from anything that is not triangular. No one would doubt that a piece of rubber which is triangular at time t_1 and not at time t_2 must in the meantime have changed in some way admitting of causal explanation.

Just so, I say it follows that if just one of two seemingly identical glasses is fragile, there is a substantial difference between them (*pace* Mackie 1972 p. 130) – apart, of course, from their numerical and spatio-temporal differences (a proviso I take as read hereafter). It follows likewise, if a glass is fragile at t_1 and not at t_2, that an event has occurred meanwhile which could have had causes and may have effects (*pace* Levi and Morgenbesser 1964 p. 11).

These claims are commonsensical enough to throw the onus of proof on whoever denies them. They would not need arguing for had they not been argued against. But they have been argued against, and these arguments have not, to my mind, been sufficiently met. So I shall have in what follows to criticise some allied, as well as some opposed, literature.

2 First consider by way of contrast a predicate like 'forty' (or 'mortal'). It applies by virtue of a past (or future) event rather than a present property. To be forty is just to have been born forty years ago. A man of forty need differ in no present respect from a man of thirty-nine; the occurrence of one's fortieth birthday calls for no causal explanation.

Dispositional predicates are not like this. We cannot suppose them to apply now by virtue of the past and future events that could then display them. There is a crucial difference here between 'fragile' and 'forty' or 'mortal'. To be mortal a man has to die in the end come what may, whereas to be fragile a thing only has to break if it is dropped (say). Many fragile things never are dropped and so never break. In that case no relevant future event is available to make just one of two presently identical glasses fragile, whereas the relevant future event (death) is always available to make just one of two presently identical people mortal.

A corollary of this difference is that being forty or mortal now has past or future consequences where being fragile or soluble does not. His past birth being what makes a man forty now, it must have made him thirty ten years ago; similarly, a man who is mortal now is bound to be mortal until he dies. We draw no such consequences from the present ascription of dispositions. A fragile glass may (or may not) be toughened by heat treatment at any time; just as Mrs Lot's present insolubility is in no way impugned by her turning to salt in 2084. Our archetypal dispositions, then, are as mutable as shapes and colours, and as much (but no more) subject to the problem of induction. If they are to be distinguished from properties like shapes at all, it will not be because they are more like age or mortality.

3 Carnap's (1936–7 pp. 52–3) account of dispositions still has a following, despite his notorious failure to accommodate their mutability. He sees that '*Sa*' ('*a* is soluble') entails some such conditional as

(1) If *a* were put in (enough)[1] water it would dissolve,

which amounts to more than

(2) *a* is put in water ⊃ *a* dissolves.

He first strengthens (2) to

(3) *Whenever a* is put in water it dissolves,

that is,

(3) $Sa \equiv (t)[W(a,t) \supset D(a,t)]$,

where '$W(a,t)$' means '*a* is put in water at *t*' and '$D(a,t)$' means '*a* dissolves at *t*'; and it is of course essential to this reading that *a*'s solubility be not a function of time. Carnap deals with (3)'s quite different defect of making everything soluble that is never put in water by replacing it with the 'reduction sentence'

(4) $(t)[W(a,t) \supset (Sa \equiv D(a,t))]$,

which entails nothing about the solubility of such things. But (4) still makes *a*'s solubility (or insolubility) immutable.

Pap (1962 ch. 15.C) undertook a Thermopylaean defence of this consequence of Carnap's account, in which he has been tacitly followed by Quine (1969). He takes dispositions to be determined by essentially immutable natural kinds. A soluble thing is a thing of the same natural kind (for example, sodium chloride) as things that actually dissolve. Now membership of a natural kind may be, while it occurs, necessary; but things can nevertheless become and cease to be members of the most plausible instances of natural kinds. A decaying radium atom changes from being of one radioelement to being of another.[2] Christ at Cana need not have added alcohol, nor destroyed

[1]Usage does not specify how much water is 'enough'; it is vague on *how* soluble a thing must be to be soluble. Similarly with 'fragile' and many other qualitative predicates related to quantitative ones. The issues raised by the vagueness of such predicates are not germane so long as it is realised that 'enough' does not just mean 'enough to dissolve'.

[2]Pap (1962 p. 281) dismisses this blatant counter-example to his thesis as only being 'meaningful in the context of a theory postulating microentities and microevents, whereas what is here under analysis is pretheoretical discourse about things and their kinds, as in qualitative chemistry'. How size could make such a difference to a necessary immutability in things is not

the water and created wine *ex nihilo*; he could just have changed the chemical nature of the selfsame stuff. In any case, sameness of natural kind does not suffice in general for sameness of disposition. Hot water dissolves more of most substances than cold water does; and a glass of water neither changes its natural kind nor loses its identity just by cooling.

Not all authors have seen that immutability is a necessary as well as an impossible price to pay for Carnap's account of dispositions. Storer (1950–1) for instance, reads (4) above as

(5) $(t)[W(a,t) \supset (S(a,t) \equiv D(a,t))]$,

in which a's solubility can vary with time. (5), however, quite fails to express Carnap's intended account of a's present solubility as its dissolving every time it is put in water. What (5) says is plausible enough (and half true; see **11** below), that whenever a is put in water it dissolves if and only if it is then soluble. But that is no news, and meets no objection to dispositions. Carnap's account is not right; but it is not trivial, as (5) is.

Reconciling Carnap with the manifest mutability of dispositions calls for a logic to make

(6) It is true at t_1 that a will dissolve if put in water at t_2

compatible with

(7) a is put in water at t_2 and does not dissolve.

Woolhouse (1973) sees a future for such a logic; I do not – but I suppose if he is right it may yet come to have one.

4 Ryle (1949 ch. 5) has none of Carnap's problem in accommodating the mutability of dispositions. For him they are not, in my sense, real properties at all. 'a is soluble' states no fact; it merely licenses the inference from 'a is put in water' to 'a dissolves' (p. 125):

> Dispositional statements are neither reports of observed or observable states of affairs nor yet reports of unobserved or unobservable states of affairs.

---

remotely clear. How small must a bacterium or virus be, one wonders, before it can turn into something else?

So two seemingly identical glasses of which just one is fragile really are identical while they are not being dropped. No event, with possible causes and effects, makes a glass become, or cease to be, fragile.

Ryle's view is a special case of his 'inference ticket' view of laws of nature. Suppose it is a law that every (suitable)[3] dropping of a certain kind of glass is followed by breakage. Then while any glass *a* is of this (fragile) kind, it follows that if it were so dropped it would break. The inferences licensed by '*a* is fragile' are thus just a subset of those licensed by the law. If an inference ticket is all the latter need be, it is plausibly all the former need be. The greater temptation to speak of '*a* is fragile' as a statement, true or false, no doubt arises from its referring to an individual, *a*, about which it ostensibly says something. There is no harm, for Ryle, in speaking so, provided truth here is accounted for in terms of the reliability of the inference. He is concerned only to deny any extra feature of the world that makes '*a* is fragile' true, over and above those that make true such statements as '*a* is being dropped' and '*a* is breaking'.

Ryle's view of laws has been resisted (e.g. by Ayer 1956 pp. 228–9), to my mind rightly. That does not, however, dispose of his view of dispositions (although it does free us from the need to accept the one as a consequence of the other). Take a stock example of a law, that all *A*-events are followed by *B*-events, where *A* and *B* are independently observable properties of events. Suppose we agree, *pace* Ryle, that the reliability of the inference from 'this event is *A*' to 'the following event is *B*' will not explain, because it can itself be understood only in consequence of, the supposed truth of 'all *A*-events are followed by *B*-events'. We have then to explain wherein the truth of the law does consist, the natural Humean explanation being in terms of the constant conjunction of events having property *A* with events having property *B*. No such explanation is available, however, for the truth of '*a* is fragile at *t*' when *a* is not then being dropped. At *t* there are *ex hypothesi* no droppings and breakings of *a* to be conjoined at all, never mind constantly. And we have already seen that it will not help to invoke the occurrence of events at any other time.

The most this shows is that a glass need not be kept fragile, and different in that respect from other glasses, by events occurring. So what? Equally no events need occur to keep a thing triangular, and different in that respect from a square. The occurrence of events can in no reasonable sense of 'event' be considered necessary to the reality of a thing's properties. Yet Ryle (1949

[3]Cf. 'enough', fn. 1 above.

p. 124) passes repeatedly from such truths as that nothing need be 'going on' when a wire that can conduct electricity is not doing so to the conclusion quoted above.

Ryle makes his argument more specious by tacitly conflating an observable-theoretical distinction with that between events and things. The display of a disposition – say a glass breaking – is an event; hence observable and so admissible into Ryle's behaviourist ontology. The disposition itself, however, is a property a thing may have without the occurrence of any event, hence *not* observable, and so not admissible.

In fact, however, the observable-theoretical distinction divides the class of events as much as it divides the class of things, as Ryle's own examples serve to show. What makes electrical conductivity a less observable property of a thing than its shape is that the relevant display, of current flowing through it, is less observable as an event than (for example) coming to the end of a process of counting a thing's corners.

Ryle thus complicates the special question (for example, how to tell if a glass is fragile) by confounding it with a general question about the empirical basis of relatively theoretical knowledge (whether of events or things). The special question, once distinguished, is easily answered: drop the glass and see what happens, just as one would count the corners of a thing to settle a doubt about its triangularity. But what about the glass when it is not being dropped? Well, what about the supposed triangle when its corners are not being counted? All properties of things are unobserved most of the time; so, come to that, are most events. The problem of induction, such as it is, is neither peculiar to dispositions nor to properties of things (as opposed to events).

The special question about dispositions, if there is one, is not epistemological. It is not how we *know* a glass is fragile when it is not being dropped, but what *makes* it fragile whether it is being dropped or not.

Ryle's view has at least the merit of keeping these two questions more plainly separate. Licensing the inference from '*a* drops' to '*a* breaks' is a matter plainly independent of the truth, and hence of knowledge, of the premise. But what gives the licence its authority? It is obviously not a logical authority and, as Ryle insists, it does not derive from the present occurrence of any event. What then could the authority be if not some present contingent property distinguishing *a* from a glass for which the inference is not licensed?

5 Such indeed is Armstrong's (1968 ch. 6.6, 1973 ch. 2.II) reaction. In his earlier book the authority is some non-dispositional property which provides a 'categorical basis' for applying the dispositional predicate. Thus a fragile glass

differs in (for example) molecular structure from one that is not fragile; a man with a belief differs in some brain state from one without it. In his later book Armstrong admits also 'ultimate potentialities' or an endless regress of dispositions. To the idea that dispositions involve potentialities I return in **9**, and subsequently to the assumption that a disposition, not being a real property in its own right, cannot be an unbased basis for applying other dispositional predicates. Here I consider Armstrong's earlier argument for the necessity of a non-dispositional basis for dispositions.

Armstrong's premise is that unless a disposition has a non-dispositional basis there can be no grounds for ascribing it between its displays. What he calls a 'phenomenalist' (such as Ryle) might admit a contingent correlation with a non-dispositional property and claim to use that to decide if (for example) an undropped glass is fragile. But, says Armstrong, a contingent correlation cannot be shown to hold in the cases for which it is needed. To establish a contingent correlation between fragility and some other property of undropped glasses, one must be able to tell independently whether such glasses are fragile. Yet it is precisely the impossibility of doing this that makes one invoke the correlation in the first place. So it must be necessary, Armstrong concludes, not contingent, that a disposition have some non-dispositional basis.

This argument, however, is no more fatal to Ryle than it is to Armstrong himself. It is necessary, for Armstrong, that a non-dispositional basis *exist*; but that any given property is the basis is quite contingent. And what one needs to distinguish fragile from other undropped glasses is knowledge of *which* property actually is the basis of fragility. But if Armstrong's argument were sound this knowledge would not be available. Armstrong's mistake is his tacit inference from the necessary existence of a correlation to the existence of a necessary correlation. This notoriously does not follow, and anyway contradicts Armstrong's earlier insistence on the contingency of a disposition's categorical basis (for example, the contingency of his mind–brain identity thesis).

Contingent identity statements have been attacked on the grounds that identity statements are, if true, necessary (Kripke 1971). One might hope therefore to salvage Armstrong's argument against Ryle by taking a disposition's identification with its non-dispositional basis to be necessary rather than contingent. These identity statements, however, even if necessary, are not knowable *a priori*; anyone taking this line has to distinguish, as Kripke does, between necessary and *a priori* truths. But Armstrong's argument needs the latter, not the former, its gist being the impossibility of *a posteriori* knowledge of the identity he says is required to make knowledge of

undisplayed dispositions intelligible. The identity's being necessary does not help to make it knowable. The argument in short needs *a priori* knowledge which all parties agree is not to be had.

6 Armstrong's problem, as his premise cited above shows, is merely the problem of induction, which we have observed is not peculiar to dispositions. Goodman (1965 ch. 2) comes to this conclusion too, but his argument *en route* has other, less palatable, consequences.

He essentially assimilates things and events to a homogeneous ontology of temporal 'segments' of things (p. 43). Where I talk of a glass, *a*, being fragile at time *t*, Goodman considers (to adapt his example to ours) the criteria for applying 'fragile' to the entity *a*-at-time-*t*. He considers the unproblematic 'manifest' predicate 'breaks', applying to a subclass of those glass segments that drop. His problem is how to 'project' this predicate, into the wider class of all glass segments, in order to yield the dispositional predicate 'fragile' (p. 44).

On this account the entities that drop are always different from those that are not dropping. Of the former Goodman thinks it sufficient to say whether or not they break; he takes the problem to be what to say about the latter. He and Quine (1960, 1969) answer essentially in terms of their relevant likeness to those dropping glass segments that break. As Quine (1969 p. 16) puts it,[4] we need to 'round out' the set of dropping glass segments that break into a natural kind by adding sufficiently similar glass segments that are not dropping.

I need not go into all the variant details of the Goodman–Quine approach. I merely note two shared, and fatal, consequences. First, no sense is given to fragility unless something breaks at some time. Unless some glasses are dropped and some (but not all) break, there is nothing for others to be relevantly similar to. No doubt the condition is satisfied in fact, but it is surely not necessary to a glass's being fragile.

Quine would doubtless retort that his object is less to analyse our present usage than to provide an adequate and clearer replacement for it. A more serious example then is that of safety precautions at a nuclear power station, based on the fuel's known disposition to explode in circumstances which the precautions are designed to prevent. It is absurd to suppose that these precau-

[4]Quine refers to things rather than to thing segments, but he must either be read this way or be convicted, with Carnap, of imposing immutability on dispositions, without Carnap's excuse that his account requires it: 'We make a set of all the sometime victims, all the things that actually did or will dissolve in water, and then we add just enough other things to round the set out into a kind. This is the water-soluble kind.' See section 3 above.

tions have no basis unless they are somewhere and sometime unsuccessful. Yet only then would the Goodman–Quine surrogate for our fuel's explosive disposition become available.

The second consequence is that projecting a manifest predicate like 'breaks' must be done on the basis of other manifest predicates. We can distinguish fragile from other undropping glass segments only by their sharing a predicate with those dropped glass segments that break; and this predicate, to be any use, must manifestly itself be manifest. As Goodman (1965 p. 45) puts it for his own example, 'we can define "flexible" if we find an auxiliary manifest predicate that is suitably related to "flexes" through "causal" principles or laws'.

Goodman and Quine, like Armstrong, thus require acceptable dispositional predicates to have some suitable non-dispositional basis. They differ, however, in the bases they propose. Quine follows Armstrong, Broad (1925 ch. 10) and many others in citing the fine (e.g. molecular) structure of things that is postulated by scientific theory. These proposals I consider in **8**. Goodman takes a different line (p. 41):

> To find non-dispositional, or manifest, predicates of things we must turn to those describing events – predicates like 'bends', 'breaks', 'burns', 'dissolves', 'looks orange' or 'tests square'. To apply such a predicate is to say that something specific actually happens with respect to the thing in question.

So Goodman supposes that, when a respectable disposition is not being displayed, some suitably characteristic events must be 'going on'. We have seen this (in **4**) to be Ryle's assumption; only he infers from it that dispositions are illusory properties, since evidently nothing need be 'going on' between their displays. And we noted there that this inference would rule out equally the most archetypally non-dispositional properties of things.

What follows, of course, is not that there must, despite all appearances, be enough events for Goodman to base his 'projections' on, but that this mysterious alchemical activity is no more needed for dispositions than it is for the transmutation of elements. Projection may, as Goodman (1965 chs 3–4) later argues, be needed to tackle his 'new problem of induction'. But the new problem of induction is no more our present business than the old one is.

7 I have so far criticised Carnap, Ryle, Armstrong, Goodman, and Quine for failing to accommodate obvious and important facts about dispositions. It is time to see if I can do better.

First, conditionals. I take conditionals – such as (1) above – to be entailed by the ascription of dispositions and so to be statements capable of truth and falsity. The alternative view of them as 'condensed arguments' (Mackie 1962) merely rephrases the question of what makes (1) true in Rylean terms as the question of what makes the argument from '*a* is dropped' to '*a* breaks' a sound one. The answer either way is: some property of *a*; and the Ryle–Mackie formulation makes it no easier to say what the property is. So I continue to talk of conditional statements; those who prefer arguments should paraphrase accordingly.

I put (1) in subjunctive mood and follow custom in referring to these conditionals as 'subjunctive'. If entailed by a dispositional ascription, however, the conditional must shed the counterfactual implications it normally has in English. To say a glass is fragile in no way implies that it is not being dropped. The crucial point is that the entailed conditional is stronger than the material conditional. Whether it happens also to be counterfactual is quite incidental (cf. Ayer 1972 p. 117). With that cautionary remark I continue to use the subjunctive as the most natural English mood. We need only bear in mind that analyses of specifically counterfactual conditionals are not going to give us the truth conditions for dispositional statements; and beware in particular of accepting any account with the absurd consequence that glasses cannot be fragile while they are actually breaking.[5]

Symbolising then our paradigm:

P: glass *a* is (suitably) dropped at *t*
Q: glass *a* breaks at *t*
R: glass *a* is fragile at *t*
S: if glass *a* were (suitably) dropped at *t*, it would break,

the *prima facie* relations between this 'specific'[6] disposition, its associated conditional, and the events that display it, are

(*A*) $P\&R \vdash Q$
(*B*) $P\&{\sim}R \vdash {\sim}Q$
(*C*) $P\&Q \vdash R$ from (*B*)
(*D*) $P\&{\sim}Q \vdash {\sim}R$ from (*A*)

[5]The caution is needed largely because of the literature's excessively epistemological preoccupation with situations in which dispositions are not being displayed and hence the entailed conditionals actually are counterfactual. Goodman, for example, refers exclusively to counterfactual conditionals in his discussion of dispositions.

[6]The term 'specific' is taken from Ryle (1949 p. 118) to signify a disposition that displays itself in just one way. Dispositions with diverse displays he calls 'generic'.

(E) $R \vdash P \supset Q$ from (A)
(F) $R \vdash S$.

That is, a glass's fragility or lack of it apparently both deductively explains and is entailed by the glass breaking or not when (suitably) dropped.[7] In the end (B) and (C) have to go (see **11** below), as does the less compelling but *prima facie* plausible

(G) $S \vdash R$.

On the other hand we do *not* want

(H) $R \vdash \sim P$
(I) $P \supset Q \vdash S$
(J) $P \supset Q \vdash R$ from (I) and (G)
(K) $\sim P \vdash R$ from (J)
(L) $\sim P \vdash \sim R$.

(H) follows from the counterfactual misinterpretation of S; (I) and (J) are offensive because they lead to (K); the objection to (K) and (L) is that they make all undropped glasses alike in respect of fragility.

The present point is that none of (H)–(L) follows from any or all of (A)–(G). Given (E) and $\sim(I)$ – namely, that S is stronger than the material conditional – we are free to suppose that two glasses can differ in fragility whether both are being dropped or neither is. It remains to say, however, what the difference can be.

8 The difference between fragile and other glasses, we have seen, does not consist in the occurrence of events. It might, however, be a difference in some non-dispositional property of the glasses, even though Armstrong's epistemological argument fails to show that it must be.

Recall that doubts about dispositions stem chiefly from their inscrutability when not displayed. This shows in the lack of criteria for a subjunctive conditional's truth when it happens to be counterfactual. Presumably the non-dispositional basis is supposed to provide the required criterion. But how could it? If it does not itself entail subjunctive conditionals relating events (however theory-laden their description), how could observation, which is an event (whatever else it is), be used to assess the ascription of the basis ? (One may decline to solve the problem of induction and still retain *some* empiricist

[7] *Pace* Fetzer (1971), whose desire to assimilate statistical dispositions leads him to deny the general case of (A) and hence even of (E). See Mellor (1981) pp. 68–70 for a critique of this treatment of propensities.

standards.) The stock candidates at any rate all do entail such conditionals, which is no doubt why Popper (1957 p. 70) and Goodman (1965 pp. 40–1) say that *all* properties of things are dispositional. Take the paradigm, molecular structure – a geometrical (for example, triangular) array of inertial masses. To be triangular is at least to be such that if the corners were (correctly)[8] counted the result would be three. Inertial mass entails only subjunctive conditionals specifying acceleration under diverse forces. It is, *pace* Mackie (1972 pp. 148–53), nothing but a 'generic'[9] disposition – that is, a conjunction of specific dispositions.

To ascribe a molecular structure (or any such supposedly non-dispositional property of things) thus at least entails these non-material conditionals, most of which will at any time be counterfactual. How then is this ascription less problematic than that of what we normally call dispositions? It will not help for the ascription to entail *more* than counterfactuals; the truth of the ascription will still depend on whether they are true. As Goodman (1965 p. 45) observes:

> Those who propose to deal with the problem of dispositions by means of classes defined in terms of the microcosmic structure of things often beg the question; for among the dispositional predicates they set out to explain lie the very predicates they need for describing those structures.

Quine (1969) overlooks this point. He thinks scientific theories can give truth conditions for the particular dispositions they explain (for example, for water solubility in terms of chemical composition), thus making the dispositions at once 'respectable and, in principle, superfluous' (p. 20). But if what makes a disposition disreputable is that its truth conditions include the holding of counterfactuals, the truth conditions provided by a theory of it are in no better case.

No doubt there are virtues in explaining properties of things in terms of other properties, especially in terms of those of their spatial parts (see e.g. Schlesinger 1963 ch. 2 on the regulative 'principle of microreduction'). And no doubt also the existence of an explanation is evidence for the truth of the statement being explained. These are certainly reasons for being more prepared to attribute fragility where we can at least sketch a structural theory that accounts for its presence. But none of this gives reason to suppose that

---

[8] 'Correctly' here refers to how the counting is done, not to whether it gives the result three.
[9] See fn. 6 above.

dispositions can or need be made respectable by being given some non-dispositional basis.

9 Yet the feeling persists, and must be exorcised, that dispositions are not quite real. One major cause of it is the curiously widespread view that they are mere potentialities (Armstrong 1973 pp. 13–14) or possibilities. Goodman (1965), for example, calls his chapter on dispositions 'the passing of the possible' and says that 'a flexible object is one *capable* of bending' and that 'the peculiarity of dispositional predicates is that they seem to be applied to things by virtue of *possible* rather than actual occurrences' (pp. 40–2, my italics). A fragile glass, it is felt, is one that *could* break, perhaps in some possible world in which, unlike the real world, it actually is being dropped. But then, as Goodman says, 'possible occurrences are for us no more admissible as unexplained elements than are occult capacities'.

I agree: if mere possibilities distinguished fragile from other glasses, fragility would be no real property, and change in it could be neither cause nor effect. The latter view has indeed been held (Levi and Morgenbesser 1964), but it is evidently false. Consider a rod so twisted that, when put in liquid helium to make it brittle, it breaks. Its becoming brittle is caused by the cooling and in turn causes it to break. Similarly with propensities, the dispositions that display themselves in the chances of events. A man may stop smoking with the intended effect of reducing his risk of death from cancer; an atomic explosion is caused by a change in the propensity of a mass of uranium and other atoms to absorb the products of their own decay.

So dispositions are real properties in a sense that rules out any account of them as mere potentialities or possibilities. But why should any such account have been thought of in the first place? Dispositional ascription entails statements of (admittedly conditional) fact, not statements of possibility. A fragile glass is one that *does* break (*if* dropped), not one that *can* break. Whether it *can* break depends *inter alia* on whether it can be dropped, and its being fragile entails nothing about that. (Bear in mind that it must be physical, not logical, possibility at issue here.) The safety precautions at our nuclear power station (**6** above) are intended to prevent an explosion by making impossible the conditions in which the fuel would explode. It is ridiculous to say that their success robs the fuel of its explosive disposition and thus the precautions of their point.

Of course, in ascribing a disposition we usually entertain the possibility of its display; but to entertain a possibility is not to assert it. Just so we usually resort to subjunctive conditionals only when we think they are counterfactual.

Thus dispositions are usually ascribed only when we regard their displays as possible and not actual. The ascription itself, though, entails neither of these things.[10]

10 Dispositions, then, no more reify possible events than they require actual ones. They do not peculiarly entail subjunctive conditionals, nor are they peculiarly subject to the problem of induction. The problem is not that they are unobservable. Still, there is an objection to them, which must be met.

I take it that the proper rôle of dispositions is to explain their displays: its fragility is what is supposed to explain the breaking of a dropped glass. The objection is to the seeming triviality of the explanation while fragility lacks other criteria than the display it is supposed to explain. Given its mutability, how, for example, is a previously fragile glass to be distinguished from one that just turns fragile upon being dropped?

Explanatory dispositions require some independent basis for their ascription between displays; but the basis need only be another disposition. Thus its mass explains a thing's acceleration under one force only because it is independently testable by its acceleration under other forces. The 'dormitive virtue' of sedatives is made respectable by the maxim that they must also differ in some chemical way from other substances.

To accept its fragility as explaining the breaking of a dropped glass is therefore to interpret *R* as saying that at time *t* glass *a* is *such that* proposition *S* is true. 'Such that' is a 'place-holder' (Levi and Morgenbesser 1964), which must be given independent, but not necessarily non-dispositional, content. A proposed dispositional content can indeed be tested only in the joint display of *S* with whatever other conditional is thereby entailed, but it is no worse off that way than any non-dispositional content. Not all consequences of a hypothesis can be tested all the time (in the case of inertial mass, the proportion testable at any time is infinitesimal), and it is a truism that nothing can be tested when it is not being tested. Induction, it cannot be too often said, is not our problem. What matters is that a glass's fragility be testable at any time by other means than dropping it.

I use here my earlier distinction between properties and predicates. Newtonian mechanics specifies impartially all the specific dispositions, to accelerate under various forces, that the single property of having a certain

[10]Propensities, the statistical special case of dispositions, indeed deal in possibilities. A biased coin only *may* land heads if tossed, the chance of heads being the measure of that physical possibility. This, which is not a point on which deterministic dispositions can be attacked, certainly provides reason to give talk of chance a basis in real properties of the coin – real, but not therefore non-dispositional.

inertial mass supports. The sense of 'fragile', on the other hand, only partly specifies the property that supports it. Its explanatory use is a promissory note drawn on an assumed background of physical theory, which both provides and limits possible complete specifications of that property. Suspicion of such explanatory use is suspicion that the note may not be honoured.

A distinction of properties from predicates calls for some principles of individuation of properties. I need not provide them *a priori*; but I do need to show how they are implicit in scientific usage. Two general principles can be stated, though. The first I have already implied: that a real property must display itself in more ways than one. This is no new idea. Nagel (1969 p. 147), for example, cites the requirement 'to characterise as physically real only things that can be identified in ways other than, and independently of, the procedures used to define those things'.

The other general principle is what Schlesinger (1963 ch. 3) calls the 'principle of connectivity', which I have developed and defended elsewhere (Mellor 1971 p. 120). Here it says roughly that two things cannot differ in just one property, since if they did the difference would be inexplicable. That is no new thought either. Clerk Maxwell's view has been cited that 'if a quantity is connected to other effects which are independently defined then it is a physical state; if not then it is a mere scientific concept' (Turner 1955 p. 231). And apropos of the social sciences (Brown 1970 p. 305):

> In order for us to determine whether a supposed property of any kind is genuine or not, we have to discover whether its existence produces any alteration in certain other properties. ... This ... allows us to distinguish ... between a genuine and a fictitious property.

Connectivity is not a deterministic principle, since it admits and applies to propensities as much as to other properties. Two radioactive atoms that differ in half-life must differ also in some other respect (for example, atomic weight). Nor is connectivity an empirical principle, since we lack independent criteria for what constitutes a single property. If it were it could well prove false: as it is, current theory is shown to be defective if it provides no explanation of a difference between two things. Connectivity is a regulative principle, which gets application only through the networks of law that prescribe and connect explanatory properties. Thus all gas laws prescribe mass, chemical composition, volume, pressure, and temperature to form a closed set of connected properties for samples of any gas. Any difference in mass between two gas samples must be accompanied (and so independently

detectable) by a difference in one of the other properties. Were that ever not so, we should conclude, not that connectivity had been falsified, but that some further connected property of gases had been overlooked.

These principles are both principles of explanation. A property must have independent displays if it is to explain them; it must be connected to other properties if differences in it are to be explained. The identity of a property is shown by its rôle in explanation. Inertial mass, for example, is shown to be a single property by the explanatory rôle Newtonian theory prescribes for it. We *could* construe a thing's inertial mass as a mere conjunction of two properties, one displayed in accelerations under forces up to (say) 1 newton, the other displayed in accelerations under greater forces. Each would have plenty of independent displays as far as our first principle goes; and any difference in one of these properties is always accompanied by a difference in the other. But we do not admit that one such difference explains the other; Newtonian theory prescribes in effect that each merely manifests one underlying difference, in inertial mass, which itself therefore calls for further explanation. We have merely independent predicates supported by the same property, just as '> 20 cm' and '< 30 cm' are alike supported by the property of being 25 cm long.

What goes for difference *a fortiori* goes for change. The identity of a property is shown in that of the events that are constituted by changes in it, which may in turn be identified by their causes and effects (Prior 1962, Davidson 1969).

It is theory that usually prescribes the changes and differences in properties that may explain and call for explanation by others, and so in effect provides their identity criteria. So as theory changes, what counts as a property may change; and there may, at any time, be borderline cases. Does the variation of water's boiling point with pressure display one or many properties? When theory prescribes the form of the variation for a variety of substances, leaving measurements at different pressures to be indifferently combined in estimates of the parameters that characterise water, the answer will be clear. Water boiling at atmospheric pressure will be merely one of many alternative displays of a more general property; just as the stretching of a wire under a particular load is merely one of many alternative displays of its elasticity.

So it is for physical theory, not for me, to identify the property that *inter alia* supports 'fragile'. 'Mere' dispositions are those for which this job has not yet been done. But the distinction is not well put as that between 'dispositions' and 'real properties', since none but dispositional predicates need be used to do

the job. No doubt common usage makes some distinction of dispositional from other predicates of things, but it is not to any present purpose.

11 A property's identity is of course related to how specific it is. Our first principle says in effect that a property cannot be totally specific – that is, cannot display itself in just one way. But how specific a property is depends on how diversely its displays are described. A thing's inertial mass is generic to its diverse accelerations under diverse forces; it is specific to the constant ratio of these quantities. But it is acceleration we want explained, so inertial mass is generic enough for the theory that makes it a single property to be acceptable. Relativity identifies it further with gravitational mass, thus making inertia merely one manifestation of a yet more generic property.

Conversely, we might discriminate various kinds among the displays of a seemingly specific disposition; for example, how many pieces a fragile glass breaks into when dropped. There might be n ways of dropping a fragile glass, the ith way always giving $i+1$ pieces. Relative to these more specifically described displays, fragility would be a generic property, providing as good an explanation of them (if not of breaking *per se*) as inertial mass does of a particular acceleration.

Moreover, suitable relative frequencies of breakages of fragile glasses into 2, 3, ... $n+1$ pieces might prompt statistical explanation. If the number of pieces correlated with ways of dropping the glass, an appropriate propensity would be ascribed to the dropping device; if not, to the glass. The display in either case is the chance distribution over the n possible outcomes of dropping the glass (*pace* Fetzer 1971; see Mellor 1971 ch. 4). The immediately relevant point is that propensities are subject to the same constraints as any other dispositions; one property indeed may have some deterministic and some statistical displays, as does the nuclear structure of a radioactive atom, or as the temperature of a gas does according to statistical mechanics.

We are now in a position to see how much of (*A*)–(*G*) from **6** survives on this account. (*A*), (*D*), and (*E*) survive. We lose (*G*) (which is no great matter), since there is after all more to fragility than breaking when dropped. (*B*) and (*C*) come out false, since a glass's breaking when dropped cannot itself entail the truth of any other conditional and hence cannot entail that the glass is fragile. Yet (*B*) and (*C*) have intuitive appeal that needs accounting for: it is very plausible to say that a glass's breaking when dropped shows it to be fragile.

The fact is that we normally presuppose the situation to be deterministic – that is, such that all suitable droppings result in breakage (or none do). But it

is after all possible that the situation is a chance one, that the glass has merely some propensity to break when dropped in the specified way. Then there would be a finite, and perhaps large, chance of the glass's not breaking when dropped, so that it is not, in any deterministic sense, fragile. The glass breaking on a particular drop is, of course, compatible with this hypothesis, and so fails to entail the fragility of the glass. The two hypotheses may be distinguished – for example, by the change the deterministic one must postulate in the glass between a drop on which it does not, and a drop on which it does, break; by discovering such changes we may come to have reason enough to eliminate the statistical hypothesis. At any rate, once it is ruled out, together with the further possibility that the situation is not governed by law at all, the inference expressed in (*B*) and (*C*) is restored. Now we are supposing a two-member family of deterministic properties – being fragile and being tough. The glass's breaking shows it is not tough and hence, by elimination, that it is fragile.

12 Finally, I briefly consider Mackie's (1972 ch. 4) criticisms of the above account. He observes that purely dispositional properties must both entail their display and yet be 'intrinsic' properties of the things concerned. Against this he brings two main objections. The first is that we are looking for a causal explanation of the glass's breaking when dropped. This a non-dispositional basis, such as molecular structure, can provide, precisely because it does not entail that the glass breaks when dropped. The connection between fragility and breakage, however, is logical, hence not causal and so, ultimately, not explanatory. The short answer to this objection is that only events need causes, and the cause of the glass's breaking is its being dropped. It is the glass's being fragile that makes dropping it the cause of its breaking; and for the hypothesis that the glass is fragile our account provides the possibility of independent evidence. What we have is an archetypal deductive explanation on the Hempelian model. *That the glass would break if dropped*, on the other hand, is just not an event at all, and so needs no cause.

Another way of putting this first objection is to claim that its non-dispositional basis makes fragility redundant. So Mackie asks (p. 137): 'Why should we insert this extra element between the non-dispositional basis and the causal behaviour?' I claim, of course, that this is a wife-beating question: fragility *has* no non-dispositional basis in the sense required.

Mackie secondly objects (p. 137) that

> the suggestion that there are such properties is in open conflict with Hume's principle that there can be no logical connections

> between distinct existences. For if fragility in this sense were an intrinsic property of the glass, then it, the being struck, and the breaking would all be distinct existences, and yet on this view the conjunction of the first two would entail the third.

To this I reply that I do not think Hume's principle applies unrestrictedly to our mixed ontology of things and events, although it may well apply within each category. The dropping of the glass and its breaking are both events, and no logical connection holds between them. But the glass's being fragile at that time is not an event: it is a property of a thing.

It is already plain that logical connections obtain between properties of things and events. Prior (1962) suggests that all events are merely changes in the properties and relations of things, and certainly many are. So no oddity need be seen in the logical connections I have been proposing between properties and the events that display them. Nor is there vicious circularity involved. It is not a matter of attempting a general definition of the notion of event in terms of that of a thing and then defining things in terms of events. We have rather a heterogeneous ontology, from which neither things nor events can plausibly be expelled, which yet sustain certain logical connections between them.

Nor is there circularity in the characterising of specific properties and events. Events are defined by changes in other properties than they themselves display. (It is notorious that one cannot prevent one's weight increasing merely by not weighing oneself.) So some properties presuppose others in which their display constitutes a change. Thus flexibility entails, since it is displayed in changes of, shape. That in itself is no stranger than comparable facts about events – for example, that a murder must also be a death. A theory might indeed identify the flexibility of a crystal (for example) with a molecular structure that also determines its normal shape, so that the two become mere manifestations of one underlying property. But whether this happens or not, the conceptual dependence of flexibility on shape illustrates no general dependence of dispositional on some special class of non-dispositional properties.

There is of course much more to be said about the relation of things to events. But that is matter for another time. It is enough for the present that nothing which may truly be said invalidates the present account.

7 *Natural kinds*

1 Some notable philosophers have recently used new arguments to revive essentialism, and have prescribed their essences for a variety of metaphysical fears and ailments. The essence of a self has been said to guarantee its ancestry; mental essence has been promoted as a sure defence against materialism; and diamonds have been warranted in all possible worlds against being paste. I mistrust these prescriptions, especially the claims made for their active ingredients: possible worlds and necessary identity. However, I don't mean here to resist all applications of these notions, nor to dispute all forms of essentialism. I mean only to supply an antidote to the natural kind essences widely advertised by Professors Kripke (1971, 1972) and Putnam (1975).

Kripke and Putnam claim that natural kinds have essential properties; that is, properties which nothing can lack and still be of the kind. The kinds involved include the traditional natural kinds: elements and compounds like gold and water, and biological and botanical species like tigers and elm trees. Modern essences, however, come in a wider range. Essential properties are claimed also, for example, for temperatures and lengths.[1] These do not traditionally form natural kinds, but it will be convenient here to stretch the term to match the doctrine.

Properties alleged to be essential typically involve the microstructure of things. Having atomic number 79 is said to be the essential property of gold (Kripke 1972 p. 327), being H_2O the essential property of water (Putnam 1975 p. 233). Genetic makeup similarly provides essential properties for animals and plants,[2] and mean molecular kinetic energy for temperature. These essential properties of natural kinds are supplied by the natural sciences: 'In general, science attempts, by investigating basic structural traits, to find the

[1] 'It is going to be *necessary* that heat is the motion of molecules' (Kripke 1971 p. 160). The proposition 'The standard meter rod (*S*) is 1 meter long' Kripke (1972 p. 275) takes to be a contingent *a priori* truth – *a priori* because *S* sets the standard, contingent because *S* might have been shorter, longer or non-existent. But if *S* in fact fixed the extension of '1 metre' as Putnam prescribes (see below), then whatever shared properties in fact make distant objects the same length as *S* would be essential properties of being 1 metre long.

[2] '... animals with the appearance of cats but reptilic internal structure ... would not be cats; but "fools' cats"' (Kripke 1972 p.321).

nature, and thus the essence (in the philosophical sense), of the kind' (Kripke 1972 p. 330).

The necessity of essential properties is metaphysical, not epistemic. The claim is that things of a kind have its essential properties in all possible worlds, not that its essential properties are knowable *a priori*.[3] In particular it is not supposed to be analytic to ascribe its essential properties to things of a kind. Kinetic theory gives the essence of temperature, not the meaning of 'temperature'. Essentialism needn't therefore dispute Quinean critiques of the analytic–synthetic distinction on the one hand, nor on the other need it plague theoretical conflict with the problems of incommensurability (Feyerabend 1962, Kuhn 1962) or indeterminacy of translation (Quine 1960):

> Note that on the present view, scientific discoveries of species' essence do not constitute a 'change of meaning' ... We need not ever assume that the biologist's denial that whales are fish shows his 'concept of fishhood' to be different from that of the layman; he simply corrects the layman, discovering that 'whales are mammals, not fish' is a necessary truth. Neither 'whales are mammals' *nor* 'whales are fish' was supposed to be *a priori* or analytic in any case. (Kripke 1972 p. 330)

So proponents of rival theories are not doomed by essentialism to Kuhn's (1970) and Feyerabend's (1970) dialogue of the deaf. We need not, therefore, continue that prolonged dialogue here.

2 Essentialism about kinds has various sources. It derives partly from the plausibility of examples such as those given above. Some among the properties common to things of a kind undoubtedly matter more than others. In particular, some will be more central than others to a theory which explains the properties and relations of things of that kind. I consider in **7** whether properties being important in that sense is in fact best explained by, and thereby lends support to, the claim that they are essential properties of the kind; and I conclude that it is not.

Individual essences are a classic source of essentialism about kinds. Some kinds may provide criteria for the reidentification of things of that kind, such that no thing of the kind could survive change in the specified respect. It is

---

[3]Kripke (1971 pp. 150–1, 1972 pp. 260–3). I have used Dummett's term for the latter notion, since I incline to accept his account, on which 'epistemic necessity is a stronger notion ... a statement may be [metaphysically] but not epistemically necessary, but the converse could not occur. Kripke, however, claims the properties of being *a priori* and being necessary to be quite independent' (Dummett 1973 p. 121).

arguable, for example, that a man could not survive the loss or drastic transformation of the body whose spatio-temporal continuity settles questions of human reidentification. I don't think that is true, but even if it were, it would follow neither that any men, nor that all men, must have bodies of the kind specified. In the first place, John's inability as a man to *become* a beetle is compatible with the possibility of his always having been one; in the second, there *could* still have been *men* (other than the men there actually are) who lacked these bodily features and would be reidentified over time by different criteria. Anyway, arguments from individual essence, if they worked at all, would not work for kinds, such as water and gold, that provide no criteria for the reidentification of things. Perhaps there are characteristics a gold cup cannot come to have, but that will not show that other gold objects cannot have them.

There are, I believe, no sound inferences from individual essences to kind essences; but that point is not new, and I need not argue it further here. My concern here is with two other, newly fashionable, arguments for essentialism about kinds. One, due to Putnam (1975), derives essentialism directly from a theory about how the extension of kind terms is fixed. The other, due to Kripke (1971, 1972), derives it indirectly *via* a theory of the singular reference of natural kind and other seemingly general terms, from whose necessary self-identity essentialism is taken to follow:

> When we have discovered that heat is molecular motion we've discovered an identification which gives us an essential property of this phenomenon. We have discovered a phenomenon which in all possible worlds will be molecular motion – which could not have failed to be molecular motion, because that's what the phenomenon *is*. (Kripke 1972 p. 326)

Putnam's theory of the extension of kind terms, and Kripke's theory of their reference, are alike in denying traditional accounts that make the reference (or extension) of terms a function *inter alia* of something like their Fregean sense.[4] As applied to kinds in particular, the new theories deny that the extension of kind terms is any function of descriptions believed by their users to be true of things of the kind (Putnam 1975 p. 221). Fregeans, who believe the contrary, need not of course deny that there are non-analytic

[4]I say '*inter alia*' because Fregean reference or extension is obviously a function also of context (e.g. in indirect speech, according to Frege, a name refers to what is normally its sense) and of what the world contains (e.g. whether 'gold' applies to my tiepin depends on whether my tiepin is gold). (Cf. Dummett 1973 chs 5, 9.)

essences of kinds (Dummett 1973 p. 117), but Fregean theories of how kind terms get their extension give no especial reason to think there are any. Fregean theories in general, and description theories of kind terms in particular, yield necessity only as a by-product of analyticity. (Had our Fregean sense of 'water' made us apply it only to what we believe 'H_2O' applies to, then – at least for us in this world – water would in all possible worlds have been H_2O. But that, as we have seen, is not why essentialists think being H_2O is of the essence of water.)

To provide essence without analyticity, an alternative is needed to the Fregean sense of what water is. Putnam to that end tells two tales designed to bury Frege, as a prelude to recommending his own essentialist account of natural kinds. In fact, Putnam buries Frege alive and well; and that fact must be shown first, before we can profitably turn to the deficiencies of Putnam's and Kripke's rival theories.[5]

3 Putnam's tales are aimed at the idea of a kind term's extension being *any* Fregean function of its users' beliefs. The tales therefore present cases where such an extension differs for two groups of users with relevantly identical beliefs. First, we suppose a Twin Earth somewhere, which is just like Earth except for a different microstructure, XYZ, of what they too call 'water'. Macroscopically XYZ is indistinguishable from H_2O, and it plays just the same part in Twin Earth life that H_2O does here. By 1950, however, it has become common knowledge on each planet that the other lives off different stuff. But back in 1750 no one knew about the microstructure of water, and each planet had identical beliefs about the stuff they so called. Yet 'the extension of the term "water" was just as much H_2O on Earth in 1750 as in 1950; and the extension of the term "water" was just as much XYZ on Twin Earth in 1750 as in 1950' (Putnam 1975 p. 224). Now the local stuff was no doubt in the extension of 'water' as used on each planet in 1750. If the stuff on the other planet was different in kind, it was presumably not in fact in the term's extension, even though the users then would have mistakenly thought it was. Hence the anti-Fregean conclusion: 'water' can have different extensions in the same world for different users who give it the same Fregean sense.

I agree that 'water' had (tenselessly) the same extension in 1750 as it had in 1950; what I deny is that at either time that extension was different on Earth and on Twin Earth. The fact that Twin Earth's 1950 beliefs about local water differed from ours doesn't begin to show that the extension of their term

[5]My arguments against Putnam's interpretation of his two tales overlap with those of Zemach (1976), but my further purposes make it desirable to restate them here in my own way.

'water' differed from that of ours. It doesn't even follow that the senses of the term differed; and if they did, the whole point of the sense/reference distinction is to allow sameness of reference (or extension) to accompany difference of sense. It is indeed quite plain to my Fregean eye that in 1950, as in 1750, 'water' had the same extension on Twin Earth as it had here. There was water on both planets alike, and there still is. We simply discovered that not all water has the same microstructure; why should it? Because its microstructure is an essential property of water? Well, that is what's in question.

Fregeans need not resort to science fiction to recommend their reading of this tale. There is a perfect precedent in the discovery of isotopes. If Zemach's (1976 p. 120) heavy waters are too rare and exotic to convince, try the two common isotopes of chlorine. Note that in these real cases the various isotopes occur together in natural samples; they aren't segregated onto separate planets. It is therefore undeniable that the extension of 'chlorine' included both isotopes before their discovery, and so presumably includes both isotopes now.[6] What Putnam must say is that chlorine and water have been found not to be natural kinds after all, but rather mixtures of natural kinds. But in that case, as Zemach (1976 p. 122) observes, it will very likely turn out that we have no natural kind terms. Anyway, the pertinent point is that the first Twin Earth tale doesn't compel that conclusion, which it would have to do to dispose of Frege. The Fregean reading that Putnam overlooks is *prima facie* at least as plausible as his own.

Putnam, however, has another Twin Earth tale that won't admit this Fregean reading. This time we suppose that aluminium and molybdenum are practically indistinguishable, that molybdenum is as common on Twin Earth as aluminium is here, and consequently that on Twin Earth all our Earthly uses of the two metals are interchanged. Moreover, like Americans, Twin Earth men don't call aluminium 'aluminium'; they (unlike Americans) call it 'molybdenum'. The term 'aluminium' they reserve for molybdenum. Now most people, both here and there, can't tell the metals apart; call these people 'laymen'. Laymen here are therefore in the same psychological state about aluminium that laymen there are about what they also call 'aluminium'. Yet the extension of 'aluminium' as laymen use the term there is indubitably molybdenum and thus quite different from the extension of 'aluminium' as laymen use the term here.

---

[6]The sense of 'chlorine' *might* of course have changed in the wake of this discovery, to make the term apply only to the more common isotope; just as the sense of 'water' might have changed to exclude XYZ or D_2O. But that is not what happened, and anyway not what Putnam needs. If changes of belief about the microstructure of kinds do produce changes of extension in kind terms, that rather recommends a Fregean view.

My Fregean waterworks, chlorinated or not, will not wash with this tale. Our laymen know that aluminium isn't molybdenum, even if they can't tell what the difference is. So we can't make our term 'aluminium' apply also to Twin Earth molybdenum as we made 'water' apply also to XYZ. The reason of course is that there are high priests as well as laymen, experts who *can* tell the difference. Our experts fix the extensions of 'aluminium' and 'molybdenum' for laymen here; experts on Twin Earth fix them the other way round for laymen there. There is, as Putnam (1975 p. 227) puts it, a 'division of linguistic labour'.

Very well. It need not be *my* beliefs that fix the reference or extension of terms which I can use quite well in my limited way. So I defer to experts, whose job it is to say what such a term really applies to. The reference or extension in any possible world of the term as we use it may nevertheless still be some Fregean function of our *experts'* beliefs. In Australia, for example, to be 'back of Bourke' is to be way out in the outback. That Bourke is at an edge of Australian civilisation is all I know about that place, certainly not enough to enable me to tell Bourke from several other places. Yet I can still refer to Bourke, as I have just done, by taking it to be whatever place would best fit more expert geographers' beliefs about it. Our knowledge of most things and places must be like this; certainly most of our knowledge of natural kinds. So no doubt the labour of reference is divided, as Putnam says, but it may be a Fregean labour for all that (Dummett 1973 pp. 138–9, 1974 pp. 530–1).

4 Fregeans can cope with Putnam's Twin Earth tales. Frege is still in the ring; so how does Putnam fare on points? His rival theory gives the extension of natural kind terms in two stages. First, archetypes in this world, paradigm specimens of the kind. Then anything, in any possible world, that has a suitable 'same-kind' relation to the archetypes. It is for science to tell us what the same-kind relation is for any category of kinds (for H_2O, Putnam proposes, implausibly enough, a same-*liquid* relation!). Generally, Putnam assumes that the relation will specify a shared microstructure. But whatever shared properties Putnam's same-kind relation picks out will be essential properties of the kind, since the relation is assumed to be an equivalence relation that holds across all possible worlds (Putnam 1975 p. 232), not just in this one. So not just actual specimens of the kind share the specified properties: nothing could be of the kind and not share them with the kind's archetypes.

Putnam's necessity is metaphysical, not epistemic. A natural kind may be known long before its essential microstructure is known. Putnam's theory is

radically anti-Fregean: given archetypes, a kind term's extension is fixed by its same-kind relation, regardless of what anyone believes.

Putnam's argument for kind essences credits kinds both with archetypes and with cross-world equivalence relations holding between all things of the kind. Real natural kinds need have neither. Take archetypes. Their rôle is to fix the kind term's extension without recourse to its Fregean sense. Putnam takes it that they must therefore be in this world, so that they can be picked out ostensively. Thus pointing out Lake Michigan as an archetypal sample of water (Putnam's example) fixes the extension of 'water' regardless of its sense (just as naming the lake 'Lake Michigan' is supposed to fix the reference of that expression regardless of *its* sense). Then the rest of the extension of 'water', in any possible world, is just what stands in the relevant same-kind relation to *this* archetype. That is why Putnam (1975 p. 234) says that kind terms are 'indexical', like 'I', 'now', 'here' and 'this'.

Ostensive reference, to just *this* archetype in *this* world, is thus essential to the mechanism of Putnam's essentialist theory. He must show, therefore, that our use of kind terms actually incorporates it. An extension of Kripke's causal theory of naming offers to show this. The theory has some irrelevantly contentious aspects, about how users of names pass them on;[7] all we need here is that, roughly, some archetype must be causally 'upwind' of any use of a natural kind term. Thus our uses of 'water' and 'aluminium' are supposed to derive causally from our (or our experts') causal acquaintance with archetypal specimens of H_2O and aluminium respectively; and that is supposed to be why H_2O and aluminium are what we refer to by those terms. The corresponding Twin Earth uses derived causally from archetypes of XYZ and molybdenum; which is why those are the kinds they refer to by the same terms.

That is the theory. Unfortunately, archetypes do not constrain our use of natural kind terms in that way. True, botanists designate type specimens of plant species, and geneticists designate cultures to exemplify gene-types. But these specimens are causally *down*wind of the usage they are supposed to constrain. They are chosen to fit botanical and genetic knowledge, not the

---

[7] 'When the name is "passed from link to link", the receiver of the name must, I think, intend when he learns it to use it with the same reference as the man from whom he heard it' (Kripke 1972 p. 302). Intention isn't enough, however: 'I am asked to name a capital city and I say "Kingston is the capital of Jamaica"; ... [I] said something strictly and literally true even though it turns out that the man from whom I picked up this scrap of information was actually referring to Kingston-upon-Thames and making a racist observation' (Evans 1973 p. 194). 'We are left with this: that a name refers to an object if there exists a chain of communication, stretching back to the introduction of the name as standing for that object, at each stage of which there was a *successful* intention to preserve its reference. This proposition is indisputably true; but hardly illuminating' (Dummett 1973 p. 151). See also Altham (1973).

other way round. They are certainly not the specimens whose classification caused the corresponding kind terms to be used in the first place; they may well indeed not even be of the same kind as those specimens. Hence, as Nicholas Jardine (in a Cambridge University Moral Sciences Club paper) and Zemach (1976 pp. 123–4) have observed, our most authoritative specimens of a kind might on Putnam's account not even *be* of that kind.

So some archetypal natural kinds have the wrong archetypes; others have none at all. Consider elements high in the periodic table, that do not occur in nature and have never been made. We have names for them, but there may never be archetypes to constrain our use of the names. Even if specimens eventually appear, the discovery, creation or synthesis of previously unknown fundamental particles, elements and compounds can surely be *predicted.* The term 'neutrino' applied to just the same particles when it was used to predict their existence as it has applied to since their discovery. Ostensive reference (say to a bubble chamber photograph) could not have fixed its extension then; why suppose exactly the same extension is fixed that way now?

Even if we were to grant Putnam his archetypes, however, his essentialism would still fail to follow. No reason is given why particular properties must be common to all things in all possible worlds that are of the same kind as the archetypes. Suppose that all samples of water in fact share ten 'important' properties, but that water could lack any one of them, so that only the disjunction of all conjunctions of nine of them is essential. In this case of course sameness of kind is not the equivalence relation Putnam (1975 p. 231) says it is, since it is not transitive: two merely possible samples of water could differ in two of the ten properties. But Putnam's account doesn't in fact provide transitivity, since what makes things water in other possible worlds is their likeness to archetypes in *this* world, not their likeness to each other. To claim that the relation is an equivalence relation, so that archetypes have to share the *same* properties with all possible samples of the kind, is just gratuitously to assume the essentialist conclusion.

5 Putnam's account of the extension of kind terms both fails to be true and fails to entail essentialism. I turn now to Kripke's derivation of essentialism from the necessary self-identity of natural kinds. In evaluating his argument it is essential to keep the reference of a kind term clearly distinguished from its extension; to which end I hereafter distinguish the supposedly singular term 'water' from the corresponding predicate '... is water', and likewise for other kind terms. Now Kripke (1972 p. 349) admits that a causal mechanism is not needed to secure the reference of natural kind terms. 'Neutrino' could be

introduced, as it was, by theoretical description and still be applied 'rigidly' in Kripke's sense. That is, we can still consider the consequences of just *that* kind of particle failing in this or that respect to satisfy the theoretical descriptions which in fact served to identify it. 'Neutrino', so understood, is for Kripke a non-Fregean rigid designator, since its reference in other possible worlds is not constrained to satisfy these theoretical descriptions, which must be supposed to provide its Fregean sense. For this to work, of course, there must actually *be* neutrinos, near enough as specified. They needn't be in the observed past, to serve as archetypes; but they do need to be somewhere in this world, past, present or future. Otherwise the requisite uniqueness of reference would not be secured. Many different kinds of particle will satisfy our theoretical descriptions of neutrinos in various possible worlds; and nothing but the reality of one of these will single it out as the unique referent of the term 'neutrino'. Were there in fact no neutrinos, the term could for Kripke no more designate a natural kind than 'unicorn' can (Kripke 1972 p. 763).

However, there are neutrinos, just as there is H_2O. Let us therefore grant for the moment that 'H_2O', like 'neutrino', can be made a non-Fregean rigid designator by theoretical description and Kripkean fiat. Grant also that 'water' is such a designator, in this case perhaps even because our use of the term derives causally from Putnamesque archetypes like Lake Michigan. What more is needed for 'water' and 'H_2O' to designate the *same* kind? And how is sameness of kind related to coextensiveness in the corresponding predicates '... is water' and '... is H_2O', which is what matters for essentialism?

We must in fact tackle the latter question first in order to answer the former. As referents of singular terms, kinds are obscure entities, not notably clearer than properties or attributes in their criteria of identity. If the necessary self-identity of kinds is to have any implications for essentialism, these criteria will have to be spelled out in terms of predicate extensions. One such criterion presents itself at once. For water to be the same natural kind as H_2O, it seems at least to be necessary for '... is water' and '... is H_2O' to be coextensive in all possible worlds: otherwise some possible world would contain something whose membership of the kind depended on what the kind was called, which seems implausible.

But then it looks as if 'Water is H_2O', construed as an identity statement containing rigid designators, already and trivially entails that all samples of water in all possible worlds are also samples of H_2O. Far from the *necessity* of this identity establishing that being H_2O is an essential property of water, that is just what must be the case for the identity claim to be true at all. It is indeed not clear what more the metaphysical necessity of 'Water is H_2O' could consist

in, since we have already used up our possible worlds in saying what makes it true. So no doubt it is necessary if true; but only because if in this world '... is water' and '... is H_2O' are coextensive-in-all-worlds, then so they are in all other worlds.[8] In short, because the identity criteria of kinds have thus to be given in terms of their extensions, there is no useful inference to essentialism from the necessity of identity, even if the kind terms involved are admitted to be rigid designators; if anything, the inference will be the other way.

One might try arguing, however, that the requirements of natural kind identity have been pitched too high. Perhaps what is needed is coextensiveness, not in all possible worlds, but in this world and in those nearest to it. After all, science is supposed to give us essences; yet the most scientists can show us in fact is lawful coextensiveness between '... is water' and '... is H_2O'. That is, we suppose they can show that not only *are* all samples of water samples of H_2O, but that if anything *were* a sample of water it would be a sample of H_2O. Now that need not be a claim about all possible worlds, since the consequent of a true subjunctive conditional need not be true in all possible worlds in which the antecedent is true: it need only be true in those worlds which are sufficiently like ours (Lewis 1973). Waiving the difficulties of characterising lawfulness in terms of truth in nearby worlds without begging the question, suppose for the sake of argument that the lawfulness of 'All and only samples of water are samples of H_2O', so characterised, suffices for the identity of water and H_2O. Will the rigidity of 'water' and 'H_2O' now secure necessity for this identity and thus coextensiveness in all worlds, however remote, for '... is water' and '... is H_2O'?

To see that it will not do the latter, we must appreciate how loose the connection is between the reference of 'water' (say) and the extension of '... is water' in various possible worlds. Obviously there could be much more water than there is, or much less. That is to say, there are possible worlds with samples of water that don't exist in this world; and others in which some real watery individuals (like Lake Michigan) don't exist at all. Consequently, in yet other possible worlds, there is water all right, although no individual sample of it is identical with any of the real samples that we have. In each of these worlds the singular term 'water' has the same reference (namely, of course, water – whatever that is), while the extension of the predicate '... is water' may differ totally from what it is here or in other possible worlds.

---

[8] I ignore as incredible accounts of metaphysical (as opposed to epistemic) necessity in which this does not follow, i.e. in which the accessibility relation between possible worlds is not transitive (Hughes and Cresswell 1968); but see below.

Now we are at present supposing, for the sake of Kripke's argument, that coextensiveness between '... is water' and '... is H_2O' in all worlds is not required for the identity of water and H_2O, only coextensiveness in this and some suitable class of nearby worlds. We now see that in other worlds the extension of '... is water' may be quite different from its extensions hereabouts; and so may that of '... is H_2O'. And whatever the necessity of identity may do to secure that 'water' and 'H_2O' have the same reference in these other worlds, it will do nothing to secure that '... is water' and '... is H_2O' have the same extensions, which is what essentialism needs.

6 Even if 'water', 'H_2O' and similar kind terms were rigid designators, therefore, essentialism would get no help from the supposed necessity of identities like 'Water is H_2O', however the reference of the singular terms involved is related to the extension of the corresponding predicates. But there is anyway no good reason to admit that these terms are non-Fregean rigid designators. Kripke (1971 pp. 146–8) implies that without rigid designators we would need Lewis' (1973) 'counterparts' in order to state counterfactuals about things (and kinds). I share Kripke's distaste for counterparts, but this is a false dichotomy. Fregean names may designate the same things or kinds in many possible worlds, namely in all those that can be specified by Fregeanly intelligible counterfactuals about them. On a 'cluster' version of the description theory (e.g. Searle 1958), taking account of Putnam's division of linguistic labour, such counterfactuals can suppose the lack of almost any property of the thing, or of things of the kind, and of all such properties attributed by any one non-expert speaker. That is quite enough to give a specious appearance of non-Fregean rigidity to what may in fact be Fregean names. Of course it is trivially true that 'water' applies to water whatever intelligible counterfactual supposition we make about it. The question is whether something like a Fregean sense of 'water' limits the range of such suppositions. I see no reason to deny that it does, and equally none to assert that anything about water itself makes the antecedents of any such counterfactuals necessarily false.

7 We have seen that essentialism can be extracted neither from Putnam's archetypes, nor from a merely stipulated rigidity of reference *via* modish truisms about identity. The existence of essences in Kripke's and Putnam's theories is no more than a gratuitous assumption on their part. Its appeal lies solely in that of the stock exemplars of essential properties; and that appeal, I shall argue in conclusion, is specious.

In biological species, for example, there is a distinct dearth of suitable properties shared even in this world. Capacity to interbreed with fertile offspring is the obvious candidate, but even that is well known to lack the transitivity Putnam's same-kind relation would need in order to yield essentialism. It usually doesn't hold between all and only members of the same species here, never mind in other possible worlds. Nor (*pace* Dummett 1973 pp. 143–5) can coming of a common stock be essential to a species,[9] which could easily have been cross-bred from independent mutations.

Elements and chemical compounds offer more scope for essentialism, since all their specimens at least share some properties in this world. Of properties supposed to be shared in all possible worlds, we have seen that microstructure provides the stock exemplars; it is worth asking why. Scientists commonly employ a principle of 'microreduction' (Schlesinger 1963 ch. 2), i.e. roughly the principle that properties of things should be explained in terms of the properties and relations of their spatial parts. Many properties can indeed be so explained; and the assumption that they can underlies standard techniques for studying things in convenient (e.g. laboratory) isolation from their normal surroundings.

Microreductive theories are thus relatively easily testable, which is a well-known Popperian virtue in science. It is therefore both good method to look for microreductive explanation of kinds of things and an importantly pervasive fact that they can be found. Where, moreover, such explanation is both comprehensive and deductive, we may be able to replace reference to things of a kind with reference to their parts. So it is, roughly, that reference to gas samples was made eliminable, by the classical kinetic theory, in favour of reference to gas particles.

Suppose we now adopt a Quinean view of ontology, and admit only kinds of things that *need* to be referred to in stating what we know. Since kinetic theory makes reference to gas samples redundant, they disappear from our ontology. Similarly, let us suppose, with water and gold. Microreductive theories make reference to anything more than H_2O molecules and gold atoms redundant. There is, we might say, nothing to water and gold but the particles of which they are composed. And this, I reckon, is the source of the idea that water would have to be H_2O in any possible world, because H_2O is all there is to water.

---

[9] I fear this idea derives partly from Kripke's (1972 pp. 312–14) claim that an offspring's parents are essential to it; a claim of which it may suffice to remark that 'If John hadn't been a Kennedy he wouldn't have been shot' is a plainly intelligible contingent statement (cf. Dummett 1973 p. 132).

The inference is specious. This way of removing items from our ontology requires deducibility. Unless all the macroscopic properties of a kind are deducible from its microstructure, reference to things of the kind is still required. And if they are deducible, then they occur in any possible world the microstructure occurs in. So if the microstructure is essential for this reason, so are all the macroscopic properties it explains.

So-called 'essential' properties are thus really no more essential than any other shared properties of a kind. They are just properties ascribed by the primitive predicates in a comprehensive deductive theory of the kind. That is presumably what makes them what Putnam (1975 p. 232) calls 'the important physical properties' of the kind; and the principle of microreduction will no doubt ensure that such properties will most commonly concern the kind's microstructure.

Our anti-Fregean essentialists must now suppose a property's 'importance', so construed, to be a feature of the world independent of our beliefs and theories. If we knew everything, perhaps that would be so, as Ramsey (1928) thought in 1928. But as Ramsey (1929a) observed in 1929, we don't know everything. Even microreductive theories cannot be construed as claiming completeness merely by virtue of claiming truth. For there may always be more smaller scale detail to the world than any such theory describes. Its presently primitive predicates may therefore cease to be primitive at any time, when it is itself explained by yet another such theory, without at all impugning either its truth or its explanatory value. Water may still be H_2O, however much we subsequently learn of the nuclear structure of hydrogen and oxygen atoms.

8 I have tried in this paper to dispose of some of modern essentialism's newer and more seductive arguments. Putnam's Twin Earth tales do not, as he supposes, dispose of Fregean alternatives to essentialist theory. His own account of the extension of natural kind terms is false of nearly all natural kinds and would not yield essentialism even if it were true. Kripke's theory of the reference of kind terms likewise fails to yield essentialism as a product of the necessary self-identity of natural kinds. The stock candidates for essential properties, moreover, are either not even shared in this world by all things of the kind, or their status is evidently more a feature of our theories than of the world itself.

In short, our essentialists' premises are false, their arguments invalid, and the plausibility of their conclusions specious. Their essences can go back in their Aristotelian bottles, where they belong.

8 *Necessities and universals in natural laws*

1 Prologue How do laws of nature differ from cosmic coincidences? This is a question very familiar to philosophers of science, and answers of two sorts still vie for their allegiance. One sort locates the difference in what laws say, the other 'in the different rôles which they play in our thinking', as Braithwaite (1953 ch. 9 p. 295) puts it. Braithwaite developed and defended a classic answer of the second sort: the difference, he says, lies in why we believe laws, not in what they say. In the quarter century since then, other answers of the same sort have been devised, e.g. by Hesse (1980). But since then also, answers of the first sort have again come into fashion. The revived fashion has mostly been for reading laws as saying how things must be; but some, more recently, have read them instead as relating not things but properties of things to each other. Hesse notes these fashions and rejects them, to my mind rightly, but she does not elaborate her reasons. It seems to me therefore that I can best complement her article by inspecting these fashions' argumentative cut, to see if they do indeed fit better than her and Braithwaite's Humean gear. Only first I shall build the problem up in my own way, to provide a lay figure to hang the garments on.

2 The problem Certified laws of nature are the primary products of scientific thought and observation. They embody the generalised knowledge which science yields; they supply explanations and predictions of events; and they underlie the design of most modern artefacts. To take just three obvious examples: our human life has been much altered in this century by the discovery and applications of laws governing plant genetics, aerodynamics and electromagnetic radiation.

Laws differ widely in their subject matter, importance and complexity. What they have in common is generality. A law says that all things or events of some kind have a certain property or are related in a certain way to something else. If the law is statistical, the property is that of having a chance of having some other property or of being related to something else. It is, for

example, a law that all light has the property of going at the same speed in a vacuum; and it is a statistical law that all atoms of the most common isotope of radium have the same chance (fifty-fifty) of turning into something else within their half-life of 1622 years.

What needs certifying about a law is its truth. We cannot know that all light goes at the same speed in a vacuum unless it truly does so. Its constant speed will not serve to explain or predict anything if its speed is not in fact constant. And it is unsafe to base the design of artefacts on what is not the case. We know of course that even a certified law may turn out to be false. But without good reason to think it true, we lack good reason to employ it as we do. This is why we do not call something a law unless we think it true, so that a false generalisation cannot be a law, although it may be 'lawlike': i.e. such that it would be a law if only it were true.

Certifying the truth of some laws presents no problem. These are the analytic laws, those whose truth follows from the meanings of the terms they are couched in. There are more reasons than one for laws being analytic. A law may be analytic because it is used to define one of its terms. Newton's laws of motion, for example, may well be analytic because between them they define the Newtonian concepts of force and mass. Or a law, not originally analytic, may become so successful and theoretically important that its terms change their meaning to make it analytic. For this reason it is now arguably analytic that light is electromagnetic radiation, although that could not have been the case when the electromagnetic theory was first conjectured to apply to light. Then, we could easily have envisaged observing light to go faster or slower, for example, than the theory can be shown (by measuring the ratio of electromagnetic to electrostatic units) to require electromagnetic radiation to go. Nowadays we should take such an observation to show some error in the theory rather than question the law that light is electromagnetic radiation.

But even if some laws are analytic, most laws are not, and these are the ones that concern me. It does not follow from the meanings of the terms involved that radium's half-life is 1622 years, nor that benzene is as insoluble in water as it is. Nothing semantic prevents a little more benzene sometimes dissolving in water, or some piece of radium having a rather different half-life. How then can we certify the truth of what the law says, namely that these things never happen? We cannot see that they never do, if only because at no time can we see that they never will do in the future. We cannot directly perceive the truth of non-analytic laws. At most, our senses can show us some of a law's past instances, and then only instances of laws about relatively observable properties of things and events. We can observe the speed of this

or that ray of light and, perhaps indirectly, the half-life of this or that piece of radium; but not all the things and events, past, present and to come, to which the law applies.

The problem then arises why a supposed law should be expected to hold in instances as yet unobserved: in short, Hume's problem of induction. Unlike Popper and his followers, I believe that induction does present a genuine and serious problem which needs a solution (such as that offered in chapter 15). But wherever its solution lies, Hume's problem does not arise only incidentally for laws of nature. On the contrary, it is an inevitable concomitant to their rôle in supplying predictions. To make a prediction is to anticipate, rightly or wrongly, the result of making an observation: to say or just to expect, for example, that a bomb will explode before we see it do so. Whatever purports, as a law does, to justify such an expectation necessarily arouses Hume's problem. Only a generalisation certified by observing all its instances would be free of inductive pretensions, and such a generalisation is not much use for predicting things. It might indeed have some use: one might accept it on someone else's authority and use it to predict some instances one had not observed oneself. But real laws are used amongst other things to predict the results of future observations, and these are not yet available to anyone to certify the law with. Real laws therefore undeniably need inductive support.

The other philosophical problem which laws of nature present is the one that concerns us. It is less obvious than the problem of induction, but perhaps more tractable: what exactly do laws say? I have taken them to be generalisations, and there is not much doubt of that. The debatable question is whether laws are more than generalisations, and if so, what more? And the problem of induction provides a reason for preferring weak readings of them: for the less a law says, the less there is to be certified in claiming it to be true.

The weakest reading seems to be the obvious one I have already given:

(1) All *F*s are *G*s,

where *F* and *G* are properties of things or events. They may be relational, comparative or quantitative properties; in statistical laws *G* will be some determinate chance of having another property. (1) is of course a very simple form of law, but it will do: it has all the relevantly problematic features of more complex forms. But before discussing its supposed deficiencies, some preliminary points need to be made clear.

First, as my examples have already illustrated, the 'are' in (1) is to be taken tenselessly. The law applies to all *F* items in the universe, past and future as well as present. The laws of radioactivity do not just give radium's present

half-life: they say what it always was and always will be. Now some of what we take to be physical constants, such as the half-life of radioelements, might indeed turn out to depend on the age of the universe. But then the true laws of radioactivity would say what the dependence was. Those laws would, like all other true laws, apply at all times, the values of our supposed constants at particular epochs being merely special cases of the general laws.

Secondly, I take it that anything in the universe is definitely either *F* or not *F*, either *G* or not *G*. This is not an uncontentious claim. Some have been led to deny it of so-called 'vague' properties like being bald, because of its seemingly absurd consequences (for example, that at some point adding just one hair to a bald man's head gets rid of his baldness). Others have been led to deny it of some things and events in the future, either because they want the future to be open, at least in some respects, to being made definite by human decision and consequent human action or because of problems raised by quantum mechanics. They think it cannot now be the case, for example, that I shall definitely either be dead or be alive next year, if it is still open to me and others to settle the matter by what we decide to do between now and then. I think that these are both inadequate grounds for denying that everything is definitely *F* or not *F*, but I shall not argue the point here. (On the first, see Cargile (1969); on the second see chapter 11. I also think my being wrong in either case would make little difference to the ensuing discussion, but I shall not argue that either.)

Thirdly, I exclude from the range of *F* and *G* factitious properties such as Goodman's (1965) notorious 'grue' (= green if the item is inspected before a specified time, otherwise = blue). I hope and believe criteria can be given to rule out these phoney properties (see e.g. Hesse 1974 ch. 3); but in any event all parties agree that they are phoney, and I shall take their exclusion for granted.

I should however emphasise that I do not mean to restrict *F* and *G* to physical, as opposed to psychological or social, properties. Some philosophers (e.g. Davidson 1970b, McGinn 1978) deny the existence of laws relating non-physical properties; but largely because they mistake laws to involve necessities of the kind I shall be concerned to dispute and which they correctly perceive to be absent from mental and social generalisations. Anyway the point should be left open here; so if I stick to physical examples, it is only to avoid irrelevant controversy, and not because I think there are no others.

With this preamble, we may now ask what, if anything, is wrong with (1) as a reading of laws of nature. To see what seems to be wrong, we must look

at (1)'s consequences in special cases, particularly the case, on which Braithwaite concentrates, where nothing in the world is *F*.

One might imagine that it did not matter what follows from (1) when nothing is *F*, but it does. Let us call a law 'vacuous' in that case. Many important laws are vacuous in this sense. The most famous one is Newton's first law of motion, that bodies acted on by no forces are at rest or move at a constant speed in a straight line. The law is central to Newtonian mechanics, but Newton's own gravitational theory implies its vacuity, since the theory says that all bodies exert gravitational forces on each other. No doubt Newton's laws of motion are peculiar, since as already remarked they may well be analytic. But Newton's first law illustrates a vacuity which is shared by many laws that are in no way analytic. There is in particular a multitude of non-analytic laws quantifying over determinate values of continuously variable determinables: for example, the laws relating the vapour pressure of substances to their temperature. Each determinate value of these determinables yields another law as a special case, such as the law giving the boiling point of water at atmospheric pressure. Now there are infinitely many different temperatures and pressures, and hence infinitely many of these derived laws, all with mutually incompatible antecedents (nothing can be wholly at two different temperatures or pressures at the same time). Although the temperature and pressure of any given mass of water will vary continuously with time, there are many temperatures which no mass of water ever reaches: temperatures, for example, so high that water would decompose before it reached them. At any rate, so far as these derived laws are concerned, it is entirely accidental whether any water ever is at the temperatures and pressures they apply to. Consequently they must certainly be so construed as to make equal sense whether they happen to be vacuous or not. (Cf. Ayer 1956 pp. 224–5.)

In particular, it seems obvious that mere vacuity should not settle the truth of a law regardless of its content. But a lack of *F*s makes 'All *F*s are *A*s' true for any *A*, including both *A*=*G* and *A*=not-*G*. If there never is any water at some temperature *T*, statements crediting all water at that temperature with any pressure whatever all come out true. That seems absurd; so vacuous laws should be read as saying something other than 'All *F*s are *G*s'. The question is what.

The obvious answer is that a vacuous law says

(2) If there were *F*s, they would be *G*s.

But there are objections to (2). One is that it appears to imply that there are no *F*s, whereas laws, even if they happen to be vacuous, certainly do not claim to be. We could in reply say that (2) is not to be read as having this implication; and this stipulation can indeed be given some independent rationale. A case can be made for saying that the implication is not part of what (2) says, but follows rather from applying general rules of discourse: namely, not to mislead, and to be as informative as possible (see Mackie 1972 pp. 75–7). These rules dictate that one should not say (1) when the law is known to be vacuous, since (1) is no more true then than is any other generalisation starting 'All *F*s are ...'. To pick out as a law the generalisation which relates *F* especially to *G* in these circumstances, one must have some reason other than its truth. The reason may not be specified, but the fact that there is one is signalled by using (2) instead of (1). Consequently, even if the law says no more than (1), (2) would normally be used when, but only when, the law is known to be vacuous. So (2) will indeed signal its user's knowledge of the law's vacuity, even though that is no part of what (2) is being used to say.

This is one of the arguments which can be used to defend Humean accounts of laws as saying no more than (1). It still leaves the problems of saying what reason there is to link *F* and *G* as a law does when there are no *F*s, and why (2) should be the right way to signal this reason. These are among the problems that have exercised Braithwaite and his Humean successors. But since my concern here is with their rivals, I shall concentrate instead on recent attempts to solve the problem of vacuous laws by giving (2) some assertable content over and above (1).

Laws, I have remarked, do not claim to be vacuous, even if they are; and ideally, they should say the same thing whether they are vacuous or not. It will hardly do to make laws say (2) if they are vacuous and (1) if they are not. A law cannot say (1) in both cases, we are supposing; can it say (2)? We have dealt with the obvious objection by removing (2)'s counterfactual implication (that there are no *F*s), which would have made all non-vacuous lawlike generalisations false regardless of their content. What can be said positively in favour of the suggestion?

Consider the universe of non-*F* things or events of which a vacuous law says that if they were *F*s they would be *G*s. It is surely immaterial to this supposed fact about these things or events that there happens to be nothing else which is *F*. So perhaps we should take the *non*-vacuous law also to say of every non-*F* thing or event that if it were *F* it would be *G*. But again, the law itself does not assert that these things or events are not *F*. It should say the

same of all things or events, whether they are *F* or not. Let us therefore take a law to say of every thing or event *x* that

(3) If *x* were *F* it would be *G*.

(Those who believe in possible as well as actual things and events may take '*x*' to range over them too.) I shall take the problem for our non-Humeans to be that of saying what (3) means in this case.

I shall not demand of them a general analysis of so-called 'subjunctive' or 'counterfactual' conditionals like (3). A general analysis would of course have to cover those that we are supposing to give the content of natural laws. But I am not convinced that other uses of these conditionals are homogeneous enough with this one to shed much light on it. In most other uses, for example, (3) might very well imply that *x* is not *F*, which we have seen it cannot do here. Or again, to make (3) true of an *x*, it may often suffice for that *x* to be *F* and also *G*. Lewis' (1973) influential analysis, for example, takes this more or less for granted, and the way he reluctantly accommodates possible exceptions (p. 29) will certainly not cope with natural laws. Yet natural laws must be exceptions: it might be a coincidence that an *x* is both *F* and *G*, and not a matter of natural law at all. So in this case it must take more than that to make (3) true of any *x*. And as our consideration of vacuous laws has shown, the extra cannot be that all other *F*s are *G*s too, for there might just as well be no other *F*s. So whether the law is vacuous or not, the truth of (1) will not suffice to make (3) true of everything. But what more than (1) can a law say?

3 Possible worlds The traditional non-Humean answer is that natural laws are or express necessities of some kind: what makes (3) true of everything is that *F*s not merely are *G*s, they have to be: (1) is not merely true, it is necessarily so. Conceptions of law as what Kneale (1949) called 'principles of necessitation' are of course by no means new. The problem with them is to justify the idea of necessity they invoke and to show how it explains the universal truth of (3). Of late years, the development of so-called 'possible world semantics' has made that problem look more tractable, and thus encouraged a revival of the idea that natural laws are necessary truths. It has done this by providing a systematic way of saying what makes statements of necessity (and of possibility) true. So in particular we might hope to find in it an acceptable way of saying what makes necessary natural laws true.

The basic concept of this semantics is that of a possible world. A possible world is a way the world might be, or might have been. There are many such ways, and therefore many possible worlds, of which the actual world is just

one. Possible worlds are distinguished by what the facts are supposed to be in them: if the supposed facts differ at all, so do the worlds. I might, for instance, die in various ways, and, for each way, at various ages. So there are many possible worlds in which I expire of, say, cirrhosis (or my counterpart in that world does so: see Lewis 1973 p. 39), and these differ amongst other things according to my or my counterpart's age at the time. In general, a statement which might be true, but fails to specify every detail of the universe, will be true in many possible worlds, differing amongst themselves in the details left unspecified.

Having in some such manner as this grasped the idea of possible worlds, and reified them, one can turn round and give, as the truth conditions of a statement, the set of possible worlds in which it is true. That is how possible world semantics offers to give the meaning of various kinds of modal statements, and in particular of statements of necessity and possibility. How enlightening this conceptual round trip is, from what might be the case, to what is the case in a possible world, and back again, is a very moot point, but one that can be waived while we see how well the concept copes with the supposed necessity of natural laws.

It follows at once from the definition of a possible world that a statement which might be true is one that is true in some possible world. Hence statements which have to be true are those which are true in all possible worlds. In particular, for (1) to be necessarily true is for it to hold in all the worlds there might be or might have been. Is that really what a natural law claims?

Suppose it is: does that solve the problem of vacuous laws and explain (3)'s being true of everything in the actual world? Consider again the case where there are no actual *F*s. The law does not say there are none, and it is tempting to suppose there always might have been. If that were so, then, on this view of laws, (1) would have to be true not only in this world, but also in worlds containing *F*s where its truth would not be the trivial consequence of vacuity it is here. And that would certainly distinguish (1) as a law from other vacuously true generalisations.

But this account depends on the possibility of there being *F*s; and, on this view of laws, there will often be no such possibility. I have cited the example of high temperature instances of the vapour pressure law for water that are vacuous because water decomposes before it reaches those temperatures. Now, that water decomposes below these temperatures is itself a natural law and so, on this view of them, necessary. Consequently these high-temperature instances of the vapour pressure law not only are vacuous, they have to be.

There could be no water at such temperatures. But that is to say there are no possible worlds in which these instances are not vacuous; and therefore none in which the truth of this instance of (1) is other than a trivial consequence of vacuity.

So the idea of laws being true in all possible worlds does not solve the problem of vacuous laws. Nor, for much the same reason, does it explain why (3) is true of everything in this world. Again, it would if anything, *a*, in this world might have been *F* even if it isn't. Then there would be possible worlds in which *a* (or some counterpart of *a*) is in fact *F*; and in all these worlds it, like every other *F*, is *G*. Where that is so, it seems to me undeniable that (3) is true of *a*. However, for any *F* there will be many *a*s of which it is quite incredible that they might have been *F*. Take the law that in a vacuum all light goes at a constant speed – which is to say that all photons do. It is true then, of anything at all, that it would go at that speed in a vacuum if it were a photon. But this is not to say of everything that it might have been a photon. There is no possible world in which I am (or any counterpart of me is) a photon; and *a fortiori* none in which, as a photon, I (or any of my counterparts) travel at the speed of light. That is not, I believe, what makes this instance of (3) true of me. Yet I believe it is true of me, since I believe the law; and there is surely no inconsistency in my combining these beliefs.

For subjunctive conditionals like (3) to be true, their antecedents do not have to be possible. This is blatantly obvious in *reductio ad absurdum* proofs, where the truth of a subjunctive conditional is actually used to prove that its antecedent is *not* possible. One and one cannot make three precisely because, if they were to, something impossible would be the case. It should be almost as obvious that conditionals which give the content of natural laws likewise do not imply the possibility of their antecedents being true. The vapour pressure example shows at least that they cannot both do this and themselves be necessary truths. And I have given elsewhere (see chapter 6) the example of safety precautions at a nuclear power station, which are supposed to make impossible the conditions under which, as a matter of natural law, the fuel would explode. It is ridiculous to maintain that the success of these precautions would disprove the very law that makes them necessary.

I am not sure why (3) should be so often thought to imply that *x* might be *F*. The reason may well be the same for taking (3) to imply that *x* is not *F*: namely, that it is customary to reserve subjunctive conditionals for use when their antecedents are believed to be false but possible. We see, however, that this custom is not invariable, and have in any case already seen reason not to make such a custom part of a conditional's meaning. So however natural the

thought may be, it is false, at least of the conditionals implied by natural laws. But the mistake is very widespread and of long standing, and it has had serious consequences. It has bedevilled the analysis of disposition statements, as I argue in chapter 6.9. It has likewise afflicted discussions of free will, in which 'I could have done *X*' is frequently equated with 'I would (or could) have done *X* had I chosen to.' But it obviously does not follow from the latter that I could have done *X*, since it obviously does not follow that I could have chosen to.

The common confounding of conditional statements with statements of possibility has thus had ill effects in more than one area of philosophy. The ill effect here has been that possible world semantics has been mistakenly thought to give sense to the idea that natural laws are, or assert, some kind of necessity.

4 Natural necessity Laws might however still be necessary even if possible world semantics fails to say what makes them so. What makes (3) true of everything might still be that nothing could be both *F* and not *G*, whether or not it could be *F*. But it is not at all obvious that this is so. Subjunctive conditionals are not in general made true by necessities. Suppose that if I were to go to London I would go by train. This does not mean that I could not go any other way, merely that I would not. Lewis' (1973) treatment of subjunctive conditionals recognises this fact about them: the consequent does not have to be true in all the possible worlds the antecedent is true in, only in those most like the real world.

Still, I have insisted that conditionals like (3) which follow from laws are a special case. In particular, it does not suffice for their truth that their antecedents and consequents are true; whereas my going to London by train may well make it true that, were I to go, I would go that way. So perhaps (3) does need some necessity to make it true of everything, even if conditionals in general do not.

But most natural laws seem to be contingent. Apart from those that are definitions, and those whose success has made them analytic, any law might have been false. We could have come across a counter-example to it; and we still could, even if we never will or would. That seems at any rate to be why we need to test our supposed laws by observation: things could be other than the law says, so we need to look and see whether or not they are. I believe, for example, that light could have gone in a vacuum at other than its constant speed, even if no photon ever does and even if nothing, were it a photon, ever

would. So on the face of it, conditionals like (3) no more exhibit necessity than does the conditional about my going to London by train.

Attempts have been made to explain away the apparent contingency of natural laws. One attempt, which need not detain us long, distinguishes logical necessity and possibility from their natural or physical counterparts. It is logically possible for *F*s not to be *G*, but not naturally or physically possible. But all 'physically possible' means is 'consistent with natural law'. So to say that something is physically necessary is merely to say that some law entails it. Whether the law says it has to happen, and whether the law itself has to be true, remain entirely open questions.

A more serious attempt distinguishes between metaphysical and epistemic necessities (Kripke 1971 pp. 150–1, Dummett 1973 p. 121); that is, between being necessary and being knowable *a priori*. Laws appear to be contingent because they cannot be known *a priori*. They cannot be proved in the way the truths of logic and mathematics can. We need to look and see what the world's laws are, and it may always turn out that what was thought to be a law really is not one. The *F*s we have seen to be *G* may mislead us into believing they all are, even though some future ones are not. It is consistent with all we have seen that there should be *F*s which are not *G*. That is the epistemic possibility of a supposed law being false; and something like it exists in mathematics. There too, special cases may mislead us into believing a mathematical generalisation to which there are in fact counter-examples. Now, recognising this possibility in mathematics does not diminish our belief in the necessity of mathematical truths: if the generalisation is true, it could not have been otherwise. It is likewise conceivable that natural laws, if true, are necessarily so, even though we may be mistaken in what we suppose the true laws to be.

The apparent contingency of natural laws could undoubtedly be explained away like this if there were good reason to think them necessary: but is there? The analogy with mathematics certainly does not give one. If Goldbach's conjecture proves true, any attempt to suppose it false will eventually lead to contradiction (that of course being one way of proving it). In that case no consistent description could be given of a world in which the conjecture was true. That is, there is no such possible world. We might therefore explain the conjecture's necessity, if true, as truth in all possible, i.e. coherently conceivable worlds: since conceivability is a notion arguably more basic than necessity and intelligible independently of it. But no such case can be made for the corresponding conception of natural necessity. As Hume insisted, there is no difficulty in conceiving a natural law to be false: since it is not analytic, no contradiction ensues. A perfectly coherent description can be given of a world

containing *F*s that are not *G*. The only ground for thinking such a world impossible would be that the law which would be false in it is not only true but necessary; and this is the very fact that needs to be established and explained.

5 Essences Arguments have recently appeared for the metaphysical necessity of some laws, namely those specifying essential properties of natural kinds. An essential property of a kind is one which nothing of that kind can lack. So if being *G* is of the essence of a kind *F*, the law that all *F*s are *G*s will be a necessary truth. The exemplars most widely touted by advocates of essences concern the microstructure of kinds: the atomic number of gold, the molecular constitution of water, the genetic makeup of plant and animal species, and the mean kinetic energy of gas particles at a given temperature. The question is: why suppose that these, or any other, properties of kinds are essential?

Two sorts of arguments have lately been adduced for essences, and hence for the necessity of the corresponding laws. Both employ possible world semantics; neither therefore proves more than that some generalisations hold in all possible worlds, and we have seen in **3** that this is not enough to serve our turn. But the arguments repay scrutiny nonetheless, since there is more to them than the possible world jargon they are couched in.

One argument, due to Putnam (1975), infers essences from a mechanism for fixing what things or events a kind predicate ('*F*') applies to, i.e. its extension. This mechanism fixes what things are, or might be, *F*s in two stages. First, there are archetypal actual *F*s (e.g. paradigm specimens of gold or water): things that have to be *F* if anything is. Second, to be *F*, anything else has to have a suitable 'same-kind' relation to these archetypes. What this means is that it has to share some property with them – apart of course from the property *F*. What the same-kind relation is, for any category of kinds, it is for empirically testable scientific theories to say. The relations are not discoverable *a priori*, and in particular they do not follow from the meanings of the predicates involved: the laws giving the essences of kinds are not supposed to be analytic. But any shared property *G* which a same-kind relation picks out will be an essential property of the kind since, Putnam assumes (p. 232), the relation is an equivalence relation holding across all possible worlds. Thus not only are actual *F*s all *G*s, all possible *F*s are: so (1) in this instance is true in all possible worlds.

I have elaborated elsewhere (see chapter 7) my reasons for rejecting this argument. Briefly, the extensions of real natural kinds do not in the first place depend on archetypes in the way Putnam's mechanism requires. And, in the second place, even if they did, his mechanism would still not produce essences.

To produce an essence, the same-kind relation must be transitive, in order to ensure that all possible *F*s share the *same* property *G* with each other. But the mechanism does not need a transitive relation, since what makes things *F*s in other possible worlds is their sharing some property other than *F* with the archetypal *F*s in this one, and there is nothing to say this shared property must be the same in every possible world. For Putnam to claim the same-kind relation to be transitive, which he does in taking it to be an equivalence relation, is for him gratuitously to assume the essentialist conclusion he is out to prove. His mechanism in fact gives us no reason to think any instance of (1) true in all possible worlds. And since in any case only those giving essences are in question, Putnam's theory, even if it worked, would not solve the general problem of distinguishing laws from universal coincidences.

The same of course is true of Kripke's (1971, 1972) argument for essences; but that too we must look at, since solving our problem for some laws would at least be better than solving it for none. Kripke's argument is quite different from Putnam's. Kripke takes laws giving essences to be identity statements: the law 'Water is H_2O' he takes to say not merely that anything, were it water, would be H_2O, but that being water and being H_2O are one and the same property. But identity is a necessary relation, in the sense that nothing could fail to be identical to itself. So being water and being H_2O are the same property in all possible worlds, not only in this one. Nothing that could be water could fail, were it water, to be H_2O.

This of course is the merest sketch of Kripke's argument. He has, for example, also to show that 'water' and 'H_2O' are what he calls 'rigid designators', i.e. that they refer to the same stuff in any possible world it exists in. Otherwise the identity statement, since it is not analytic, might be true without being necessary, as for example 'Water is the most powerful solvent' is. ('The most powerful solvent' is not a rigid designator: it refers to whatever the most powerful solvent is, which in a world restricted largely to oil products would not be water.) As I have abbreviated Kripke's argument, so I shall abbreviate the objections raised in chapter 7. The chief objection is that the argument, like Putnam's, blatantly begs the question: for being water and being H_2O to be the same property at all, never mind necessarily, the predicates '... is water' and '... is H_2O' must already be coextensive in all possible worlds. This is not a conclusion to be derived from the necessity of the identity: it is built into the identity as a premise. Granted, 'Water is H_2O' states a true law, and it has the form of an identity statement. But it is clearly only a variant of 'All water is H_2O', which does not have that form. At any rate, the identity of these properties only follows if 'water' and 'H_2O' are rigid designators, i.e. could

not refer to different properties. But to believe this, one needs already to believe what the argument from this premise is supposed to show: namely, that there could not have been some samples of water of a different molecular constitution.

Kripke, like Putnam, fails to establish the existence of essences. The microstructural exemplars which give their doctrine its spurious appeal indeed have a special status in science, but not the status of essences. They are special because they are central to our current scientific theories; but that, I have argued elsewhere (chapter 7.7), is quite a different matter from being necessary features of the world.

6 Universals The properties of being water and being H_2O do not stand in the necessary relation of identity. Perhaps, however, as Armstrong (1978 ch. 24), Dretske (1977) and Tooley (1977) have suggested, these universals stand in some contingent relation which makes it a law that all water is H_2O. This relation, that is to hold between the properties *F* and *G* whenever 'All *F*s are *G*' is a law, Armstrong and Tooley call 'nomic necessitation'; I shall call it '*N*'. *F* and *G* have to be differently related if the law is that no *F*s are *G* or that all *F*s have a chance *p* of being *G*. But *N* will do for now: if it works, the other relations will; if not, nor will they.

This suggestion requires a realist view of at least those universals which are related by natural laws: for *N* to relate *F* and *G*, these properties must exist. That of course is debatable, but suppose for the moment it is true. Then *FNG* is by definition the fact that makes 'All *F*s are *G*s' a law. This is a contingent fact, and not only because *F* and *G* might not exist. *F* and *G* could quite well exist without 'All *F*s are *G*s' being a law: laws do not relate every property to every other property. Being water and being at 100°C, for example, are properties that enter into laws, yet no law relates them to each other. But though it is not, it might have been a law that all water is at 100°C; just as it might not have been a law that all water boils at that temperature at atmospheric pressure. Apart from analytic laws, therefore, it is quite contingent that *N* relates any particular *F* and *G*.

To do its job, *N* has not only to make *G*s out of actual *F*s, it has to make (3) true of everything, i.e. to be such that anything, were it *F*, would be *G*. Since this is all it has to do and be, one might think that postulating *N* is more a relabelling of the problem than a solution to it. But that would not be a fair response. There is a dearth of candidates for making (3) universally true, as we take it to be. If *F*, *G* and *N* would between them make it true, that may well, as Tooley (1977 p. 262) urges, be reason enough to believe in them.

After all, we already invoke properties to make conditionals true. The inertial mass, m kilogrammes, of a thing a at time t makes true all the conditionals of the form:

(4) If a were subjected at t to a force of f newtons, it would then accelerate in the direction of the force at f/m metres/second^2.

Any of these conditionals is in fact a generalisation about events, namely that, if they were subjectings of a to a (specific) force f at t, they would be (or be shortly followed by) accelerations of a of magnitude f/m. These generalisations are just like the conditionals (3) entailed by laws, except that they are restricted to the individual a. We think them true, and a fact is needed to make them so: and the requisite fact is that a has mass m at time t. This is all the property of inertial mass amounts to: a truth-maker, as Tooley puts it, for conditionals like (4); and I have argued elsewhere (see chapter 6) that all properties of things are just truth-makers for such conditionals. But if we believe in properties F and G because they are needed (and suffice) to make conditionals like (4) true, why jib at accepting N when it is likewise needed and (with F and G) suffices to make conditionals like (3) true?

Here, however, the crucial difference between (3) and (4) emerges: (4) entails that a exists. Without a, 'a' would have no reference, and I do not see what (4) could then mean, nor how in particular it could be true. So if (4) is true, a exists; so the fact that a has mass m is always available to make (4) true. But it is by no means so clear that F and G exist whenever (3) is true of everything. The law that all F are G, it is agreed, may be vacuous: and if it is, there are no Fs. Now if, as many (including Armstrong) suppose, properties and other universals need instances, then without Fs there will be no F. But without F there will be no fact FNG to make (3) true of everything; and the problem of accounting for vacuous laws will remain unsolved.

Perhaps then universals need no instances. Concepts certainly do not: we can have the concept of a unicorn without there being unicorns. But universals are not concepts: concepts, if anything, are parts of our thought or our language; whereas universals, if anything, are parts of the world whether or not it contains any thought or language or concepts. No doubt concepts are closely related to universals, but it is not safe to assume that universals can dispense with instances just because concepts can. That remains an open question.

Tooley (1977) takes it for granted that universals need no instances, since he uses a particular example of a vacuous law to argue by elimination 'that it

must be facts about universals that serve as the truth-makers for basic laws without positive instances' (p. 672), going on to ask rhetorically: 'if facts about universals constitute the truth-makers for some laws, why shouldn't they constitute the truth-makers for all laws?' Armstrong (1978 vol. 1 p. 113), by contrast, holds a 'Principle of Instantiation: For each *N*-adic universal, *U*, there exist at least *N* particulars such that they *U*'; but he offers nonetheless to cope with Tooley's example of a vacuous law. Now if Armstrong really can supply enough universals to make vacuous laws true, without violating his principle of instantiation, we may not have to decide whether universals do in fact need instances; but can he?

In Tooley's example, as Armstrong puts it, just two out of several types of particle happen never to meet; so the law governing their interaction is vacuous. Nevertheless particles of other types meet, so that the universal, meeting (*M*), exists; as do these two mutually evasive particle types (*A* and *J*). Armstrong can claim therefore that the law, despite being vacuous, 'holds in virtue of the universals [*A*, *J*, *M*] being what they are' (p. 157). This solution, however, is only available for special cases of vacuous laws. For a start, it only works here because particles of other types do meet, thereby ensuring the existence of the universal *M*. Now if the law governing *A* and *J* particle interactions does not ensure their meeting, it can hardly ensure the meeting of other types of particles; and if *A* and *J* particles can fail to meet, so can others. If no particles ever met, the laws of all their interactions would still be true, but the universal *M* would not, for Armstrong, exist to make them so.

More seriously, not only might there be no meetings, there might be no *A* or no *J* type particles. Yet the law could still be true and important, even if there were nothing it applied to. We have seen that to be the case with Newton's first law of motion, and the vacuous instances of vapour pressure laws. For these, and indeed for the bulk of vacuous laws, Armstrong's principle of instantiation does deprive him of the universals he needs as truth-makers. The Tooley case he discusses happens to be of the only sort he can cope with, and it is worth drawing out what makes cases of this sort amenable to Armstrong's treatment: namely, that in them there exist things with properties which make certain generalisations true. These are in fact generalisations of conditionals like (4) above. Consider that for many determinate values of the determinable force *f*, (4) is vacuous: *a* can only be subjected to one (net) force at any one time, and there will be many forces *a* never experiences. Yet *a*'s always having mass *m* suffices to make all these vacuous generalisations true. And so it is with Armstrong's *A* and *J* particles. While they exist, they are disposed to interact as the law says, whether they ever

actually meet or not. Armstrong's universals *A* and *J* are just conjunctions of such dispositions (see chapter 6), and can thus be truth-makers for those laws whose vacuity results merely from the dispositions of actual things failing to display themselves. But not for the more important cases in which laws are made vacuous by the non-existence of the things themselves; and hereafter I will reserve the term 'vacuous' for such cases.

Since vacuous laws will in fact defeat the Armstrong-Dretske-Tooley account if universals need instances, we have after all to consider whether they do. I follow Ramsey (1925) in taking particulars and universals to be simply parts of facts picked out in order to generalise. For example (pp. 19–20):

> It is not '*aRb*' but '(*x*).*xRb*' which makes *Rb* prominent. In writing '(*x*).*xRb*' we use the expression '*Rb*' to collect together the set of propositions *xRb* which we want to assert to be true.

To recover a proposition from this set, we need to know what an instance of *xRb* is, i.e. we need criteria for identifying the items, such as *a*, which have been quantified over. But we do not need separate criteria to identify *Rb*. Given *a*, *Rb* is just the remainder of the fact *aRb*. If it were an independently identifiable constituent, then, as Ramsey says (p. 14), *a*(*Rb*) would differ from (*aR*)*b* and *a*(*R*)*b*, because these facts would have different constituents; and this is absurd.

Similarly, if we form the doubly general '(*x*)(*y*).*xRy*', we must regard the universal *R* as just the common part of all the facts thus collected: e.g. the fact *aRb* minus *a* and *b*. And since in forming the law that all *F*s are *G*s we at least collect whatever facts such as *Fa* there may be, the universal *F* must likewise just be the common part of all such facts. At least in laws, therefore, a universal must be regarded as derived from the particulars which are its instances and the facts that they are so. To regard them, as extreme realists do, as a primitive kind of entity, distinct from particulars but able to combine with them to yield facts, is to put the universal cart before the factual horse. It does nothing but pose such ancient but manifestly dotty conundrums as: why there are these two different kinds of entity, particulars and universals, and what the difference between them is; why two entities of the same kind (two particulars or two universals) cannot combine to form a fact; what the relation is between a particular and a universal that are so combined. The last question on its own is fatal to this view, since any answer to it immediately generates Bradley's (1897 ch. 3) notoriously vicious regress; a regress not avoided just by Armstrong's (1978 vol. 2 p. 3) ingenuous device of calling the relation in question a 'union ... closer than relation'.

The fact is, as Ramsey showed, that we have no *a priori* reason to suppose that universals are fundamentally different in kind from particulars. What we think of as particulars are merely the kinds of entity we can most readily individuate, typically by appeal to their spatio-temporal location (cf. Braithwaite 1926); and a universal is just the common residue of a set of facts about such individuals. So there is really no mystery about what relates particular to universal in a fact, nor about why a fact has to contain at least one of each. Nor is there any general reason why residual universals cannot themselves be individuated and so admitted in their own right as entities to be quantified over – thus, for example, leaving the particular *a* as the common residue of a set of facts about *a*'s properties. Nominalism therefore is not the only, nor the most sensible, alternative to an extreme realism about universals.

From this Ramseyan account of universals it does however follow that they need instances: Armstrong's principle of instantiation is quite right. In the law that all *F*s are *G*s, the property *F* is just the residue of such facts as *Fa*. If the law is vacuous, there are no such facts; and no facts leave no residue. If there are no *F*s, there is no *F*. So there will be no fact *FNG* to make such a vacuous law true, and the Armstrong-Dretske-Tooley theory fails. Whether there are 'real connections of universals', as Ramsey (1929a p. 160) put it, I do not know: like him, 'I cannot deny it; for I can understand nothing by such a phrase; what we call causal laws I find to be nothing of the sort'.

7 Epilogue I have considered two attempts, seriously undertaken of late years, to make natural laws say more than generalisations; both fail. The law that all *F*s are *G*s is given the force it needs neither by taking it to say that 'All *F*s are *G*s' is true in all possible worlds or is in some other sense necessary, nor by taking it to assert a contingent relation between *F* and *G*. Neither construal can cover the crucial case of vacuous laws which Braithwaite rightly stresses. There are no doubt likewise aspects of the problems of laws which solutions of Braithwaite's Humean cut also have difficulty covering; only they, to my mind, are more readily patched up. Those patches, however, must be woven elsewhere.

9 *Laws, chances and properties*

1 Introduction Not all laws of nature are deterministic. For example, the laws of radioactivity don't say that atoms of a radioelement *R will* decay (*D*) in any time *t*, only that each such atom has a probability $p(D)$ of doing so, where $p(D)=1-e^{-\lambda t}$ and λ is a decay constant characteristic of *R*, often re-expressed as the half-life *T* ($=\ln 2/\lambda$), the time for which $p(D)$ is exactly 0.5 (see e.g. Delaney 1962). The analysis of laws should therefore include an analysis of the probabilities they contain (probabilities which for brevity, and to distinguish them from others of arguably different kinds, I shall call 'chances'). But the analysis both of laws and of chances is controversial, and the two are rarely related. They should be, because the right account of what chances are, and of how they and other universals are involved in laws, can meet serious objections to existing accounts of laws. In particular, it yields better answers to the main questions about laws: namely, how they differ from regularities, how they make regularities in some sense necessary, and how they support conditionals that regularities don't support. The object of this paper is to derive and recommend those answers.

2 Laws and law statements First I must distinguish laws from law statements, a distinction many philosophers overlook (not all: see e.g. Armstrong 1983 ch. 1.3). Taking laws to be statements may well stem from taking them to be God's edicts, analogous to the edicts of our legislatures. But what makes a piece of legislation a law is not the edict that states it but something else (the will of God or the sanctions of a legal system) that's meant to enforce it: i.e. to make the statement of it true. What makes it a law that *Thou shalt not kill* is not that the statement 'Thou shalt not kill' is a law: that statement is called a law only to tell us that it *is* a law that Thou shalt not kill, in order to make us conform to that law. But since we have to conform to the laws of nature anyway, there's no such excuse for calling statements of them 'laws'.

I need to distinguish laws from law statements in order to distinguish predicates from properties and relations, by which I mean real universals

(Armstrong 1978 chs 1–4), i.e. non-particular constituents of atomic facts (Ramsey 1925): not concepts, or the meanings or actual or possible extensions of predicates, or sets of particulars. This is because, on my account, the law that all *G*s are *H* involves the properties *G* and *H*, not just the predicates '*G*' and '*H*'. One might of course reject the account for this reason, as nominalists will (since they don't believe in universals): but even to reject it, one must be able to state it, and hence to distinguish properties from predicates and thus laws from law statements. So by 'law' hereafter I shall mainly mean (e.g.) the law that all *G*s are *H*, not the law statement 'All *G*s are *H*' – although I may call that a law too when the difference doesn't matter or it's clear what's meant (and conversely, I shall refer for brevity to laws 'entailing' regularities, meaning that statements of them entail the corresponding generalisations).

3 Laws and regularities As laws stand to law statements, so regularities, like the fact that all actual (past, present and future) *G*s are *H*, stand to the generalisations ('All *G*s are *H*') which state them. And just as the first question about law statements is whether they *say* anything more than the generalisations they entail, so the first question about laws is whether they *are* anything more than the regularities that make those generalisations true.

This question has generated a vast literature, which I shan't even try to review. I agree with Armstrong (1983 ch. 5) that the best 'regularity' account is Lewis' (1973 ch. 3.3) development of Ramsey's (1928) view of law statements as the 'consequences of those propositions which we should take as axioms if we knew everything and organised it as simply as possible in a deductive system' (Ramsey 1929a p. 150). But I also agree with many of Armstrong's criticisms of the Ramsey–Lewis account, while finding the best non-regularity account, of Armstrong (1983 pt 2) and others (Dretske 1977, Tooley 1977, Swoyer 1982), of a law as a relation of 'nomic necessitation' between universals, unacceptable on other grounds (see chapter 8). Here, taking chances to be universals, I offer a unified account of both deterministic and indeterministic laws which, by invoking neither simplicity nor nomic necessitation, meets serious objections to both its rivals.

I start by asking how my first question applies to indeterministic laws: with what regularities might they be identified? The answer may seem obvious: for to the law that the chance of a *G* being *H* is $p(H)$ there obviously corresponds the statistical 'regularity' that the fraction $f(H)$ of *G*s that are *H* is also $p(H)$. But the trouble with this answer is that the law doesn't entail that regularity. Take radioactivity again. For any radioelement *R*, it's a law of nature that there is a time (T) in which the chance of any *R*-atom (*R* for short) decaying

naturally is exactly 0.5. But this doesn't entail that exactly half of the large but presumably finite number n of Rs in the universe at any time t_0 will decay before t_0+T. (It can't, if only because n might be an odd number.) It doesn't even entail that the fraction $f(D)_n$ of those Rs that do decay in this time will lie between any two values between 0 and 1. All it entails, via the laws of large numbers (Kingman and Taylor 1966 ch. 13.3), is that the *chance* of $f(D)_n$ differing from 0.5 by more than a very small amount is also very small, and gets smaller as n gets larger.

But for no finite n is this chance ever zero. G's half-life being T is consistent with *all* Rs decaying within time T, with *none* of them doing so, and with any fraction in between. The laws of radioactivity don't *entail* any statistical regularity at all in the fractions of actual atoms decaying in any finite time. They may *explain* why about half of many such atoms decay within their half-life: indeed, the fact that the chance of any such atom decaying in that time is 0.5 is what *causes* about half of them to decay (by making the chance of that happening greater than it would have been had the half-life been different: see chapters 12 and 14.4). But the explanation isn't deductive, and the causation isn't deterministic. And similarly for all indeterministic laws: they may explain some facts about statistical regularities, but they won't entail them. So the standard question, whether laws are *more* than regularities, seems not to apply to these laws, which aren't *even* regularities.

But in fact, as we shall see, the question does apply. Indeterministic laws do entail regularities, and differ from deterministic laws in this and related respects much less than they seem to. But to see how and why this is, we must look more closely at the chances that occur in them.

4 Chances, tendencies and possibilities Everyone agrees that chances are probabilities: i.e. that they satisfy the standard probability calculus (Kolmogorov 1933). But that doesn't tell us how to interpret the calculus in this case: it doesn't tell us what chances are.

Two interpretations may be quickly dismissed. First, *pace* de Finetti (1931) and others, chances can't be merely subjective probabilities. A G's chance $p(D)$ of decaying in time t can't be identified with anyone's degree of belief in the proposition D that it will decay (in time t, a qualification I shall hereafter take as read). Facts about the fractions of Rs that decay may, as I've said, be caused, or at least statistically explained, by $p(D)$. But they're certainly not caused or explained by how much anyone expects Rs to decay, if only because they would decay as they do even if no one existed to expect anything. So we can't identify $p(D)$ with even an expert scientist's degree of belief in D. His or

her degree of belief should of course coincide in value with the chance, since to know the chance of anything is *ipso facto* to know how much to expect it. (As I (1971) and Lewis (1980) have argued, it's analytic to the concept of chance that my degree of belief in any proposition Q, relative to the evidence that its chance of being true is $p(Q)$, should also be $p(Q)$.) But even experts aren't infallible, as they show by disagreeing and changing their minds: social phenomena which no one can really take to affect the stability of atomic nuclei. And even if, sceptical of objective truth, we defined truth in general *à la* Pierce as whatever scientists eventually agree on, that still wouldn't justify identifying $p(D)$ with their agreed degree of belief in D. The present physical decay of R-atoms can hardly be explained, causally or otherwise, by a hypothetical future mental consensus that will probably never occur. (I am indebted for this argument to an unpublished paper by J. T. Whyte, which raises a similar causal objection to the coherence theory of truth.)

Nor, for the same reason, can $p(D)$ be any kind of merely epistemic probability. For if the decay of Rs can't be caused or explained by subjective probabilities which people *do* have, it certainly can't be caused or explained by ones they *don't* have (but would have if only they satisfied the canons of some inductive logic). Of course $p(D)$ *entails* an epistemic probability, since, as I've said, knowing the chance of anything tells you how much to expect it. But then there must be more to chances than the epistemic probabilities they entail; and we can see what more by looking briefly at their rôle in decision theory.

Take the well-worn case of smoking and cancer. Suppose I'm a smoker who'd rather smoke even if he gets cancer. But smoking raises my chance of getting it so much that the expected utility of my smoking, calculated from this chance, is less than that of my giving up. So of course I should give up. But, as I've argued elsewhere (see chapter 16) against Jeffrey (1983 ch. 1) and others, I shouldn't give up if my chance of getting cancer is a merely epistemic (or subjective) probability, raised by smoking only because (e.g.) smoking and cancer are both statistical effects of a common genetic cause. For it's clearly not worth giving up just to lessen the *evidence* that I'll get cancer if that won't actually lessen my *prospects* of getting it, which is what its chance is meant to measure. But then a 'prospect' must be something more than an epistemic possibility. But what?

The obvious answer is that a prospect is a real ontological possibility, which comes by degrees and is contingent, in this case on my smoking. And that, according to Popper (1990), is precisely what chances are: degrees of contingent ontological possibility. My chance of getting cancer is the degree of

possibility that I'll get it, contingent on my smoking; and $p(D)$ is the degree of possibility that an *R*-atom will decay, contingent on facts about its nucleus.

The obvious objection to this answer is that it makes chance incompatible with determinism. For suppose it's also a law that smokers get cancer if and only if they also have some further metabolic property *A*. Then as a smoker, my chance of getting cancer will also be 1 if I'm *A* or 0 if I'm not. But if chances are possibilities, this means I have either no possibility of getting cancer or no possibility of not getting it. In neither case can my intermediate chance of getting cancer, contingent on my smoking, measure a real possibility. Whereas, assuming no such 'hidden variable' distinguishes all and only *R*-atoms that decay from those that don't, $p(D)$ can indeed measure a real ontological possibility of an *R*-atom decaying.

These and related considerations made me argue in Mellor (1971) that chance is not compatible with determinism. But I have since been persuaded by Salmon (1979) and others that it is (see chapter 14.11). The existence of the property *A* in no way impugns the probabilistic law linking smoking and cancer, nor does it make subjective or epistemic interpretations of the chance the law contains any more credible. That law still gives smokers a quantitative tendency to get cancer – *via* a tendency to have the property *A* – a tendency that measures a real prospect of getting cancer which it makes sense for them to reduce by giving up smoking.

I conclude that chances are tendencies which, in the absence of relevant determinism, are also real contingent possibilities. Moreover they are (or display: see Mellor 1971 chs 3–5) real properties of actual particulars, changes in which have real causes and effects. There is no contradiction here: on the contrary, what could make statements of ontological tendency or possibility contingently true if not real properties and relations of actual particulars? (In general, I agree with Goodman (1965 p. 57) that possibilities must lie within the actual world, not *vice versa*, though I can't argue the general point here.) In particular, what makes it worth giving up smoking, even if I'd rather not, is that doing so will cause a real change in me: namely, a reduction in my chance of getting cancer. Similarly, it may be worth the cost of assembling a critical mass of *R*s to increase their chance of decaying quickly (thus causing an atomic explosion) even though, whether they decay quickly or not, we'd rather not incur that cost. But as with giving up smoking, this makes sense only if decaying is more than a merely epistemic possibility. An *R*-atom must have a real tendency to decay, embodied in an intrinsic property of *R*s, which we can affect by bombarding them with the products of their own decay.

This view of chances as tendencies gains further support from well-known objections to frequency views of chance. We've seen already that chances can't be actual frequencies (i.e. fractions): the laws of large numbers prevent $p(D)$ being identified with the fraction $f(D)_n$ of n *R*s that do in fact decay. The only fraction $p(D)$ could be is a hypothetical one: a limit $f(D)_\infty$ to which $f(D)_n$ would tend if n increased without limit. But what could fix $f(D)_\infty$'s value? It can't be fixed by anything knowable *a priori*, because chances are contingent. Nor can it be fixed by any actual $f(D)_n$, which will always be consistent with any value of $f(D)_\infty$ from 0 to 1 inclusive: since any $f(D)_n$ would always be outweighed by an infinite number of merely possible *R*s decaying and/or an infinite number of them not decaying, which could make $f(D)_\infty$ 1 or 0 or anything in between, regardless of $f(D)_n$.

If $f(D)_\infty$ exists at all, the only thing that can fix its value is some property of an R such that there is a value to which the decaying fraction $f(D)_n$ of such *R*s tends as $n \to \infty$. But then this property just *is* the chance $p(D)$: provided $p(D)$ is indeed an intrinsic property of each R and therefore logically and epistemically independent of any facts or information about anything else. For then the laws of large numbers will give $f(D)_n$ a chance 1 of tending to $p(D)$ as $n \to \infty$.

And whether chances measure possibilities or only tendencies, this in turn will entail that $f(D)_n$ actually *will* tend to $p(D)$. For just as the least degree of possibility is impossibility and its greatest degree necessity, so the greatest degree of a tendency is inevitability. And just as nothing can occur that is in any sense impossible, so nothing can fail to occur if it is in any sense necessary or inevitable – e.g. if its chance is 1.

In general, therefore, for any proposition Q, $p(Q)=1$ must entail Q and $p(Q)=0$ must entail $\sim Q$. But this consequence of chances being tendencies poses some problems. For a start, the laws of large numbers don't entail it: for they, as we've seen, would let any number of *R*s decay even when $p(D)=f(D)_\infty=0$. Moreover it looks at first as if that must be right. For $p(D)$, the chance of an R decaying in any time interval t is $1-e^{-\lambda t}$, which for $t=0$ is 0. In other words, an R's chance of decaying at any one instant is 0. But how then, if this fact prevents it decaying at that instant, can it ever decay? Similarly with any continuous distribution of chances over quantities like temperature, pressure, momentum, position, etc. Take a spinning pointer with an equal chance of stopping in any equiangular sector of the circle it marks out. Its chance $p(\theta)$ of stopping in any sector of angle $\theta°$ is $\theta/360$, which again for $\theta=0$ is 0. In other words, the pointer has a zero chance of stopping at any given point on the circle: yet it must stop somewhere. So how then can $p(Q)=0$ entail $\sim Q$?

One apparent answer is to let atoms and pointers have infinitesimal chances of decaying or stopping at spacetime points (Skyrms 1980a appendix 4, Lewis 1980, 1986 §B). This might indeed let the chances of these mutually exclusive possibilities be non-zero without violating the calculus by adding up to more than 1. But it still contradicts the laws, $p(D)=1-e^{-\lambda t}$ and $p(\theta)=\theta/360$, which make $p(D)=0$ when $t=0$ and $p(\theta)=0$ when $\theta=0$. In short, infinitesimals are irrelevant (as Timothy Smiley has pointed out to me), since all they do is give *R*-atoms non-zero chances of decaying, and pointers of stopping, in infinitesimal regions of spacetime, not at points.

The real answer is to deny, as I (1971 ch. 6) and others (Teller 1979, Urbach 1988 §2(e)) have done on other grounds, that things can have point values of continuous quantities. An atom can't decay at just one of a continuous array of points in spacetime, any more than a pointer can stop at just one of a continuous array of points on a circle; and similarly for all other continuous quantities. Atoms can only decay, and pointers stop, in some finite (or perhaps infinitesimal) region of spacetime, and can therefore always have a finite (or at least infinitesimal) chance of doing so. This, I believe, is why nothing ever needs to happen that has no chance of happening, so that chances really can be tendencies.

5 Chances and properties But $p(D)$, an *R*'s chance of decaying in time t, is not only a tendency. On my account, it is also a real property. But what property? All I've said so far is that $p(D)$ is a property such that the fraction $f(D)_n$ of n such *R*s decaying in time t tends to the value $p(D)$ as $n\rightarrow\infty$. In other words, $p(D)$ is a disposition like solubility, a property such that, if anything is put in water while it has this property, it dissolves. But note the qualification 'while it has this property'. Calling an object a 'soluble' doesn't entail the unqualified conditional 'if a were put in water it would dissolve', since a might only be put in water if it were *in*soluble. So to call a soluble is not to assert this conditional outright, but to make an existence claim: a has a property S satisfying a description '*W*', namely that anything would dissolve if put in water *while it is S*. (This is a subjunctive version of what Carnap (1936–7 pp. 52–3) called a 'reduction sentence'.)

Similarly, to credit an R with the chance $p(D)$ is not to assert the conditional that $f(D)_n$ would tend to $p(D)$ if the number n of *R*s were to increase without limit: since those *R*s might not have that chance of decaying. It is to make an existence claim: the R has a property C satisfying a description '*J*', namely that the decaying fraction $f(D)_n$ of n *Rs with the property C* would tend to $p(D)$ if $n\rightarrow\infty$.

But the descriptions '*W*' and '*J*' may not suffice to identify the properties *S* and *C*. By this I don't mean that they must be identified with non-dispositional properties (see chapter 6), merely that '*W*' and '*J*' may need supplementing respectively by further descriptions '*X*', of the specific dispositions of *a*'s molecules that make water loosen their bonds, and '*K*', of the nuclei of *R*-atoms that make them unstable. But other substances may be made equally soluble, or radioactive, by quite different microscopic dispositions. So the properties *S* and *C*, which in *a*, and in *R*-atoms, are such that $W(S)\&X(S)$ and $J(C)\&K(C)$, may in other substances be different properties altogether. The predicates '*S*' and '*C*' are therefore not so much names of properties as abbreviated descriptions ('*W*' and '*J*') of them, which more than one property may satisfy.

'*W*' and '*J*' may still of course be satisfied by single properties in single substances, and for present purposes we may assume that they are. Indeed in the case of radioactivity we may assume more than that. For as we've seen, an *R* has not one but an infinity of chances $p(D)$ $(=1\text{-}e^{-\lambda t})$ of decaying in different times *t*, each of which generates a logically independent description '*Jt*', namely 'the fraction $f(D)_n$ of *n* *R*s with property ... decaying in time *t* tends to $1\text{-}e^{-\lambda t}$ as $n\rightarrow\infty$'. But these different descriptions needn't be satisfied by *different* properties *C*, and they're not. They are all entailed by an *R*'s having at every instant a constant chance-density λ of decaying per unit time. This property (λ for short) is a single instantaneously-possessed property satisfying all the '*Jt*'s – just as, for example, a velocity *V* is a single instantaneously-possessed property satisfying all the descriptions '*Vt*' of how far any object with that property would get in any time *t*. In short, the property *C* just *is* λ.

In taking λ to be a single property, I am not of course applying any *a priori* criterion of identity for properties. There's no such criterion for empirical universals, any more than there is for empirical particulars. All we can say *a priori* is that for *G* and *G** to be the same property, the predicates '*G*' and '*G**' must be coextensive in all possible worlds, since otherwise whether some possible particular is *G* will depend on what *G* is called, which is absurd. But that doesn't tell us much about *G*'s identity. The real work of identifying empirical universals is done by scientific theories, construed as Ramsey sentences (Ramsey 1929d, Carnap 1966 ch. 26): e.g. 'There is a property – namely λ – such that $(t)(Jt(\lambda))$.' Sentences like this are what describe, distinguish and thus individuate the empirical properties and relations whose existence they assert. And in return, as we shall see, those properties enable the statements that describe them to state laws.

6 Chance, necessity and inevitability Our indeterministic law of radioactivity does therefore correspond to a regularity after all: the regularity that all actual *R*s are always λ. But that's because, so far as λ is concerned, the law (that all *R*s are λ) is deterministic. So the difference between deterministic and indeterministic law statements is really one of content, not of form: since law statements of both kinds are (or at least entail) 100% generalisations of the form 'All αs are β'.

But how then does the content of indeterministic laws differ from that of deterministic ones? The difference may seem obvious: indeterministic laws involve chances and deterministic ones don't. But I think deterministic laws do involve chances: chances of 1 or 0. For if there can be a law that all *G*s have a chance 0.5 of being *H*, there can surely be a law that all *G*s have a chance 1 of being *H*. And I don't see how this law can really differ from the deterministic law that all *G*s are *H*.

For consider how a deterministic law is supposed to differ from its corresponding regularity, in this case the fact that all actual *G*s are *H*. Philosophers disagree about exactly what the difference is, but not that there is one, nor that part of it is (roughly) that, if it's a law that all *G*s are *H*, the corresponding regularity is not just a coincidence: that even if it's contingent, the fact that all actual *G*s are *H* is also in some sense necessary. Hence the question 'How can something be contingently necessary?', a question to which calling it 'physically' (or 'naturally') necessary, meaning simply that it's entailed by a law, clearly provides no answer at all.

But chance provides an excellent answer. As I've argued already on other grounds, to have a chance 1 is to be, if not always contingently necessary, at least inevitable. What better explanation therefore could there be of every *G*'s being *H* than that it had no chance of not being *H*? The explanation isn't question-begging, as just calling the coincidence 'physically necessary' is. Nor is it trivial: for, as we've seen, all *G*s can be *H* even if a *G*'s chance of being *H* is less than 1 (or doesn't exist at all). But then the regularity obviously *is* a coincidence, which it certainly isn't when every *G*'s chance of being *H* is 1.

I conclude therefore that the law that all *G*s are *H* is in fact the law that all *G*s have a chance 1 of being *H*; and similarly that the law that no *G*s are *H* is the law that all *G*s have a chance 0 of being *H*. In other words, I agree with Skyrms (1980a IB) that deterministic laws are just limiting cases of indeterministic ones, where the chances involved are 1 or 0. But then all laws involve chances: 1 or 0 if they're deterministic, and something in between if they're not. And conversely, the above account of the inevitability that

deterministic laws entail applies to indeterministic laws as well. For as we've seen, indeterministic laws also entail regularities: e.g. the regularity that all actual *R*s have the property λ. But now, on my account, this law itself must really be that all *R*s have a chance 1 of being λ. So the fact that every *R* has a chance $p(D)$ of decaying in time t isn't an accident either, since it has no chance of not having this chance of decaying.

But doesn't this generate an endless regress, which if not vicious, is at least wildly implausible? For I have argued that, in general, for a proposition Q to have a chance p of being true is for something to have some property C, which we may write $C_p(Q)$. Thus when an R has a chance p of decaying in a time t, the property is $C_p(D)$; when a G has a chance 1 of being H, the property is $C_1(H)$; and so on. But then a chance 1 of being $C_p(D)$ will itself be a property: $C_1(C_p(D))$, or C_1C_p for short. But this makes it a law that all *R*s are C_1C_p, hence that it's a law that they have a chance 1 of being C_1C_p, hence that it's a law that they are all $C_1C_1C_p$, and so on to infinity.

All this is true: but it doesn't generate an endless regress, vicious or otherwise. For C_1C_p, $C_1C_1C_p$, etc. needn't be *different* properties, and they aren't. Just as a single property λ can satisfy all the descriptions '*Jt*' that define C_p for all times t, so a single property $C_1\lambda$ can satisfy the descriptions that define $C_1C_1\lambda$, $C_1C_1C_1\lambda$, etc. Indeed, not only *can* it do so: its doing so is what entails (and thus explains) the fact that the chances in these other descriptions are all 1.

7 Laws, conditionals and properties So far so good. But we must go further. For laws do more than make the regularities they entail inevitable. They also support counterfactual conditionals. If it's a law that all *G*s are *H*, then not only must all the actual *G*s be *H*, it must also be true of anything, x, whether it's *G* or not, that if it *were G*, it *would* be *H*. Moreover, this would be true even if there were rather more (or different) things than there actually are: so the variable 'x' here ranges over more things than there actually are. Hence the other question about laws: what makes them support such outrageously far-reaching conditionals? How does my account answer that question?

First, as we've seen, the laws of large numbers make chances support equally far-reaching conditionals. To have a chance p of being H is to have a property $C_p(H)$ such that, if there *were n C*-things (whether there are or not), then the fraction, f_n, of them that were H would tend to p as $n\rightarrow\infty$. And if p=1, then since having a chance 1 of being H entails being H, not only will this limit be 1, so will f_n for all values of n. In other words, anything would be H

if it had the property $C_1(H)$, whether or not it actually has that property, and whether or not it even actually exists.

So far again so good; but it's still not far enough. This isn't the conditional we need. The one we need says that anything would be H if it had the property G, not if it had the property $C_1(H)$. But G and $C_1(H)$ may be the very same property. For example, consider again an R's property λ. That, I said earlier, is a property which also satisfies some description 'K' of an R's nucleus. But having a nucleus satisfying the description 'K' may be just what makes something an R. In other words, to be an R and to be λ may be to have the very same property. But then not only will all actual Rs have a chance $1\text{-}e^{-\lambda t}$ of decaying in time t, so would anything, if it were an R, whether or not it is one, and whether or not it even exists.

But now we've gone *too* far. For as I noted in **5**, if $R{=}\lambda$, the predicates 'R' and 'λ' must be coextensive in all possible worlds. But this makes the regularity, that all Rs are λ, not just inevitable or contingently necessary but necessarily necessary; and although it has been argued (e.g. by Swoyer 1982) that laws are indeed necessary in this sense, the arguments are not good (see Armstrong 1983 ch. 11) and I see no independent reason to believe it. Maybe some natural kind laws (e.g. that all water is H_2O) are necessary, though again I don't believe it, and Kripke's (1971, 1972) and Putnam's (1975) well-known arguments certainly don't show it. (See chapter 7: Kripke in particular begs the question by taking the statements of these laws to be identity statements – 'water=H_2O' – which they're not.) But even if a few laws are necessary, most are clearly not. And if the law that all Rs are λ isn't, then in some possible world (in which it isn't a law) something will be R but not λ. But then 'R' and 'λ' won't be coextensive in all possible worlds and R can't be λ.

Perhaps it can be $C_1\lambda$, since a chance of 1 is after all only a contingent necessity? Unfortunately not. For now 'R' and '$C_1\lambda$' must be coextensive in all possible worlds. But since being $C_1\lambda$ entails being λ, this also requires all Rs to be λ in all possible worlds. The law that all Rs are λ remains necessary.

How then *can* the law that all Rs are λ be contingent? Consider again the property (or properties) S (solubility) such that $W(S)$, i.e. such that anything would dissolve if put in water while S. For every S, it's a law that any object put into water (P) while S would dissolve (be Z). But that law, I've argued, is really the law that every such object has a chance 1 of being Z: i.e. has a property $C_1(Z)$ which entails that anything (actual or merely possible) with this property would also be Z. In short, every property S embodies a contingent law that certain Ps (namely Ss) are all Z. In other words, just as chances are properties embodying contingent tendencies, so the properties that

satisfy $W(S)$ embody contingent laws of solubility by making the conditionals they entail contingently true.

And as for S, so for other properties. Every (value of) inertial mass M embodies an infinity of special cases of Newton's second law of motion. That is, for every M, it's a law that any object of mass M to which any force F is applied has a chance 1 of an acceleration $A=F/M$: i.e. has a property $C_1(A)$ such that anything (actual or merely possible) with it would have that acceleration. And as for M, so for λ. Every (value of) λ embodies an infinity of special cases of the laws of radioactivity. That is, for every λ, it's a law that any object whose decay constant is λ has for any time t a chance 1 of having a chance $p(D)=1-e^{-\lambda t}$ of decaying in that time: i.e. has a property $C_1(C_p(D))$ such that anything (actual or merely possible) with it would have that chance of decaying in that time.

Yet again, however, so far so good, but not far enough. For S, M and λ are independent of the properties linked by the laws they embody. In other words, no law makes all objects dissolve when put in water, accelerate alike under any given force, or have the same chance of decaying in any given time. But suppose it did. Suppose it was a contingent law that all Ps (objects put in water) are Z (dissolve). No single property S can embody that law. For it would only be entailed by the law that all Ps are Z if they're S if it were also a law that all Ps are S; and we're no further forward. Similarly with the law that all Rs are λ. Imagine a property U such that any R would be λ if it were U. This only entails the law that all Rs are λ if it's also a law that all Rs are U. So even if there is such a property U, it only shifts the problem: it doesn't solve it. Some contingent laws can be embodied in single properties, but only if others, from which they follow, are not. Those are the laws we have yet to account for.

Let us assume therefore that these laws include the law that all Rs are λ. What, if not a single property, can make this law contingent? The answer is simple: the *pair* of properties R and λ. The law statement 'All Rs are λ' is really part of a Ramsey sentence asserting the existence of distinct properties R and λ such that, among other things, $L(R,\lambda)$ – namely, that if anything were R it would be $C_1\lambda$ – the other things including of course the fact, noted in **5**, that $(t)(Jt(\lambda))$. And as for this law, so for all contingent laws. The law that all Gs are H is made contingent by the existence of distinct properties G and H such that $L(G,H)$.

8 Laws and nomic relations But doesn't this make the law that all Gs are H a relation (L) between the universals G and H? L looks remarkably like

Armstrong's (1983) relation of nomic necessitation. But it isn't, because it isn't a universal. It may be a concept or the meaning of a predicate, but it's not a non-particular constituent of any atomic fact (see 3). Nor therefore is it a part of any law, of whose instances it would then also have to be a part (Armstrong 1983 ch. 6.6). The law that all *G*s are *H* contains only *G* and *H* (and $C_1(H)$, an addition I shall hereafter take as read), and the particulars that instantiate it are only *G* and *H*: they aren't also *L*(*G*,*H*).

For consider again the property or properties *S* that embody the laws of solubility by being, among other things, such that *W*(*S*), i.e. such that anything put in water (*P*) while it's *S* would dissolve (be *Z*). '*W*' is indeed a predicate applying to one or more properties, just as, for example, 'neither wise nor just' is a predicate applying to one or more people. But that doesn't make being neither wise nor just a constituent of any atomic fact, and it isn't (Ramsey 1925 p. 22). And nor is *W*. *W* isn't a real property of *S*, a further constituent of the law that all *S*-things dissolve in water. The law needs no such constituent: all it needs, besides *P* and *Z*, is *S* itself.

What makes it seem to need more is an apparent dichotomy between dispositions and their 'categorical bases' (e.g. Armstrong 1968 ch. 6.6), and hence between '*W*(*S*)' and some non-dispositional description '*X*(*S*)' (e.g. of molecular structure) defining the categorical property *S* that satisfies '*W*'. But a solubility law can't be embodied in a property *S* which is only such that *X*(*S*): hence the apparent need for *W* to be a further property of *S*.

But the dichotomy is false. For as I (see chapter 6) and others have argued, most if not all properties satisfy only dispositional descriptions. Take mass. As we've seen, a mass *M* just *is* a property such that (among other – e.g. gravitational – things) any object with it would be accelerated by any force *F* at *F*/*M*. This dispositional description is what identifies masses if anything does: which is why the laws of motion need no further properties to embody them. Similarly therefore for solubility, since descriptions of molecular structures can hardly be less dispositional than those of the masses they include. In short, '*W*(*S*)' is no more dispositional than any other description of *S*, and no less entailed by it. Like *M*, *S* just *is* a property such that (among other things) *W*(*S*): which is why the laws of solubility need no further property *W* to embody them.

And as for the laws of solubility, so for other contingent laws. The properties *R* and λ just *are* the properties such that, among other things, *L*(*R*,λ); just as *G* and *H* are the properties such that, among other things, *L*(*G*,*H*). That's why the laws that all *R*s are λ and that all *G*s are *H* can be embodied in those properties, and need no further universal *L*.

9 Law statements and Ramsey sentences In my view therefore, *pace* Armstrong (1983) and others, laws of nature include no relation of nomic necessitation between the universals that embody them. But my view also entails, *pace* Ramsey (1928) and Lewis (1973 ch. 3.3), that what makes law statements state laws is not that they are general axioms or theorems of the simplest deductive theory of everything. Yet it does entail that they will in fact be that, and we should see why.

Law statements, I have said, are entailed by Ramsey sentences asserting the existence of universals: R, λ and $C_1(\lambda)$ such that $L(R,\lambda)$, $(t)(Jt(\lambda))$, etc.; S, P, Z and $C_1(Z)$ such that $W(S)$, $X(S)$, etc.; G, H and $C_1(H)$ such that $L(G,H)$; and so on. Now for Ramsey (1929d p. 131), statements like '$L(G,H)$' within Ramsey sentences are not 'strictly propositions by themselves', because 'G', 'H', etc. are not predicates but predicate variables bound by an existential quantifier. But I claimed in **5** that such sentences are what identify the universals whose existence they assert. In other words, provided we include *all* the laws containing R and λ (and hence $C_1(\lambda)$), they will be *the* properties such that $L(R,\lambda)$, $(t)(Jt(\lambda))$ etc.; and similarly for all other empirical universals. The Ramsey sentence (Σ) of the conjunction of all law statements would therefore be a definite description of all the properties and relations whose existence it asserts: there is no more to them than the totality of laws they embody.

To this, however, many will object that even Σ still only asserts the *existence* of universals satisfying dispositional descriptions: it doesn't tell us what universals they are. R, λ, etc. remain properties 'we know not what'. But what don't we know about R and λ? We know *which* properties they are: Σ tells us that by distinguishing them from all the other empirical universals. Moreover, as I argued in **8**, any description that told us more *about* R and λ would not differ in kind from Σ's dispositional descriptions of them. But then, by definition, Σ already contains it. The only thing Σ doesn't tell us about R and λ is what particulars have these properties: i.e. what the extensions of 'R' and 'λ' are. But no one thinks telling us that will tell us what R and λ are. I conclude then that Σ really does identify all the universals it asserts to exist, so that 'R', 'λ', 'G', 'H', etc. really are predicates, and law statements like 'All Gs are H' really are 'propositions by themselves'.

Back to the Ramsey–Lewis account of laws. Suppose we did know all empirical facts and organised our knowledge in a deductive system. The system would have to yield: (i) all the law statements in Σ, and their consequences; and (ii) all true *singular* statements saying (e.g.) what

particulars are R, λ, G, H, etc., and *their* consequences – which would of course include all accidentally true generalisations. So law statements would indeed be the general axioms and theorems of such a system. But what gives these statements that position in the system is the fact that they state laws, not the other way round. What makes statements state laws is not that they're axioms or theorems in a deductive theory of everything, and *a fortiori* not that they're axioms or theorems in the simplest possible such theory. Being a law has nothing to do with being statable in a simple way.

I don't of course deny that simplicity is a virtue in statements. Indeed the main point of systematising any knowledge is to simplify the statement of it. We do therefore want to state laws as simply as possible, which on my account will include referring as simply as possible to the universals that embody them. But the simplest way of referring to anything is to name it, as opposed to describing it. Thus the simplest law statements are those that name the relevant universals ('R', 'λ', 'G', 'H', etc.) instead of describing them in incidental terms. And those who think laws *are* statements will certainly want them to name the universals involved, if only to avoid multiplying laws beyond necessity. Hence, I conjecture, the specious appeal of the Ramsey–Lewis criterion of simplicity for law statements and hence for laws.

10 Conclusion So much for the Ramsey–Lewis view of laws. But rejecting it doesn't entail denying that laws are regularities: that remains an open question. My own view does indeed prevent the law that all Gs are H being the regularity that all Gs are H. But might I not let it be the regularity that all Gs are $C_1(H)$, i.e. have a chance 1 of being H? No: for as we've seen, that will only entail the law's counterfactual conditionals if G is either H or $C_1(H)$ (which makes the law necessary) or if G and H are such as to embody them. But either way it takes further facts about G and H, not just the regularity, to make it a law that all Gs are H.

I conclude therefore that all laws, deterministic or indeterministic, contingent or necessary, are more than regularities. Yet I can still explain the truth in the regularity view, stripped of its questionable concept of simplicity. And in saying what makes laws more than regularities I invoke nothing *ad hoc*, as my non-regularity rivals do: no primitive notion of natural necessity, nor of relations of nomic necessitation between universals. Those notions are *ad hoc* because there is nothing more to them than what they're defined to do, namely make laws necessary and enable them to support their counterfactual conditionals. Whereas the only concepts I invoke, of chance and universals, have many other applications, independent of their rôle in my account of laws.

Finally, I accept of course that what I've said by no means answers all the questions laws of nature pose. I haven't, for example, shown how to apply my account to laws more complex than those of radioactivity. But that shouldn't be too hard to do. More seriously, I still can't explain the possibility of laws without instances, like Newton's first law of motion – a serious problem which stems from the fact that, unlike concepts, universals like G, H and $C_1(H)$ don't exist if they have no instances. (See chapter 8 and Armstrong 1983.) But then no other account of laws I know of explains that either – so the problem is not so much an objection as a challenge.

10 *Properties and predicates*

Introduction I share Armstrong's (1978, 1989) realism about universals. I agree with him that properties and relations exist, just as the particulars exist which have those properties and relations. I also agree with him (1978 Introduction) that universals are not to be understood semantically as the meanings, references or extensions of predicates. This does not of course prevent there being obvious connections between universals and predicates. For example, to every property there obviously corresponds a possible predicate applying to all and only particulars with that property. But it does not follow from this, and is not obviously true, that to every actual predicate there corresponds a single property or relation. Perhaps 'given a predicate, there may be none, one or many universals in virtue of which the predicate applies [and] given a universal, there may be none, one or many predicates which apply in virtue of that universal' (Armstrong 1978 vol. 2 p. 9). So the questions remain: how do universals relate in general to our predicates, and how in particular do they relate to what those predicates mean?[1]

In order to tackle these questions I shall make some assumptions which I shall defend only briefly if at all. Some are uncontroversial, some merely terminological. Others are more serious, but I shall still not argue for them at length because, although I do believe them, my main interest here is in what they entail.

First of all, although for brevity I shall refer only to properties, I shall take it for granted that what I say also applies *mutatis mutandis* to relations. On the other hand, I shall not assume that what I say applies to all properties, and specifically not to apparently necessary properties of abstract particulars like numbers and sets, such as the oddness of the number 3. I am interested

[1]Besides the works mentioned in the text, my answers to these questions have been considerably influenced by Austin (1939), Wilson (1982) and an unpublished paper by Greg McCulloch. Earlier versions of this paper were discussed at a Conference on Truth and Reference held at the Inter-University Center in Dubrovnik in September 1990 and at the Cambridge University Moral Sciences Club in November 1990. My subsequent revision of it has been much indebted to comments made in those discussions, especially by Alexander Bird, Jeremy Butterfield, Mike Martin, Greg McCulloch and Roger Teichman, and also to comments made later by Jeremy Butterfield, Tim Crane, Alex Oliver and Peter Weatherall.

here only in contingent properties of so-called 'concrete' particulars: i.e., roughly, particulars which have causes and/or effects and are more or less localised in space and time. 'Concrete', however, is a bad name for them, since the particulars that concern me may well include events (such as explosions) and processes (such as fires and long walks) as well as material objects (such as planets and people). Whether particulars of all these kinds exist and, if so, how they are related to each other are of course contentious questions, but fortunately not ones I need to tackle here. What matter here are contingent properties, not how many or what kinds of particulars have those properties.

Next, I take existence, and the having of properties, to be tenseless but not modal. In other words, I restrict them to the actual world, but not to the present as opposed to the past or future. This assumption too is contentious, and it does affect some of my conclusions, but only in obvious and uninteresting ways that anyone who disagrees with it can easily work out. In what follows, therefore, I shall take it for granted that the class of real people does not, for example, contain the merely possible Danish Prince Hamlet, but does contain all the human ancestors and all the as-yet-unconceived human descendants of everyone alive today.

And as for all these actual human beings, so for the property (if any) of being human which they all share. As a realist about universals, I take the actual properties of actual past, present and future particulars to exist, and to do so whether or not they ever have been or ever will be conceived of by us or by any other thinkers. That is, I reject both nominalism and conceptualism about universals, although again I shall not discuss my reasons for doing so, nor for adopting any specific version of realism about universals. What I am going to discuss is what properties there actually are – an open question for realists, just as what particulars there are is an open question for nominalists – and what if anything those properties contribute to the meanings of our predicates.

Properties I shall start by elaborating on my rejection of the obvious answer to these questions: namely, that properties just are (or are given by) the meanings of our predicates. One reason for denying this is of course that, if they were, they could not *give* our predicates their meanings, any more than particulars could give the meanings of names or other singular terms if that was all they were. But of course they're not. No one thinks the planet Mars just is, or is part of, or defined by, the meaning of the word 'Mars' which we use to refer to it. We may indeed give a referential account of that word's meaning, i.e. one which takes the planet Mars to be part or all of what the

word 'Mars' means. But what makes this a serious thesis about the meaning of that word is precisely that it takes for granted the planet's independent existence and identity: we are using the planet Mars to give the meaning of the word 'Mars', not the other way round.

Similarly with the word 'red'. A referential theory of its meaning might take the property of being red to be part or all of what the word 'red' means. But this again will be a serious thesis about the meaning of that word only if it takes for granted the property's independent existence and identity, i.e. if it uses the property to give the meaning of the predicate rather than the other way round.

We may of course reject these referential accounts. We may deny that Mars itself is any part of what our word 'Mars' means, perhaps because we think its meaning is given by a definite description (such as 'the red planet') which any planet might satisfy. But this will not make us deny Mars' existence or query its identity. There is more to Mars than its semantic rôle, and we have more than semantic reasons to believe in its existence. Indeed that is an understatement. The planet Mars does not depend on its semantic rôle at all, either for its identity or for its existence: which is why a referential account of the meaning of the word 'Mars', whatever else may be wrong with it, is neither trivial nor viciously circular.

Similarly, I maintain, for the property, if any, of being red. But not everyone will agree. Some philosophers still think that properties, unlike particulars, do depend on their semantic rôles: that a property is nothing if not all or part of what some predicate means. Unless the meaning of the predicate 'red' is, includes or entails a corresponding property, then no such property exists. And if it does exist, its identity is given by its rôle in the meaning of the predicate, not the other way round.

I, like Armstrong, disagree. I think that in this respect properties are just like planets. We have good non-semantic reasons for believing in them, and there is more to them than their semantic rôles. Indeed I think that is another understatement. A contingent universal's existence, like that of Mars, does not depend on its having any semantic rôle, and its identity does not depend on what that rôle is. Which is why referential accounts of the meanings of predicates, whatever else may be wrong with them, are also neither trivial nor viciously circular.

But what then are the non-semantic reasons for believing in contingent universals, and what, if not the meanings of predicates, fixes their identity? I take the main reasons for believing in contingent universals to be the rôles

they play in causation and in laws of nature, and those laws are what I take to give those universals their identity.

Now one might think that causation needs universals just because it is one: namely, as Davidson (1967a) and others maintain, a relation between particulars events, as in 'The explosion *caused* the fire'. I don't think that, because I think that what the causation in such cases primarily links are facts, not particulars (see chapter 12). So I would rather report this example of singular causation by saying 'There was a fire *because* there was an explosion', which represents the causation not by a predicate ('caused') but by a connective ('because'). This of course is yet another contentious claim, but again it's not one I need to defend here: since even if causation is not a universal, it will still need universals. For just as Davidson thinks that causation only links particulars with properties that make them instantiate laws of nature, so I think it only links facts which have just such properties as constituents. And if so, then causation will need universals anyway, and we need not discuss whether it itself is one.

But is this so, and if so, why? Why must causes and effects have, or contain, properties that figure in laws? I think the reason is that singular causation entails physical probabilities, or chances. Suppose for example the causation in this case is deterministic, so that in the circumstances an explosion is both sufficient and necessary for a fire. This means that in the circumstances the chance of a fire occurring is 1 with an explosion and 0 without one. But this, I have argued in chapter 9, entails that in sufficiently similar circumstances (i) anything sufficiently like the actual explosion would always produce something like the actual fire and (ii) nothing else would ever do so. And this I take to be an existential proposition, entailing that these sufficient similarities exist: in other words, that there are properties C, F and G, of which the actual circumstances, explosion and fire respectively are instances, such that it's a law of nature that in C-circumstances, all and only F-events are followed by G-events.

That, briefly, is why I think causation always instantiates laws. Again the argument is contentious, but again I needn't defend it. For all I need is its conclusion, which is much less contentious – and even that contention I shall now try to disarm by disclaiming some common but contentious claims about causation and laws which I don't accept and to which nothing here will commit me.

First, I am not committed to physicalism. Nothing I have to say about causation and laws, or about the particulars and properties involved in them,

requires them to be physical. Nor does it require them not to be. Nothing in what follows will entail either physicalism or its negation.

Next, I am not committed to causal determinism. Causation does not entail deterministic laws, because its connotations don't require causes to be either sufficient or necessary for their effects. I do think they require causes to raise their effects' chances, but they needn't raise them to 1, and they needn't raise them from 0 (see chapter 13). So although individual circumstances, causes and effects will always need some properties *C*, *F* and *G* to make them instantiate laws, those laws need not be deterministic: they need only entail, for example, that in normal circumstances (e.g. in the presence of oxygen and inflammable material and the absence of other causes of fires) fires have a greater chance of occurring when explosions do than when they don't.

Finally, I am not committed to laws having or entailing any kind of necessity, natural or otherwise – except of course in the common but trivial sense in which calling something naturally or physically necessary just means that it's entailed by a law, or has a chance of 1.

So much for what I am not committed to. What I am committed to is a distinction between laws and statements of laws (see e.g. Armstrong 1983 p. 8). This distinction is easily, and often, overlooked (e.g. by me in chapter 8): as when Humeans say that laws are just true generalisations, like the statement 'All *F*s are *G*s'. But they could equally well say that all it takes to make this generalisation state a law is the fact that all actual *F*s are *G*s, and call that fact the law. Now whether we think of laws as true statements (or sentences or propositions) or as the facts, Humean or otherwise, that make those statements true often doesn't matter, which is no doubt why the difference is often overlooked. But here it does matter, because what law statements contain are predicates (or their meanings), whereas what the facts that make those statements true contain are properties. And clearly what causation needs are the facts, with their constituent properties, not the statements with their predicates. That's why in what follows it is the facts, and not the statements, which I shall call laws.

I don't of course deny the close connection between the properties that laws contain and the predicates we use to state those laws. On the contrary, that connection is, as we shall see, much closer than it is between properties and most other predicates. But this is not because the predicates that occur in law statements define the corresponding properties. It's the other way round. For the fact is that we have no semantic (or any other *a priori*) criterion of identity for the contingent properties that laws contain, any more than we have for contingent particulars. As noted in chapter 9, the most we can say *a priori*

is this: for F and F^* to be the same property, the predicates 'F' and 'F^*' must be coextensive in all possible worlds – since otherwise whether some possible particular is F will depend on which predicate we use to say that it's F, which is absurd.

But this modest *a priori* truth won't enable us to identify F, or any other contingent property. Properties are identified *a posteriori* by scientific theories, construed as Ramsey sentences: i.e., as saying for example that *there are* properties C, F and G such that in C-circumstances all F-events have such-and-such a chance of being followed by G-events. If that statement is true, then there are such properties, and there is such a law, of which those properties are constituents. And being a constituent of some such laws is, as I have argued in chapter 9, all there is to being a property. There is no more to temperatures than the thermodynamic and other laws they occur in; no more to masses and forces than the laws of motion and of motion's gravitational and other causes; and so on. In other words, if we stated all the laws there are in a single Ramsey sentence Σ, the properties Σ would quantify over are all the properties there are.

And this means that, with one possible exception, Σ would provide a definite description of all actual contingent properties. The exception (pointed out to me by Jeremy Butterfield) would be a pair of symmetrically related properties P and P^* – like being left- and right-handed – identifiable only by ostension. Whether two properties could really differ like this without entering differently into *some* law is not clear to me; but even if they did, Σ would still quantify over both of them. So even then the contingent properties there are could still be just those Σ quantifies over; and that from now on is what I shall assume they are. So our question now becomes: how do these properties relate to the meanings of our predicates?

Predicates Suppose I see that some thing, a, is red, i.e. that the predicate 'red' applies to it. What has happened? Clearly something about a has caused me to believe this. But what? In particular, is it just the fact that a has the property of being red?

But what does this question mean? What is it for a to have the property of being red? What is it indeed for there to be such a property? If actual properties are those that our Ramsey sentence Σ quantifies over, what makes one of them the property of being red? Well, suppose that anything which anyone sees to be rightly called red always has a certain property P, and that its being P is what causes them to see that. Whether there is any such property P is, as

we shall see, a very moot point. But suppose for the moment there is. Then clearly, if anything is the property of being red, *P* is.

But how, if it is, does *P* contribute to the meaning of the word 'red'? Suppose we agree to start with that a predicate like 'red' may also be used as a singular term referring to the property (if any) that all and only the particulars it applies to share: as in 'red is a warm colour'. Then if the singular term 'red' refers to anything, it refers to *P*. So suppose it does refer to *P* – and does so even though no one knows which property *P* is, because no one knows enough laws of nature to distinguish *P* from all other properties.

Some philosophers may deny this possibility because it conflicts with what Evans (1982 ch. 4) calls Russell's Principle, namely 'that a subject cannot make a judgment about something unless he knows which object his judgment is about' (p. 89). But this principle is false. We can easily make judgments about (and hence refer to) things without anyone knowing which they are. For example, I can easily judge that it's raining now without anyone knowing what time it is, i.e. which time my token 'now' refers to. Similarly, when I measure an object's temperature *T*, I judge in advance that *T* is what my thermometer will say it is: so I refer to *T* even before I know which temperature it is, and even if my thermometer fails me and no one ever knows which it is. But if we can use 'now' and '*T*' as singular terms to refer to times and temperatures without anyone knowing which times or temperatures they are, we can certainly use 'red' as a singular term to refer to *P* without anyone knowing which property *P* is.

So let us suppose we do that. How does our use of 'red' to refer to *P* relate to our use of 'red' as a predicate? In particular, how does it relate to the predicate's extension, i.e. to the particulars it applies to. Obviously they can't *be* the property *P*, since they are many and *P* is one. Nor can *P* be the set of all *P*-things. For since *P* is a contingent property, there could be more or fewer *P*-things, and hence more or fewer red things, than there actually are. But if *P* were its own extension, and hence that of the predicate 'red', there couldn't be: so it isn't.

Indeed *P* can obviously not be any set of *P*-things, precisely because being *P* is what *makes* things members of such sets. What *P*, like any other universal, is, I maintain, is a constituent of atomic facts, like the fact that *a* is *P*. I admit of course that what this amounts to – and especially what links *a* and *P* – are hard and long-standing questions: to which I can only respond here by asserting, with Armstrong (1978 pt 2), that they do have answers, and that those who deny that facts have universal constituents face even harder questions.

But those are not the questions I want to discuss here. The question here is this. If 'red', used as a singular term, refers to *P*, and this is what makes 'red', used as a predicate, apply to all and only *P*-things, then what makes 'red' refer to *P*? In particular, if 'red' refers to *P* by having a *sense* which makes it do so, what gives 'red' that sense?

The obvious answer is that 'red' gets its sense from a kind *K* of visual sensation which *P*-things give us when they make us call them red, so that 'red' refers to the property of things which causes us to get sensations of kind *K*: namely, *P*. But although this could be how we apply the predicate 'red', it notoriously needn't be. We can learn to see when to apply the predicate 'red' without the *P*-things it applies to giving all of us sensations of the same kind. Being *P* must make a difference to how things look to us, but the difference needn't be the same for everyone. I can learn to apply the predicate 'red' by learning to associate it with whatever kind of visual sensations I get from the things which existing users tell me are called red. It is this learned use of the predicate that fixes which kind (or kinds) of sensation this will be for me, not the other way round. And what fixes this learned use, and hence the extension of the predicate 'red', is the property *P*: since instances of *P* are in fact what we learn to respond to by applying the predicate 'red'.

But this makes *P* look less like a referent than a sense, the sense of the predicate 'red': namely, that which fixes its extension. But if it is, then again it can't be necessary for us, or any authority we defer to, to know what or which sense this is. For as we've seen, *P* can fix the extension of our predicate 'red' in this way without anyone knowing which property *P* is. But this no more stops *P* being a sense than it stops it being a referent: since we no more need to know which senses our words have than we need to know which things they refer to. For suppose the sense of the predicate 'red' was in fact given by sensations of a certain kind *K*, which it certainly could be, even if it isn't. This wouldn't require us to know which kind of sensation *K* is. It would only require us to respond reliably to *K*-sensations by applying the predicate 'red' to the things that caused them. But if that's enough to make *K* the sense of 'red', then *P* can also be the sense of 'red'. For all it takes for *P* to fix the sense of 'red' is for our eyes to make us respond reliably to *P*-things by calling them red.

Now 'sense' is of course a term of philosophical art, and for some Fregean artists it takes more than this for a word to have a sense (see e.g. Evans 1982 ch. 1). But if it does, then the predicate 'red' needn't have a sense at all. Yet something about us will still fix the extension of the word as we use it: namely, our having learned to let a thing's being *P* cause us to call it red. So I prefer to

stick to the minimal sense of 'sense' as that which fixes the reference or extension of our words, and allow *P* to be the sense of our predicate 'red' even though no one knows which property *P* is.

Suppose then that some property *P* is, in this minimal sense, the sense of our predicate 'red'. How much of the meaning of the word 'red' does this fix? Not much, and certainly not enough to give us our *concept* of red. For even with words linked as closely to our perceptions as colour predicates are, there is more to understanding them than being able to apply them. To know what 'red' means it isn't enough to know when something is red. We must be able to draw some inferences from that: our predicate 'red' does have some connotations. How does *P* help to provide them?

The short answer is that it doesn't, at least not directly. *a*'s being *P* can make us call *a* red without inclining us to infer anything from that fact. For as we've seen, *P* can enable us to apply the predicate 'red' without our knowing any of the laws in which *P* figures. We needn't even know the laws of reflection that make *P* the property which causes us to call things red, let alone the laws of chemistry that determine what chemical properties will make things *P* and therefore red. *P* need not give our predicate 'red' any connotations at all.

But *P* will constrain its connotations. For we do want our inferences to preserve truth, and when we see that they don't, we give them up. So the inferences we persist in, and eventually make part of the meanings of our predicates, will mostly preserve truth – or at least, they will when their premises and conclusions can be verified by our senses. So at least the verifiable connotations of 'red' will not contradict the laws that *P* figures in. Indeed the fact that these connotations generally do preserve truth will generally follow from some of those laws. The connotations of our predicate 'red' will therefore certainly be constrained to some extent by the laws that make *P* the property it is.

But they will equally certainly not be constrained enough to make *P* part of our concept of red. For even if a single property *P* is in fact what makes us apply the predicate 'red' as we do, this fact is obviously not one of that predicate's connotations. It is no part of our concept of red that all red things share any one property (in my non-semantic sense), let alone the property *P*. And rightly so, since there need be no one such property that all red things share. For laws need not, and mostly do not, take the simple form 'In *C*-circumstances, all and only *F*s are (followed by) *G*s', where *C*, *F* and *G* are single properties. In particular, the laws on which our senses rely when we use them to apply predicates like 'red' will almost always have much more complicated antecedents. At the very least, they may easily make things need a

negation ($\sim P$), or a disjunction ($P \vee Q$), or a conjunction ($P \& Q$) of properties to make us call them red – and $\sim P$, $P \vee Q$ and $P \& Q$ will not be properties on my account, since the Ramsey sentence Σ that quantifies over P and Q will not also quantify over them.

But maybe I should allow the existence of complex properties like $\sim P$, $P \vee Q$ and $P \& Q$. For even if a and b need only be P or Q to be red, it still seems to follow from their both being red that there is something they both are. But that something can't be P or Q, since a may be P but not Q and b Q but not P. So it looks as if the complex property $P \vee Q$ must exist to be the something that both a and b are. But not so: any more than an actual person (Nobody) must exist in order to be what two empty rooms c and d both contain. The only sense in which it obviously follows that there is something that a and b both are is substitutional: they are both truly called red, just as c and d are both truly said to contain nobody. But on an objectual interpretation of the existential quantifier, it no more obviously follows that there is an actual property which a and b share than that there is an actual occupant whom c and d share.

Yet even if it doesn't follow, it may still be true. It is less obvious that there are no such properties as $\sim P$, $P \vee Q$ and $P \& Q$ than that there is no such person as Nobody. But there really are no such complex properties, as the following argument (taken from Ramsey 1925 pp. 14–15) shows. For suppose there are, i.e. that there are properties U, V and W such that $\sim P = U$, $P \vee Q = V$ and $P \& Q = W$. Then Ua and $\sim Pa$, for example, are the very same state of affairs. But they can't be, because they have different constituents: the first containing U but not P, the second P but not U. And similarly for Va and $Pa \vee Qa$, and for Wa and $Pa \& Qa$. So there are no such properties as U, V and W – which is not of course to deny the existence of the predicates 'U', 'V' and 'W'.

Armstrong (1978 ch. 15.1), however, while agreeing that there are no negative or disjunctive properties like $\sim P$ and $P \vee Q$, does think there are conjunctive ones like $P \& Q$. And conjunctive properties are indeed more credible than disjunctive and negative ones, just because they do sustain the existential inference: a and b being P and Q does entail that a and b share a property – indeed two properties, namely P and Q. But this hardly shows that they share a third property $P \& Q$. Nor does Armstrong succeed in showing that. His claim that 'it is logically and epistemically possible that all properties are conjunctive properties' (p. 32) just begs the question, while the intuition behind it (that nature may be infinitely complex, so that, for example, there may be no limit to the small-scale structure of matter) needs no complex properties: since there need be no limit to the number or complexity of laws

of nature, nor hence to the number of properties over which Σ has to quantify. And Armstrong's only other argument (p. 35) is a *non sequitur*. The fact that *Pa&Qa* may have effects which don't follow from those of *Pa* and *Qa* doesn't show that *P&Q* is a property, merely that laws of the form 'All *P&Q*s are ...' need not follow from laws of the forms 'All *P*s are ...' and 'All *Q*s are ...'.

I conclude then that there really are no complex properties, and therefore that there need be no one property that all red things share. And once we see that there need be no such property, it is obvious that in fact there isn't. For the property or properties of light that make it red will clearly differ from all the other equally different properties of objects that make them respectively reflect, transmit and emit red light. So our application of the predicate 'red' must in fact rely on at least four laws, with the same consequents – the forming of a belief that something is red – but different antecedents, involving four properties ($P_1 \ldots P_4$) of light and of reflecting, transmitting and emitting objects respectively, which in four corresponding kinds ($C_1 \ldots C_4$) of observational circumstances make them cause such a belief.

There is thus no such property as being red, i.e. no property that all red things share. No one property gives our predicate 'red' its sense in even the minimal sense of fixing its extension, let alone in any more substantial sense of fixing its connotations and hence our concept of red. Even the minimal extension-fixing sense of 'red' must be at least a disjunction, $P_1C_1 \vee \ldots \vee P_4C_4$, of four conjunctions, of the properties of the four different kinds of red things that make us call them red with those of the circumstances in which they do so. But this doesn't make the predicate 'red' ambiguous. For since, as we've seen, we needn't know what the sense of 'red' is that fixes its extension, we needn't know, or even think, that it is disjunctive. Our concept of red no more requires red things to differ in their relevant properties than it requires them to be the same.

Nor will the lack of a property that all red things share make it any harder for us to learn how to apply the predicate 'red'. It obviously won't if what makes us call things red is that they all give us sensations of some kind, even if that kind varies from person to person. For then it is the similarity of those sensations that makes us call all the different things that cause them red, not that of the properties which make those things give us those sensations. So if this is what makes us call things red, it obviously doesn't matter whether all red things share a property or not. But it doesn't matter anyway, even if this isn't what makes us call things red. We can still learn to respond reliably to instances of a complex combination of properties like $P_1C_1 \vee \ldots \vee P_4C_4$ just as

easily as to instances of a single property *P*. For as we've seen, our calling things red isn't an inference from our seeing them to be *P*, or $P_1C_1 \vee \ldots \vee P_4C_4$: it's a direct effect that we learn to let those things have on us in those circumstances. And however complex the combination of properties involved, we can learn to be affected in this way without having any prior concept either of red or of those properties. For we can learn by example, being corrected by existing users of the predicate, just as the network of parallel distributed processors described by Churchland (1988 ch. 7.5) can learn by example to use sonar to tell underwater mines from rocks. And just as Churchland's network can learn without containing any representation either of mines or of the properties by which it learns to detect them, so we can learn what to call red without having or acquiring any concept either of red or of the properties to which we learn to respond by calling the things that have them red.

Nothing therefore in our learning or use of the predicate 'red' requires it to correspond to any one property of the things we apply it to. And as for 'red', so for almost all our predicates – except some of those we use to state laws of nature. For if, as I have argued, actual properties are those quantified over by the Ramsey sentence Σ that states all laws, then predicates corresponding to Σ's predicate variables will in turn correspond to properties. And this gives us reason to think that the simple predicates we use in our law statements – e.g. those ascribing masses, temperatures, energies, chemical and biological kinds, mental states and kinds of sensation – correspond to properties. We may of course be wrong: not only because our supposed law statements may be false, but because discovering more laws may convince us that predicates we thought were simple (like 'chlorine') are really complex. But even when a predicate '*P*' really is simple, the property *P* will still not give us its connotations. For even if Σ identifies *P*, *P* doesn't identify Σ. The laws of nature are contingent: they aren't entailed by even the totality of properties they contain, let alone by any one of them. And although, as we have seen, the laws that contain *P* will somewhat constrain the connotations of the predicate '*P*', they certainly won't provide them. For not only do we not know all those laws (and probably never will), but even if we did, that would still not determine which of them we do or should build into the meanings of the predicates involved.

In short, contingent universals contribute little to the meanings even of the scientific predicates we use to identify them. And to the meanings of most other predicates they contribute even less: not even their extensions. Thus what Wittgenstein said of the predicate 'game' may well be true of every

ordinary predicate: no one property is shared by everything it applies to. But that doesn't dispose, as some have thought, of universals and hence of the problems they present: it provides no excuse for nominalism or conceptualism in either metaphysics or semantics. What it does do is show the need to recognise how much our concepts, and the meanings of our predicates, differ from the actual properties and relations of things, but also how much, and in what complex ways, they depend on them.

11 *McTaggart, fixity and coming true*

1 Introduction Some events are past, some present and some, I expect, are still to come. These are at once the most obvious, the most basic and the most disputed facts about time. I am one of those who dispute them. I maintain with McTaggart (1908, 1927 ch. 33) that in reality nothing is either past, present or future. Since, however, I part from him by thinking that reality need not be tensed to be temporal, I am not led, as he is, to deny the reality of time itself. Indeed I believe that, paradoxically, time needs to be both real and tenseless to explain how and why people come to think of events as being past, present and future.

These propositions are, I fear, still contentious, so they will have to be defended in what follows. But my main object is not merely to promote and sugar McTaggart's pill. I want also to prescribe it: specifically, for Jeffrey's (1980a) 'conceit that the world grows by accretion of facts'; or, in other words, that only when an event happens does the proposition saying so 'come true' (p. 253). It will also serve to purge Mackie's (1974) closely related conceit that events become 'fixed and settled and unalterable' (p. 178) as soon as their 'preceding sufficient causes ... have occurred' (p. 181). These are serious conceits, though not new ones: McTaggart (1927 §337) himself appeals to the second while disposing of Broad's (1923 ch. 11) version of the first. But as they have been newly reconceived, so they need renewed purgation. They are, I shall argue, only trivially true if time is tenseless. And rather than tax my distinguished colleagues with triviality, I prefer to conclude that they are wrong.

2 Time without tense First, however, we must get rid of tense, and I will not pretend that this is easy. Consider for example the fundamental relation '... is earlier than ...' (or its converse, 'later than'). What makes this relation temporal? One persuasive answer is: one event, e, being earlier than another, e', implies such tensed facts as that sometime e' is present and e past but never *vice versa*. What makes the 'earlier' relation temporal, in other words, is that it determines the order in which the events it relates become successively

present and then past. But if there are in reality no such tensed facts as events being present or past, something else must make 'earlier' temporal – and it is no easy task to find something else that will do the job. As McTaggart saw, it is not enough for a tenseless relation between events merely to reproduce the order in which they appear to become present. If, for example, everything in the universe was always at the same temperature at the same time, but always cooling, the 'cooler' relation would do that: but that would not make 'cooler' a temporal relation.

Advocates of tenseless time have, I admit, mostly shirked the task, e.g. of saying what is temporal about the non-spatial dimension of their four-dimensional Minkowski manifolds. Their 'block' universes have no more real time in them than McTaggart's does – the difference being that McTaggart sees this and they, by and large, do not. I too will shirk the task here, but I do acknowledge it, since I am not willing to give up real time, and I undertake to tackle it elsewhere (see Mellor 1981). All I can say here is that the materials I will use are the direct perception of one event being later than another, which occurs whenever we see something move or change in some other definite way, and the role causation plays in that perception.

There is another task, however, which I must attempt here: namely, to give a tenseless account of change. Time is essentially the dimension of change, and any theory of time has to account for that fact. Now McTaggart thought that change needed tense, since he thought change to be impossible without events moving from the future *via* the present to the past, a movement I shall call 'McTaggart change'. Without real tense, of course, McTaggart change does not exist, so a tenseless account of change must find a way of doing without it. My account derives from Russell (1903): 'Change is the difference, in respect of truth or falsehood, between a proposition concerning an entity and the time t, and a proposition concerning the same entity and the time t', provided that these propositions differ only by the fact that t occurs in the one where t' occurs in the other' (§442). This is what Geach (1979) has called 'Cambridge change' and, as he says (pp. 90–2), actual change is only one species of it. To adapt an example from McTaggart (1927 §309), 'the fall of a sand-castle on the English coast' effects a Cambridge change in the Great Pyramid, by changing a relation in which it stands to the sand; but clearly the Pyramid itself does not actually change as the sand does. The difference between actual and what Geach calls 'merely' Cambridge change is causal: actual changes are events, with spatio-temporally contiguous effects, and merely Cambridge changes are not (see Mellor 1981 ch. 7).

I follow Davidson (1969) in taking events, including changes, to be individuated by their causes and effects. But not all events are changes; nor do events themselves change. Change occurs in things, i.e. individual substances, in one standard sense of that term. (The difference between things and events I take to be that whereas events, if extended in time, have temporal parts, things do not. People are things in this sense, and so are common objects such as tables and chairs – and McTaggart's (1927 §§313–15) poker. For a longer list, and some reasons why the thing/event distinction matters, see Mellor 1982 §6.) A thing may have a non-temporal property at one date incompatible with those it has at earlier or later dates; and when such a fact constitutes an event, with effects spatio-temporally contiguous to the thing, the thing has undergone an actual change between these dates. We may indeed use this as a criterion for distinguishing real from merely apparent properties of things, thus ruling out such spurious properties as being forty, famous, the tallest man in the room and 'grue' (Goodman 1965 ch. 3). Real properties of things and people, loss or gain of which is actual change in them, rather than the merely Cambridge variety, include temperatures, masses, colours, shapes – and both physical dispositions such as solubility (see chapter 6.1–5), and mental states such as particular beliefs and desires (see chapter 3).

Now suppose some thing, a, has a pair, G and G^*, of such incompatible real properties (e.g. temperatures) during two separate stretches of time t and t^*: i.e.

(1) a is G during t

and

(2) a is G^* during t^*

If a were an event, it would have different temporal parts containing a-during-t and a-during-t^*, and the supposed change would reduce to these different parts having different properties:

(3) $G(a\text{-during-}t)$

and

(4) $G^*(a\text{-during-}t^*)$

But that different entities differ in their properties does not amount to change, even if one is earlier than the other and both are parts of something else. (3) and (4) would no more constitute a case of change than would a's spatial parts

differing in their properties – e.g. McTaggart's poker being hot at one end and cool at the other.

I take change to require a difference between the state of a *whole* thing at two different times. That is, real changeable non-temporal properties of a thing are in fact relations it has to the various times and stretches of times at which it exists. That is, (1) and (2) should be read as

(5) $G(a,t)$

and

(6) $G^*(a,t^*)$

Treating temperatures, colours, shapes etc. as relations between things and times may seem odd, but it is only a way of making two indubitable points about facts like (1):

(i) Both the contexts

'... is G during t'

and

'a is G during ...'

are transparent, i.e. (1) remains true however a and t are referred to.

(ii) For (1) to be true, both a and t must exist. (This need not of course imply a Newtonian conception of absolute time: it does not follow that time could exist without events – times may still need specifying by events, such as Christ's birth, and their temporal relations, such as the earth's period of rotation on its axis and about the sun.)

I should emphasise at once that (5) and (6) in no way beg the question against tenses. Nothing prevents t and t^* taking tensed values like 'yesterday' and 'tomorrow' as well as tenseless ones like '9 January 1994' and '10 January 1994'. Nor do (5) and (6) conflict with the use of sentential operators which Prior's work has made usual in tense-logic; i.e., in this case,

(7) During t, Ga

and

(8) During t^*, G^*a

On the contrary, a relational reading of tensed facts is standardly used to supply 'semantics' for these operators (see e.g. McArthur 1976 ch. 1.3). In other words, even tense-logicians take (7) and (8), with appropriately tensed t

and t^*, to be made true by the corresponding relational facts as stated in (5) and (6).

However, as an advocate of tenseless time, I will restrict t and t^* to tenseless values. Change, I maintain, consists in a thing's having a real non-temporal property at one date which it lacks at others, i.e. respectively having and lacking, to those dates, the corresponding real non-temporal relation.

McTaggart would not agree; but not because he disputes – he does not draw – my distinctions between things and events and between actual and merely Cambridge change. For McTaggart, (1) and (2), however construed, would not constitute change because they are themselves unchanging facts about *a*. His poker being 'hot on a particular Monday' and cool thereafter (McTaggart 1927 §315) is no change in it, he says, since it always was and always will be a fact that it is hot that Monday and cool thereafter. And as McTaggart says, neither this 'nor any other fact about the poker change[s], unless its presentness, pastness, or futurity change'. McTaggart change, in other words, is the only kind of change tenseless facts are capable of. But why, in order for a change to be a fact, must that fact also change? I see no reason to believe it must, nor hence any good argument from real change to McTaggart change and hence real tense. We can quite well deny both, and still insist that McTaggart's poker changes as it cools.

3 Tenses and dates We can, I believe, account for time and change without real tense: but why should we try to? Because real tense implies McTaggart change, and that, as he showed, is a myth – the 'myth of passage' as it has been called (Williams 1951). But it is a very powerful myth, and undoubtedly expresses something real and important about time. As the persistent rejection of McTaggart's own sound and simple disproof of it shows, its grip will not be broken until something better is put in its place. In what follows, therefore, I shall put up a tenseless surrogate for it; to which end, I must first lay down more precisely the specification the surrogate has to satisfy.

The myth of time passing, i.e. of McTaggart change, combines two ways of locating events in time: by their dates, and by their temporal distance, past or future, from the present. These two ways locate events in two series of temporal positions which McTaggart called the '*B* series' and the '*A* series' respectively. McTaggart change consists in the relative motion of these two series. Events of given date become less future or more past, as the present time moves from earlier to later dates.

(There may in fact be several *A* and *B* series. In both, events get the same location just in case they are simultaneous; and relativity theory may make the

simultaneity of distant events depend, within causal limits, on an arbitrary choice of a so-called 'reference frame' to settle what is to count as being at rest. Physical fact may fail to settle that question: so different but equally good reference frames may make quite different celestial events simultaneous with the terrestrial events of 1 January 1994, for example, thus filling that *B* series position quite differently. But the same goes for the *A* series: whatever celestial events get that terrestrial date will *ipso facto* then count as temporally present. So there is, as McTaggart (1927 §323) conjectured, a distinct *A* series corresponding to each distinct *B* series. For present purposes, however, I can afford to ignore these relativistic complications, since I am concerned only with the apparent relative movement of corresponding *A* and *B* series. In referring to 'the' *A* and *B* series, then, I shall henceforth mean any relativistically acceptable *B* series, and the *A* series corresponding to it.)

Positions in the *B* series I shall call 'dates', stretching that term to cover locations of all sizes from nanoseconds to millennia. Thus B.C. is a date, and so is the first p.m. second of 1 January 1994. (Events have any date that includes all their temporal parts, just as things have any spatial location that includes all their spatial parts. Thus, the end of World War 2 has, *inter alia*, the dates A.D., the twentieth century and 1945, just as London has the locations Earth, Europe and England. When I refer to 'the' date of an event, I mean the shortest date that includes all its temporal parts.) Dates may be regarded as intervals of *B* series instants, such as noon precisely on 1 January 1994, ordered by the 'earlier' relation. I do not of course mean by this that instants exist: if there are any such things, they will be spatio-temporal entities – spacetime points – not purely temporal ones. Instants are no more than convenient theoretical devices for generating indefinitely divisible systems for dating events.

Positions in the *A* series I shall reluctantly follow custom and call 'tenses', though they are mostly marked, not by verbal inflection but by adverbs and phrases such as 'today', 'ten days hence' and 'last year'; and given these, verbal tenses are redundant – 'last year' already implies the past tense, as 'today' implies the present. Tense in the sense of *A* series position must therefore be sharply distinguished from verbal tense, which is merely one very crude way of marking it; and the former, not the latter, is what I shall mean by 'tense' hereafter unless I explicitly say otherwise.

Tenses, like dates in the *B* series, may be regarded as intervals of instants, and these are likewise ordered by the 'earlier' relation. McTaggart (1927 §305) characterises the *B* series as ordered by 'earlier', as opposed to the *A* series, which is ordered by degrees of pastness or futurity; but this is a false

contrast. 'Earlier' orders both series. *Ten days ago*, which is an *A* series position, is earlier than today in just the same sense in which 1 January is ten days earlier than 11 January. In fact, the *A* and *B* series have exactly the same temporal structure. They use the very same 'earlier' relation to order the very same collections of simultaneous events. Fix which *B* series instant is the *A* series present instant, and either series is immediately definable in terms of the other. To every *B* series instant there then corresponds the *A* series instant which is that much earlier or later than the present instant; and hence to every date (i.e. interval of *B* series instants) there corresponds a tense, and *vice versa*. Thus, when it is now noon on 1 January 1994, 10 a.m. is two hours past, 11 January is ten days hence, and the next century is the twenty-first.

Seeing that the *A* and *B* series are so similar, and so simply interdefinable, what is the difference between them? The difference is that whereas an event's dates are fixed, its tenses are not. By this I mean that its tenses vary with time (this of course being just what McTaggart change is), and its dates do not. Suppose for example that it is now May 1994 and the Queen is 68. That is, she was born 68 years ago; in other words, that event has the tense: 68 years past. The tense of this event obviously varies with time: in 1984, the Queen was 58 years old; in 2004, she will be 78. Note that the event's tense varies just the same if the time itself is reckoned by tense rather than by date: thus, ten years ago, the Queen's birth was 58 years past; ten years hence, it will be 78 years past. These facts, of course, follow from each other, the general study of such temporal entailments being the business of so-called 'tense-logic'. The reason there is no comparable 'date-logic' is simply that an event's dates, unlike its tenses, do not vary with time, whether the time be reckoned in tenses or dates. The fact now, in May 1994, is that the Queen was born on 21 April 1926; and that always was and always will be the date of her birth. (Some indeed think that before 1926, when the Queen's birth was future, it did not yet exist, and so had no date at all. But no one thinks it ever had, or ever will have, any date other than 21 April 1926.) Date-logic, then, is not studied, because it is too simple. Temporal operators, be they dated or tensed, and however they are iterated, have no effect at all on the classical truth value (if any) of '*e* occurs at *t*'.

Dates, unlike tenses, are outright, temporally unqualified properties of events. That is the essential characteristic of the *B*, as opposed to the *A*, series – and why, provided tenseless sense can be made of 'earlier', it is the fundamental series. The *B* series is definable as the definite temporal structure of all the world's events (on a relational view of time), or of all instants (on an absolute view). The *A* series is neither: it has to be defined in terms of the *B*

series plus a present instant. And the present instant has to move: there has to be McTaggart change, or the *A* series would bc identical with the *B* series. Past, present and future, therefore, as aspects of reality, stand or fall with McTaggart change. They fall – as we shall see in the course of constructing something tenseless to put in their place. But first let us look at the reasons that support them.

4 Tensed truth, tenseless fact There are two chief reasons for believing in real tense, and in particular in a real present. One is experiential, the other linguistic. The former is what many take to be an irreducible experience of events being present as they happen to us (or, in the case of actions, as we perform them); in other words, its sheer presentness seems to be an undeniable part of our every experience. A credible surrogate is needed for this. To produce it, however, I must first dispose of the latter, linguistic reason for believing in real tense: namely, that our judgments about the tenses of events are generally either objectively true or objectively false, and real tenses are needed to make them so. In May 1994, for example, it is objectively true to think or say that the Queen is 68. What makes that true seems to be that she is then 68, i.e. that at that date her birth really does have the tense: 68 years past. But if reality has no tense, there is no such fact, and we must give this indisputably objective judgment alternative tenseless truth conditions. And once that has been done, explaining away the apparent presentness of our experience will turn out to pose no great problem.

The truth conditions I need are really quite obvious, and also quite indisputable. Even if events have tenses, it turns out that these have nothing to do with making what I shall call 'tensed judgments' about them true or false. The truth value of a tensed judgment is determined entirely by how much earlier (or later) it is than the event it is about. A judgment that the Queen is *N* years old, for example, is objectively true just in case its date is between *N* and *N*+1 years later than that of her birth. It is quite immaterial whether the Queen's birth, or the tensed judgment about it, is past, present or future.

The truth conditions of all tensed judgments are fixed in reality by dates. A present tense judgment is true if, and only if, it differs no more in date from the event it is about than the span of tense it ascribes to that event. E.g. '*e* occurs today' is true just in case it is said or thought on the same day as *e*; '*e* occurs this week' just in case it is said or thought the same week. Past and future tense judgments are true if and only it they have dates as much later or earlier respectively than the events they are about as the tenses they ascribe to them are than the present. Tensed judgments can of course be more complex

than the simple ascription of an *A* series position to an event. There are, for example, the judgments commonly expressed in English by verbal tenses such as the future perfect. But the truth conditions of these too are fixed by how much earlier or later their dates are than those of events they are about and other dates definable from these. 'Next year the Queen will have reigned 43 years', for instance, is true just in case the Queen is still Queen the year after that judgment is made, and that year is 43 years later than her accession. And similarly for tensed judgments of any complexity. The real usefulness, indeed, of the standard 'semantics' of tense-logic referred to in **2**, is that it shows how to derive any tensed judgment's truth conditions from its date in this sort of way.

Dates are not only sufficient to fix the truth conditions of tensed judgments, they are also necessary. Suppose a tensed judgment, e.g. that the Queen is 68, had no date – being, perhaps, one of God's judgments if, as some have said, He is 'outside time'. What could make it true? Not that the Queen really is 68 when the judgment is made: for, given that the Queen was born in 1926, that gives the judgment a date, namely 1994. Without a date, in short, a tensed judgment has no definite truth conditions; and with one, its truth conditions contain no tenses. These facts seem to me to make the idea of real tense not merely redundant, but incredible. Try to suppose that there really is in 1994 such a tensed fact as that the Queen is 68. This supposed fact turns out to be no part of what makes the corresponding judgment true: what does that job is simply the date of the Queen's birth being 68 years earlier. Now a fact which has nothing to do with making any tensed judgment true is surely no tensed fact. But these supposed facts are by definition tensed. Yet in reality no such supposedly tensed facts make any tensed judgment true. So I conclude that in reality there are no such facts: there is no real *A* series, and therefore no McTaggart change.

5 Experience and indexicals But what then of our experience of tense and of McTaggart change? Tenseless truth conditions seem not to dispose of that. Consider Prior's famous example: 'Thank goodness that's over!', said after a painful experience (Prior 1959). 'That's over' is indeed true if and only if said or thought later than whatever experience the 'that' refers to. But why thank goodness for such a tenseless fact, which could be recognised as such at any time, before or during, as well as after, the pain in question: surely the thanks are given in sheer relief for the pain's becoming past and thereby ceasing to be present?

Not necessarily. 'Thank goodness' certainly expresses relief, and is thus appropriately said or thought just when relief is appropriately felt. But when is that? Prior says it is when a pain is past, as opposed to present or future; whereas I say it is just after the pain, as opposed to during or before it. I cannot see that Prior's tensed account of when relief is appropriate is any better than my tenseless one. And mine does make sense of the whole remark: since 'thank goodness', said of a pain, is appropriate just when 'that's over', said of it, is true, it is always right to say both (or neither) at the same time.

This account of Prior's case gives the clue to a tenseless analysis of the apparent presentness of experience. Like his case, it involves self-consciousness; only here one is making tensed judgments of experience as it occurs, rather than afterwards. Now simultaneity with its subject matter is the defining truth condition of a present tense judgment, as opposed to a past or future tense one; so if I am thinking of my actions or experiences as happening *while* I am thinking of them, I am *ipso facto* thinking of them as being present. And that, I suggest, is all there is to the much vaunted presentness of our experience. Experiences in themselves, like events of every other kind, are neither past, present nor future. It is only our simultaneous consciousness of them, as being simultaneous, which necessarily both has, and satisfies, the tenseless truth conditions of present tense judgments.

Our being trapped for ever in the present is not a profound metaphysical constraint on our temporal location: it is a trivial consequence of the essential indexicality of tensed judgment. It is like everyone being condemned to be himself and, wherever he is, to being – as he sees it – here. The judgments 'I am *X*' and 'Here is *Y*', made respectively by person *X* and at place *Y*, are as objectively and inevitably true for all *X* and *Y* as 'It is now *T*', made at time *T*, is for all *T*: but not because *X* and *Y* have respectively such real properties as 'being me' and 'being here'. Obviously there are no such personal and spatial equivalents of our supposed tensed facts; and if there were, they would, like tensed facts, be no part of what makes the corresponding judgments true. 'I am *X*' is true if and only if *X* judges it; 'Here is *Y*' is true if and only if it is judged at *Y*. So anyone who judges, of the place he is at, that it is here, is bound to be right, wherever he is; and similarly, *mutatis mutandis*, for judgments of one's own first person identity. That is all the inescapability of being oneself and being here amounts to: and so it is with the inescapability of the present.

I conclude that neither our experience of time nor the objective truth of tensed judgments requires, or indeed admits of, real tense. Tensed judgments are simply a kind of indexical judgment, with tenseless truth conditions. But

this does not mean either that tensed judgments themselves are really tenseless, or that we could do without them. Tense may not be an aspect of the world; but, as Perry (1979) has shown, it is, like personal and spatial indexicality, an irreducible and indispensable aspect of our thought.

That a tensed judgment is not equivalent to any tenseless one is easily seen. If it were, it would be equivalent to the tenseless judgment that its own truth conditions obtain. For example, a particular judgment *J*, a token of the type 'It is now *T*', is true if and only if it is made at *T*. Let *J'* be the tenseless judgment that this is so, i.e. '*J* is made at *T*'. *J* is true if and only if *J'* is. But they are not the same judgment. In particular, if *J'* is true at all, it is true whenever it is made, whereas *J* is only true at *T*.

In other words, as upholders of tense have rightly insisted, tensed truths cannot be translated into tenseless ones. Neither the sentence type 'It is now *T*', nor Prior's 'Thank goodness that's over', nor any other tensed sentence type, means the same as any tenseless sentence. That is because tensed sentence types are indexical: it is part of their meaning that the truth conditions of their tokens vary with time, which is not true of tokens of tenseless types. But there is no tense in the truth conditions themselves; just as the truth conditions of tokens of 'Here is *Y*' are (literally!) neither here nor there, despite its being different from any non-indexical spatial judgment.

Not only is indexical judgment untranslatable, it is also indispensable. To suppose that we could make do with a tenseless language is as much a mirage as is real tense itself. Suppose I want to do something at *T*. Some change in my state of mind is needed to prompt me to act at *T* rather than some other time. The change of course is my coming to judge 'It is now *T*', where before I judged, 'It is not yet *T*'. And for this kind of change of tensed belief there is no tenseless substitute. Because the truth value of tenseless beliefs does not change with time, mere lapse of time gives no cause to change them. But it does give us cause to change our tensed beliefs if we are to keep them true, which it is the object of all our belief to be. And these changes, especially changes of belief from the future to the present tense, are the immediate and indispensable causes of our actions. Whether they cause us to act in time is of course another matter: our mental clocks are as fallible as any others. But without them, i.e. without making tensed judgments, we should have no cause to act at all.

This is my surrogate for the myth of passage: the tensed judgments we need to have, and therefore continually to change, in order to be capable of timely action. This is the truth behind the myth. The error is to misread the tense of these judgments as part of their non-indexical content, and hence to see it as an

extra, ever-changing aspect of the objective world. Having exposed the error, we may hope at last to break the myth, and begin to repair the havoc it has wreaked in the philosophy of time.

6 Fixity and coming true Tense has not wreaked all its havoc under its own name. Jeffrey's conceit, of propositions about events 'coming true' as the events happen, is stated explicitly in tenseless terms; and Mackie's, of events acquiring 'fixity', easily can be. Nonetheless, these specious happenings are nothing if not kinds of McTaggart change. Without real tense they are trivial; and with it, impossible, as I will now attempt to show.

Jeffrey gives events no tenses, only dates; but says that before the date of an event its happening is no fact. In other words, the corresponding tenseless proposition is not then true; though it may be 'ineluctable', if its 'final truth' is determined by the facts to date. As time goes on, therefore, propositions come true, and the number of facts increases: 'the world grows by accretion of facts'. What is wrong with this picture?

For a start, Jeffrey's use of 'true' and 'finally true'. In calling a tenseless proposition 'finally true', he means what most of us would mean by calling it plain 'true'. At any rate, what he calls 'final truth' is what our tenseless judgments aim at, and that is what matters. Given his 'final truth', what he calls 'truth' is entirely immaterial. Suppose I do not know whether the third cricket Test Match in a (current) 1994 Australian series has finished yet, and so am unsure what tense to give my judgment that England win it. My judgment still has a perfectly definite tenseless content, and attains its object provided England do win, whether they have done so yet or not. That question, whose answer decides whether my judgment is 'true' in Jeffrey's sense, is of no interest to me whatever: 'final truth' is all I am after.

More seriously, suppose that at the end of 1994 I make some tenseless judgment about an event (picked out by a non-temporal description) that happens in a distant galaxy after the light I see left it and before its reflection would return there. If that event is as I judge it to be, my judgment attains its object: it is 'finally true'. Whether, for Jeffrey, it is also 'true' depends on the event's date not being later than 1994, which, according to relativity, may be a matter of an arbitrary choice of reference frame (see **3** above): a matter which concerns me not at all, and is certainly not one I can credit with marking the boundaries of objective fact (see Mellor 1974: this objection is not met by the modification Jeffrey proposes in his fn. 1, p. 259).

I propose to restore 'true' to its customary and proper use, for the intended attribute of all our judgments, tensed and tenseless alike. That is, I shall call

'true' what Jeffrey calls 'finally true'. So I need another term for what he calls 'true'. Since he applies the term to tenseless propositions just when it should be applied to the corresponding past and present tense ones, I shall take the liberty of saying instead that they have 'come to pass'.

I have no objection to Jeffrey's use of 'ineluctable'. By it he means 'necessary', in the sense in which 'it is necessary that p is true if the present state of affairs makes it certain that the p-event will occur, or again if the p-event has already occurred' (Ackrill 1963 p. 139). The peculiarity of this sense of 'necessary' (in which, for example, p entails its own necessity) is quite enough to justify Jeffrey's preference for 'ineluctable'. It is also what Mackie (1974 ch. 7) means in ascribing 'fixity' to past and present events and the future events they determine. Ackrill and Mackie put the matter in tensed terms, but that, as Jeffrey shows, is by no means essential: an event is 'fixed', we may say, only on and after the date it, or an earlier sufficient cause of it, happens. The tenseless proposition that it happens is likewise 'ineluctable' only on and after the date it, or some other true proposition that determines its truth, 'comes to pass'.

Events therefore, and true tenseless propositions about them, are credited with the ability to undergo at least two sorts of change: (i) the events happen and the propositions come to pass, and (ii) then or earlier the events become fixed and the propositions ineluctable. What sort of sense can be made of these supposed changes? Tenseless propositions, after all, are normally thought to be unchanging; and while in **2** I have admitted that some events *are* changes, I have denied that events themselves change. Nevertheless, sense can be made of (i) and (ii) – only not, as we shall see in **7**, a sense sufficient for their authors' needs.

Suppose an event e happens at date T. Let H be the property of having happened, and let t and t^* be any dates entirely earlier or later respectively than (every temporal part of) e. Then the change (i) consists in the facts that

(9) e is $\sim H$ during t

and

(10) e is H during t^*

for all t and t^*.

Do (9) and (10) constitute a change in the sense of **2**? Certainly, even though e itself is an event and not a thing, (9) and (10) do not reduce to any difference between temporal parts. The parts that would be required, e-during-t and e-during-t^*, are not parts of e, since t is by definition earlier than

every temporal part of *e*, and *t** is later. They would have to be parts of some *ersatz e*-thing, say *E*, which changes from being ~*H* to being *H*. But since *t* is *any* date before *e*, and *t** any date after it, *E* would have to span the whole history of the world (except perhaps when *e* itself is). And in reality there are obviously no such things. An everlasting whole of which World-War-2-during-5000-B.C., and World-War-II-during-20,000-A.D. are temporal parts, for example, is not a credible substitute for World War 2 itself.

So (9) and (10) must be read along the lines of (5) and (6), not (3) and (4): i.e. as

(11) ~*H*(*e*,*t*)

and

(12) *H*(*e*,*t**)

H is thus some relation that any event *e* has to every date later than itself, but lacks to any earlier date. The relation is, of course, a familiar one: 'earlier' is its common name! For an event to 'happen' at a date is simply for it to be earlier than all later dates, and later than all earlier ones.

Now this is not of course a change in *e*, as it would be were *H* a real non-temporal relation. Instances of (5) and (6) are indeed taken to imply that *a*'s temporal location includes both *t* and *t**: it exists at both dates, and at some time between them changes from being *G* to being *G**. But (11) and (12) imply no such thing about *e*: on the contrary, they imply that *e* is *not* located at *t* and *t**, or it would not be later and earlier respectively than those dates. So *e* is not an everlasting thing, existing during all the dates *t* and *t** and changing at *T* in respect of having happened. Put like that, I dare say no one thinks it is. But there is evidently a recurrent temptation to harbour an equivalent thought: namely, that *e*'s happening is another event, distinct from *e*, and constituting some sort of change in it. Not so: *e* is all there is, and talk of it happening at *T* is just a way of saying that *T* is its date, its temporal location – i.e. that *e* is later than all times earlier than *T* and earlier than all later times.

T being *e*'s date is also all there is to the proposition which says that *T* is *e*'s date 'coming to pass' (and hence all other true propositions about *e* doing so). Let *p* be this true tenseless proposition, and *C* be the supposed property of having come to pass (i.e. of being 'true' in Jeffrey's eccentric sense). As before, *t* and *t** are any dates earlier and later respectively than *e*. Then the facts are that

(13) *p* is ~*C* during *t*

and

(14) p is C during t^*

Now Jeffrey in effect construes (13) along the lines of (3) and (4), not (5) and (6); i.e. he credits propositions with temporal parts:

(15) $\sim C(p\text{-during-}t)$

and

(16) $C(p\text{-during-}t^*)$

Once these temporal parts have come to pass, Jeffrey accumulates them into what he calls 'stages': 'Stages do duty (in the formal mode of speech) for all the facts so far' (p. 253). We may reconstruct his stages from (15) and (16) as follows. For any t (before or after e), let p-through-t be the whole whose temporal parts are p-during-t' for all dates t' containing no instants later than t. Let C^* be the property such that p-through-t is C^* if and only if some temporal part of it is C. Then for any given t, the conjunction of all C^* p-through-t is the stage of the world at t's last instant. In other words, as true tenseless propositions come to pass, they become parts of all later stages of the world.

The mundane facts behind this formal farrago are actually more visible in the relational reading of (13) and (14):

(17) $\sim C(p,t)$

and

(18) $C(p,t^*)$

Tenseless propositions, unlike events, admittedly have no dates; so C cannot just be the 'earlier' relation, i.e. H. But H suffices to define it:

(19) $C(p,t^*) =_{df} H(e,t^*)$

In other words, e's being earlier than t^* is the fact that makes p have come to pass at that date. p's coming to pass at T, like e's happening then, is in reality nothing more than T being e's date.

So much for (i); what of (ii)? (ii) in fact depends on (i), since fixity depends on events happening, ineluctability on propositions coming to pass. The mere happening of an event fixes it, if the earlier happening of a sufficient cause has not already done so. And no event is fixed until it, or some preceding sufficient cause of it, has happened. Now we have seen that for an

event to have happened by a certain date is simply for it to be earlier than that date. The supposed property, H, of having happened is in reality just the 'earlier' relation between events and dates. The supposed property, F, of being fixed is likewise in reality a relation events have to dates: a relation entailed by H but not entailing it, since the earlier happening of a sufficient cause may fix an event before it happens. F is thus definable by H, and by the relation S (= 'is a sufficient cause of'):

$$(20) \quad F(e,t) =_{df} H(e,t) \vee (\exists e^*)(H(e^*,t) \& S(e^*,e))$$

As for events, so for propositions. A proposition's coming to pass suffices to make it ineluctable, if it has not already been made so by the earlier coming to pass of a proposition that determines its truth. And no proposition is ineluctable until it, or some such determining proposition, has come to pass. The parallel between propositions and events here is obvious and exact. By definition, p becomes ineluctable just when e becomes fixed: i.e.

$$(21) \quad I(p,t) =_{df} F(e,t)$$

so the reality of ineluctability is just that of fixity, *viz* the conditions given in (20). All that fixity and ineluctability need are events, their dates, and the tenseless relations 'earlier' and 'sufficient cause'. (And if, as Hume thought, there is in reality no such relation as S, the second disjunct of (20) is always false, and both fixity and ineluctability reduce to events happening, i.e. to their having dates.)

7 Fixity, coming true and tense I have given the simple relational conditions of happening, coming to pass, being fixed and being ineluctable. These conditions are undeniable, but they will hardly satisfy the authors of these conceits. Jeffrey (1980a), for example, is trying to conceive the world as 'growing by accretion of facts'. But the reality of his accretion turns out to be nothing more than the truism that the later a date is, the more events are earlier than it. There is no growth in that fact, any more than there is shrinkage in the fact that the earlier a date is, the more events are later than it. Jeffrey must be after something more.

So must Mackie (1974). He hopes to find 'in this notion of fixity a basis for the concept of causal priority' (p. 183). Specifically, causes are distinguished by being fixed at times when their effects are not, but not conversely (p. 180). Since events are fixed at the latest when they happen, this is supposed to explain why causes mostly precede their effects (the exception being later causes fixed before their effects by the still earlier happening of sufficient

causes of them). But for this to be an explanation, fixity must not itself be defined by the very fact Mackie wants to derive from it. But in (20) it is. When two causally related events e and e' have no preceding sufficient causes, e is fixed when e' is not just in case e is earlier than e'. So Mackie's definition of causal priority reduces in this case to the cause being the earlier of two causally related events, which is just what he is trying to explain. And when e does have sufficient causes, the arbitrary restriction in (20)'s second disjunct, to e*s earlier than e, likewise begs the question it is supposed to answer. Later events, after all, exist no less than earlier ones, and are as capable of being sufficient causes of e. If any are, the restriction in (20) discriminates without reason against them; and if none are, it is superfluous.

The fact is that Mackie's theory, like Jeffrey's, is useless and trivial unless having happened and being fixed are something more than the relations I have reduced them to. H and F must be real non-relational properties of events, acquired at times that are their, or their sufficient causes', dates, for the facts of causal priority to be explained by them. And similarly for C and I, the coming to pass and becoming ineluctable of Jeffrey's true tenseless propositions. Real accretion must be more than a relational fact: more, at any rate, than the different relations events have to different dates. Can we meet these seemingly modest demands?

Whatever these non-relational properties H, F, C and I are, their ascription will still have to satisfy the relational conditions I have stated. Maybe 'earlier', as a relation between events and dates, should be defined by 'has happened' rather than *vice versa*: but either way, their equivalence must follow. And even if (19), (20) and (21) will not do as definitions, they must still come out as necessary truths.

What this comes to is that, for example, any judgment to the effect that an event e has the property H must come out true just in case e is not later than the date of the judgment itself. But this is to say that the judgment is indexical: specifically, that its truth conditions are those of the simultaneous judgment that e is past or present. In other words, the non-relational property H simply is that rather imprecise tense: to have happened is to be either past or present.

Ascriptions of fixity are indexical in a slightly more complex way. A judgment that an event e has the property F is true if and only if its date is not earlier than e or some sufficient cause of e. For e to be fixed, therefore, is just for it, or a sufficient cause of it, to be past or present.

Jeffrey's properties C and I likewise turn out to depend on tense, despite his tenseless pretensions. If p says that e's date is T, I judge truly that p is C if and only if I do so no earlier than e itself. So for p to have come to pass is for

e to be past or present. Similarly, for *p* to be ineluctable, either *e* or a sufficient cause of *e* must be past or present.

Mackie and Jeffrey thus both require events to have positions in McTaggart's *A* series, and the changes they postulate are a species of McTaggart change. Events happen and become fixed, propositions come to pass and become ineluctable, as the tense of events changes from future to present. Jeffrey's world growing by accretion of facts is Broad's (1923 ch. 11) world growing by accretion of present facts.

We can now, therefore, use the results of **2–5** to extract the truth in Mackie's and Jeffrey's conceits from their error. The truth is that non-relational ascriptions of *H*, *F*, *C* and *I*, because they are indexical, do not mean the same as non-indexical statements of the relational facts to which I have reduced them. A judgment *J*, that *e* is *H*, is never the same as the simultaneous judgment *J'*, that *e* is earlier than *J*. Yet they both have the same truth conditions, namely those stated by *J'*. And such truth conditions consist entirely of events, including judgments, having dates and being more or less earlier than, or simultaneous with, each other. In the real world that makes these judgments objectively true or false, the non-relational *H*, *F*, *C* and *I* do not figure at all. Because there is in reality no tense, so there is no real happening of events (apart from the events themselves) and no acquisition of fixity by them; no coming to pass, or becoming ineluctable, of true tenseless propositions.

Fixity, then, since it does not exist, cannot be the real basis of causal priority, nor can the world really grow by accretion of facts. In their intended substance, these conceits will have to go. Still, they will go in good company. Three quarters of a century after McTaggart demolished them, much writing, in many areas of philosophy, still appeals to real, non-relational non-indexical differences between past, present and future. All of that will have to go too. But not from here; despatching so great a multitude of errors must be matter for another place.

12 *The singularly affecting facts of causation*

1 Few now agree with Russell (1912) that modern science can do without the notion of cause. It may tell us that causes need only make their effects more probable, not determine them, but that just extends the notion. That there are causes is a hypothesis science still has need of, if only to make sense of the experiments by which it is tested and the technologies in which it is applied.

Physicalists will thus seek a physical basis, such as energy transfer, for causation itself (see e.g. Castañeda 1984). The basis may in turn need causes, but that need be no problem. Energy transfers, for example, could always be linked by other energy transfers to their causes and effects. Or the basis could vary from case to case, as the molecular basis of solubility and other dispositions does. Causation can always have a basis without always having the same basis.

But what basis? That looks like an empirical question, as it does for dispositions: finding out what makes salt soluble is a job for scientists, not philosophers. Before seeking bases for dispositions, however, we must ask if they need bases, and if so of what kind. Those are philosophical questions, whose answers turn on what dispositions are. Broad's (1925 ch. 2) dispositions need microstructural bases; Ryle's (1949 ch. 5) need nothing; and Goodman's (1965 ch. 2) need events. The bases, if any, that we ascribe to dispositions depend on how we conceive them (see chapter 6).

Similarly with causation. The search for its bases, if any, must be guided by a conception of it, and may be misguided by a misconception. One such misconception is Davidson's (1967a) influential thesis that singular causation relates events rather than facts. Suppose for example Don's climbing rope breaks half way up a cliff, he falls 200 metres onto rocks, and dies instantly as a result. Davidson would say that the causes and effects here are the events, Don's fall and his death, not the facts, that Don fell and that he died. I say that is wrong. Don's fall did cause his death, but only because Don died because he fell. Causation relates those events only because it relates those facts; and most causation relates facts without relating events at all. Why that is so, why it

matters, and how it affects what causation is and what its bases may be, are what this chapter is about.

2 First the questions need clarifying. For a start, they only concern singular cases of causation like the one above, not, for example, the claim that falls like Don's are a general cause of death. Whatever singular causes and effects are, their general counterparts (falling, dying) will be the same: neither events nor facts but properties (see Papineau 1986, Sober 1986). So we need not discuss general causation, and by 'causation' hereafter I shall mean the singular kind.

Next I must say what I take events and facts to be. 'Event' especially is now a term of art, and different artists use it differently. I, like Davidson, take events like Don's fall and his death to be particulars: entities picked out by names, definite descriptions or other singular terms. In that respect they are just like people and things – like Don himself, for example, and his hat.

Facts by contrast make whole sentences, statements, thoughts or propositions true. This is not, I should say, a correspondence theory of truth: facts are defined in terms of truth, not *vice versa*. Facts might perhaps be further analysed, but not here. All that matters here is that facts are not events, because they are not particulars of any kind.

Facts and particulars are, I believe, causation's only relata. They certainly encompass their obvious rivals. States of affairs, for example, like Don's falling and his dying, must obtain in order to be causes and effects: that is, the corresponding propositions, that Don falls and that he dies, must be true. In other words, those states of affairs must be facts (see Taylor 1985 ch. 2). So, *pace* Sanford (1985 p. 290), must 'event aspects', for example those emphasised in 'Don's falling *fast* caused him to die *instantly*', which says that Don's death was instant because his fall was fast (see **7** below). This indeed differs from saying that Don died because he fell, but the cause and effect are still facts: the facts that Don fell fast, and that he died instantly. Aspects of events, which are properties if they are general, are facts if they are singular. And so, I submit, are all causes and effects that are not particulars: what else could they be?

But events, I agree, are particulars. The events – Don's fall and his death – differ from the facts that he fell and that he died. So it is a real question which causation links, and if both, how the links are related. That in turn will depend on how such facts and such events are related, a relation which I agree with Davidson (1967b p. 135) that Ramsey (1927 p. 37) rightly stated thus:

> 'That Caesar died' is really an existential proposition, asserting the existence of an event of a certain sort, thus resembling 'Italy

> has a king', which asserts the existence of a man of a certain sort. The event which is of that sort is called the death of Caesar, and should no more be confused with the fact that Caesar died than the King of Italy should be confused with the fact that Italy has a king.

Having thus prevented both confusions, we can return to our question, which is not whether causation relates events as opposed to people and other things, but whether it relates particulars of either kind. It is put in terms of events only because there are too few things like Don and his hat to go round. Causes and effects can all be particulars only if events like Don's fall and his death are also particulars, and some would deny Davidson's thesis by denying that. But not I: I deny only that particular events supply the primary relata of causation; and thinking that they do is not the only reason for believing in such events (see Davidson 1967b, Mellor 1981 ch. 8).

I believe causes and effects include both facts and particular events. But causes and effects are, confusingly, also called events by some authors for whom they are really facts (e.g. Kim 1973, Taylor 1985). Now these in themselves are just rival uses of 'event' which I could simply disclaim. But backing them are identity criteria, for causally related entities like Don's fall and his death, which make them facts, and which need rebutting; not because better criteria make them particulars, but because such entities, whether facts or events, neither have nor need any special criterion of identity. However, since Davidson (1969), Kim (1973) and many others think they do, I must digress to say why they don't.

In denying particular events a special identity criterion I am not saying that they are just like things, still less that both things and events are 'individuated to perfection by spatiotemporal coextensiveness' (Quine 1985 p. 168). Nor do I agree with Davidson (1985b p. 176) that events differ from things merely in grammar. The grammatical distinction has a basis in reality: events, when extended in time, have temporal parts, and things do not (see chapter 11.1 and Mellor 1981 ch. 7). That is what lets events be changes in things (Davidson 1970a, 1985b), and thus be the actions (Davidson 1967b) and the other happenings that we associate with verbs.

But this does not make events 'a fundamental ontological category' (Davidson 1969 p. 180). Fundamental to grammar it may be, but not to ontology. Maybe, *pace* Ramsey (1925), particulars and universals form two fundamental categories, and maybe concrete particulars and abstract ones like numbers and sets form two more. But not concrete particulars with and without temporal parts. At any rate, the 'category' of events needs no criterion

of identity to defend it. How could it, when the paradigm 'category' of things has none? Things after all range from quarks, through molecules, cells, organisms, societies, mountains and planets, to clusters of galaxies. There is no one way to individuate particular things of all these kinds: each kind has its own criterion. And so do the equally multifarious kinds of particular events, ranging from quantum events, through molecular and cellular interactions, births, wars and avalanches, to explosions of supernovae. And as Quine (1985 p. 168) says, events like explosions are as well individuated by their criteria as things like mountains are by theirs: one cannot deny particularity to events on that score without also denying it to things.

There are criteria of identity for events of all kinds, but they are not all the same. And they do not distinguish events collectively from things. What does distinguish events – their having temporal parts – is not a criterion of identity. Nor is Davidson's (1969 p. 179) thesis that events are distinct (i) when, and (ii) only when, some of their causes or effects are – a thesis that does not even distinguish events from facts. For (i) is just the 'diversity of discernibles': entities of any kind that differ in some cause or effect are distinct, because all entities are distinct that differ in any way. And (ii) is as likely to be true of things and facts with causes and effects as it is of events: how, after all, would we distinguish causal entities of any kind that do not differ in cause or effect, when we only perceive anything *via* its effects?

Davidson's 'criterion' in short is as trite and as ineffective as it is needless. But some of its rivals are all too effective, because they stop events being particulars at all. Thus Kim (1973), who calls causes and effects 'events', defines an event as a thing having a property at a time. But that cannot individuate *particular* events, because it makes events facts, by making them make sentences true: namely, those ascribing properties to things at times. Hence Kim's variant use of 'event' for a kind of fact. His use is not of course improper, 'event' being as it is a term of art. But it is misleading, because the word has also long meant a kind of particular, not only in Davidson but way back, long before my 1927 Ramsey quotation above. And for authors now to apply 'event' with both connotations to causes and effects, while disputing whether they are particulars or facts, makes the word seriously ambiguous. But not hopeless, provided each of us says what he means by it. And what I mean by it is what Davidson means: my events, like his, are all particulars.

3 I will take it then that Don's fall and his death are particular events, if only to give Davidson a run for his money. The question is whether causation relates those events or relates the facts that Don fell and then he died. But how

do we decide? English usage is no guide: 'Don's fall caused his death' and 'Don died because he fell' are equally idiomatic. Most statements of causation sound as well in either form:

(1) '*c* causes *e*',

which represents cause and effect as events, referred to by the singular terms '*c*' and '*e*'; or

(2) '*E* because *C*',

in which sentences '*C*' and '*E*' state the cause and effect – and state them as facts, since '*E* because *C*' always entails both of them. (Everyone will agree that Don cannot have died because he fell unless it is a fact that he fell and a fact that he died.)

For the present (1) and (2) will be my canonical forms of causal statement, representing causal relata respectively as events and facts. But before using them in my argument I must defend them, (2) especially, against some common but misplaced objections. The obvious objection is that many causal statements are not of either form, like the example in **2** above: 'Don's falling fast caused him to die instantly'. But we can always recast them to fit, as we saw in that case, simply to show what we take their relata to be: events or facts. We must just see that the fit is not too procrustean, that the recasting begs no relevant questions.

Some recasting must be in order. For example, since we see causation – like everything else – only when it is past, we habitually report it as past: 'Don's fall *caused* his death'; 'Don *died* because he *fell*'. But we are considering causation at all times, not just past ones. So in (1) and (2) I assume an atemporal present tense, making the canonical forms of our examples:

(3) 'Don's fall *causes* his death';
(4) 'Don *dies* because he *falls*'.

These forms beg no questions about causation's temporal implications: it remains an open question, for example, whether (1) and (2) respectively entail '*c* precedes *e*' and '*E* after *C*' (see **8** below).

But I go further, as I did in **2**. I count as type (2) many statements seemingly of mixed type, like 'Don's falling caused him to die', or of type (1), like 'Don's fast fall caused his instant death'. In such cases even Davidson (1967a p. 162n) agrees with Vendler (1962) that the singular terms are really 'occurrences of verb-nominalizations that are fact-like or propositional', and admits that 'caused' is being used as a connective. But when it is being used as

a connective, that is what it is. These statements really *are* type (2): trivial variants of 'Don dies because he falls' and 'Don dies instantly because he falls fast'.

Some recasting I do resist, especially of my type (2) into Davidson's (1967a p. 151)

(5) 'The fact that *C* caused it to be the case that *E*'.

(5) is not only turgid, it misrepresents the thesis that causation relates facts. The thesis is not that causation relates entities referred to by singular terms like 'the fact that *C*': that makes facts look like a kind of particular, which is what is being denied. It is that causes and effects are reported by sentences like '*C*' and '*E*', and – as is undeniable – that for (2) or (5) to be true, '*C*' and '*E*' must be true. The facts that (2) or (5) say are related are whatever make '*C*' and '*E*' true; and any account of facts which accounts for truth will account for them. Causal facts need no special criterion of identity, such as Kim's in **2**, to justify (5)'s gratuitous reference to them. Why should they, when (as we saw in **2**) events have none, and when none is needed, either for events or for causal facts, to make them identical just when their causes and effects are?

The only appeal of (5) is that it reports nothing but causation, whereas statements of type (2) are also used to report proofs (for example, 'there is no greater prime number, because if there were there would be a greater one, because ...') and generally to give explanations, not all of which are causal. 'Because' is not an exclusively causal connective.

Why should it be? We are not trying to define a *causal* connective, only to say whether causation is reported by *a* connective. And if it is, 'because' is the obvious one. It clearly has a causal use, some of whose rules are easy enough to state: for instance, that in a causal '*E* because *C*' (as opposed to a proof), '*C*' and '*E*' must be logically independent and thus contingent. No doubt there is more to it than this, but we need not say what. All we need is the fact that 'because' has a distinctive causal use. Given that, we can restrict type (2) statements to causal ones by fiat: which I hereby do. And that, given (5)'s other defects, makes (2) the incontestably canonical way to report causal relations between facts.

4 However, statements of type (2) cannot be used to report causal relations between facts if, as Davidson claims, there are none to report. But what else might such statements be used to do? Davidson (1967a p. 162) says they are used to give causal *explanations*: the causal connective 'is best expressed by the words "causally explains"'. That is, his type (5) is really:

(6) 'The fact that *C* causally explains the fact that *E*'.

For Davidson, facts are the relata, not of causation itself, but of causal explanation.

Now we may all distinguish causation from causal explanation; but to explain away causal statements of type (2), Davidson needs more than a distinction: he needs a dichotomy. His causal *explanantia* can neither be nor correspond to causes, nor his causal *explananda* to effects. This is hard to believe. That causes generally explain their effects is a strong connotation of causation – one which explains why we use the explanatory connective 'because' to report it, and why (5), while differing from (6), can yet entail

(7) 'The fact that *C* explains the fact that *E*'.

Denying that causal explanations report causal relations makes a mystery of all this, and that counts against Davidson's thesis.

There is worse to come. Let us ask what makes (6) differ from *non*-causal instances of (7): that is, what makes (6) a *causal* explanation? Davidson cannot appeal to the causal relation (which I claim (6) reports) between the fact that *C* and the fact that *E*. For him, the causation must come from a causal relation between particulars *c* and *e* that are somehow suitably related to these facts.

In our example, the particulars, and their relation to the explanatory facts, are obvious enough. If Don dies [*E*] because Don falls [*C*], that explanation can be made causal by Don's fall [*c*] causing his death [*e*]. '*C*' here is related to *c*, and '*E*' to *e*, as 'an existential proposition asserting the existence of an event of a certain sort' is related to 'the event which is of that sort' (Ramsey 1927) – the sorts being respectively falls by Don and deaths of Don. The causation linking *c* and *e*, which by being of these sorts verify the existential '*C*' and '*E*', is what for Davidson makes '*E* because *C*' a causal explanation.

So far, so good for Davidson. But '*C*' and '*E*' need not assert the existence of events. They may deny it. Suppose Don managed to hang on when his rope broke, and so did *not* die, because he did *not* fall. That would be as causal an explanation as 'Don dies because he falls'. But '*C*' and '*E*' now assert that no falls by Don or deaths of Don exist. They are negative existential statements, verified by the *non*-existence of such events, and *a fortiori* of causally related ones. Where is the relation between events that Davidson needs to make this explanation causal?

We could of course recast the explanation to remove the negation in it, for example:

(8) 'Don survives because he hangs on'.

This seems to generate causally related events: Don's hanging on and his surviving. But these are really negative events, existing only by definition, when real events do not, to be the particular relata of causation that Davidson needs and would otherwise lack. Don's survival is simply his non-death; his hanging on, his non-fall.

And negative events cannot be particulars, because there *are* no negative particulars. Although the non-existence of negative people and things is well recognised (see Ramsey 1925, Dummett 1973 ch. 4), that of events is less so – doubtless because, as we saw in **2**, many so-called 'events' are not particulars at all. But Davidson's are, and he can no more have negative events than negative people, as I will show using Ramsey's comparison from **2**.

First, people. The existential statement 'Italy has a married King' entails 'Italy has a King', because no one can be both a King of Italy and married without being a King of Italy. However, the entailments of negative existential statements go the other way: 'Italy has no King' entails 'Italy has no married King' – *and* 'Italy has no unmarried King'. The reason is obvious: if no particular person is King of Italy, no married one is, and no unmarried one is. But now suppose 'Italy has no King' is made true by a single 'negative person', the 'non-King' of Italy, who exists just when Italy has no King. To make 'Italy has no King' entail both 'Italy has no married King' and 'Italy has no unmarried King', the non-King will have to be both married and unmarried. But he cannot be both; so he does not exist.

Similarly for events. The existential statement 'Don dies instantly' entails 'Don dies', because no event can be both a death of Don and instant without being a death of Don (Davidson 1967b). But here too the entailments of negative existential statements go the other way: 'Don does not die' entails both 'Don does not die instantly' – *and* 'Don does not die slowly'. Again, the reason is obvious: if no particular event is Don's death, then no instant one is, and no slow one is. But now suppose 'Don does not die' is made true by a single 'negative event', Don's non-death, which exists just when Don is not dying. To make 'Don does not die' entail both 'Don does not die instantly' and 'Don does not die slowly', Don's non-death will have to be both instant and slow; but it cannot be both, so it does not exist.

In short, if deaths and falls are particulars, non-deaths and non-falls are not. One could of course debate which the particulars are – deaths and falls, or survivings and hangings on – but it would not help. For either way only one of the explanations:

(4) 'Don dies because he falls';
(8) 'Don survives because he hangs on';

can be made causal by a causal relation between particular events. Yet each looks as causal as the other. It is most implausible to say that if (4) is a causal statement, (8) cannot be, and *vice versa*. But that is the inevitable consequence of limiting causes and effects to events.

The causal parity evident in many pairs of cases like (4) and (8) has elicited many of the literature's rival relata, including those referred to in **2**. Thus Kim (1973) can easily let both (4) and (8) be causal. Dying and surviving are both properties Don can have, so his having either at a time counts for Kim as a causal event. But then, as we saw, Kim's 'events' are really facts, not particulars, and so are Sanford's (1985) 'event aspects'. And that is what lets them cope with (4) and (8).

Both (4) and (8) can report causal relations between facts because 'negative facts' are still facts. Whenever '*C*' and '*E*', entailed by a causal

(2) '*E* because *C*',

are false, '~*C*' and '~*E*' are true; and the facts they state may then be causally related and

(2′) '~*E* because ~*C*'

may be true.

(2) and (2′) cannot of course be true together, but only because '*C*&*E*' and '~*C*&~*E*' cannot be. Nothing stops (2) being true when '*C*' and '*E*' are true, and (2′) being true when they are false. Thus in our example, though (4) and (8) cannot both be true, (4) may well be true if Don falls and dies, and (8) if he hangs on and survives – and both can be *causal*: that is, made true, when they are, by causal relations between the facts they entail. It makes no odds which of those facts is 'positive' and which 'negative'. Whether (4) or (8) is the instance of (2) (or there is no fact of that matter), the causation will be the same. As a type of causal statement, (2′) is no different from (2).

The ability of (2) to make (4) and (8) equally causal strongly suggests that causation relates facts as well as events. But a notorious argument of Davidson (1967a pp. 152–3) purports to prove that it cannot. We must now see what is wrong with that argument.

5 Davidson's proof starts from a truism, that '*E* because *C*' cannot be a complete truth function of '*E*' and '*C*'. It is of course, as we've seen, a partial truth function: false whenever '*C*' or '*E*' is false. But when '*C*' and '*E*' are true, it is neither always true nor always false. If it were, either all facts would be causally related or none would be, and no one who thinks causation

relates some facts thinks it relates them all. '*E* because *C*' must be true for some true '*C*' and '*E*' and false for others, the real problem of causation being to say for which. But, Davidson argues, a *causal* '*E* because C' would *have* to be a complete truth function of '*C*' and '*E*'; but it cannot be. So there is no such thing: '*E* because *C*' is not the 'logical form' of causal statements.

Davidson's argument fails to show that a causal '*E* because *C*' must be truth-functional, because at least one of his two assumptions is false. These assumptions are that a true causal '*E* because *C*' would never be falsified by replacing either (i) '*C*' or '*E*' by logically equivalent statements, or (ii) singular terms by others referring to the same particulars. And Davidson also assumes (p. 149) that since (i) and (ii) concern only the logical form of causal statements, they are independent of how causation itself is analysed.

That is his first mistake. It is a logician's conceit that causation's logical form can be fixed without analysis. How could it be? How, if not by analysis, can the logic of a concept be uncovered? In particular, consider how probabilistic analyses of causation may affect Davidson's assumption (i). Many of us (e.g. Cartwright 1979, Skyrms 1980a, Suppes 1984, Salmon 1984) appeal to probability to say for which true '*C*' and '*E*' '*E* because *C*' is true, and I think a necessary condition is this:

> (*P*) '*E*' must be more probable (that is, more probably true) than it would be in the circumstances if '*C*' were false.

I shall not try to prove (*P*) here (see chapter 13), but it at least shows what an analysis of causation can yield; and if it is true, (i) will be true only if logical equivalents never differ in probability. But that will depend on whether the probability is subjective or objective. I dare say the objective probabilities of logical equivalents never differ. But subjective ones do, because subjective probability measures the strength of our beliefs; and unless I *believe* that two statements are equivalent, I may well believe one more strongly than the other.

So the truth of (i) may well turn on the kind of probability (if any) that causes give their effects, and that depends on the analysis of causation and probability. Many analysts (e.g. Ramsey 1929a, Blackburn 1980) argue, like Hume, for subjective analyses that would make (i) false. I disagree: I think causation needs objective probabilities for several reasons (I give one in 7), and that (i) may well be true. But it is not just a question of logical form.

However, (ii) is false, even as a matter of form, whatever kind of probability causation needs. Davidson's only argument for it is that in two cases replacing one singular term by another referring to the same particular

does not falsify a causal statement, but that hardly shows it *never* does. And, as Timothy Smiley has made me see, when '*C*' or '*E*' in '*E* because *C*' is a contingent identity statement, of the form (say) 'the *F* is the *G*', it must do. For since '*E* because *C*' entails '*C*' and '*E*', it entails here that 'the *F*' and 'the *G*' refer to the same particular. So if (ii) were true, replacing either by the other in a true '*E* because *C*' would not falsify it. But it would, by generating (for instance) 'the *F* is the *F* because *C*', which is false, since 'the *F* is the *F*' is a necessary truth, and necessary facts have neither causes nor effects.

But do real causal statements ever take this form? Well, suppose in our example that several climbers fall, but that Don falls first because his rope is the weakest. That is,

(9) 'Don's fall is the first fall because Don's rope is the weakest rope'

is true. It follows that Don's is indeed the first fall, and that his rope is the weakest. So 'Don's fall' and 'the first fall' refer to the same event, and 'Don's rope' and 'the weakest rope' refer to the same thing. So for (ii) to be true, all the following would have to be true:

(10) 'Don's fall is Don's fall because his rope is the weakest';
(11) 'The first fall is the first fall because Don's rope is the weakest';
(12) 'Don falls first because his rope is his rope';
(13) 'Don falls first because the weakest rope is the weakest rope'.

Yet (10) to (13) are clearly false.

I cannot believe that (9) and many other such cases, which the reader can easily think of, are not causal. Nor could Davidson account for them by calling them causal explanations. For they, like (8), cannot be made causal by a causal relation between particulars. (9) admittedly, unlike (8), does refer to particulars, but we still cannot get from it a relevant truth of type (1). In

'Don's rope being the weakest caused his fall to be the first',

for example, 'Don's rope being the weakest' is a nominalised sentence, not a singular term. It does more than refer to Don's rope: it asserts in the context that his rope is the weakest. And so it must, for what (9) says is that each of two particulars satisfies two given descriptions – and that one does so because the other does. That is why this causal claim depends for its truth on how these

particulars are referred to, which a report of a relation between them would not do.

All this, incidentally, is explained by the probability condition (*P*), since the opacity of (9) is a feature of subjective and objective probabilities alike. (*P*) says that, if (9) is true, the probability of Don's falling first must be greater than it would be in the circumstances if his was not the weakest rope; and that is doubtless true. But the probability of Don's fall being Don's fall, and of the first fall being the first fall, will be 1 whether Don's is the weakest rope or not: (10) and (11) fail condition (*P*). And so do (12) and (13), because (*P*) makes their truth depend on the probability *p* of Don's falling first if (say) his rope were not his rope, being less than it actually is (say 0.9). Now the status of such counterfactuals, with logically impossible antecedents, is debatable: they may be true for all values of *p* (Lewis 1973 p. 16), or false, or have no truth value. But anyway *p* will have no unique value, or range of values, less than 0.9 or even than 1.

The opacity of (9), and (*P*)'s explanation of it, do more than refute Davidson's assumption (ii). They show where and how his argument for the truth-functionality of '*E* because *C*' fails. The argument is this: (i) '*E* because *C*' entails any statement got by replacing '*E*' (or '*C*') by any logical equivalent. But '*E*' is logically equivalent to '$\{x{:}x{=}x\&E\}{=}\{x{:}x{=}x\}$', because '$\{x{:}x{=}x\&E\}$' refers to the set of everything, $\{x{:}x{=}x\}$, just when '*E*' is true. So '*E* because *C*' entails '$\{x{:}x{=}x\&E\}{=}\{x{:}x{=}x\}$ because *C*'; but (ii) this entails any statement got from it by replacing a singular term by another referring to the same particular. But for any '*T*' sharing '*E*''s truth value, '$\{x{:}x{=}x\&T\}$' and '$\{x{:}x{=}x\&E\}$' refer to the same particular: namely, the set of everything if '*E*' is true, and the null set if not. So '*E* because *C*' entails '$\{x{:}x{=}x\&T\}{=}\{x{:}x{=}x\}$ because *C*' and hence, applying (i) again, '*T* because *C*'. And since the argument applies equally to replacements for '*C*', '*E* because *C*' entails '*T* because *S*' for all '*T*' and '*S*' that share '*E*''s and '*C*''s truth values. So a causal 'because' would be a complete truth function – which it cannot be.

The falsity of (ii) here, and thus the argument's invalidity, is made obvious by the analogue of (9). For even if (i) is true, and '*E* because *C*' entails '$\{x{:}x{=}x\&E\}{=}\{x{:}x{=}x\}$ because *C*', that no more entails '$\{x{:}x{=}x\}{=}\{x{:}x{=}x\}$ because *C*' than (9) entails (11). The fact that the set of everything is the set of everything no more has a cause (or an effect) than do the facts that Don's fall is his fall and his rope is his rope – and for the same probabilistic reasons.

The falsity of (ii) moreover lets many contingently true substitutes for '*E*' and '*C*' falsify a true '$\{x{:}x{=}x\&E\}{=}\{x{:}x{=}x\}$ because *C*' or a true '*E* because $\{x{:}x{=}x\&C\}{=}\{x{:}x{=}x\}$', and again (*P*) explains why. Probabilities are not truth

functions – truths are not all equally probable, even if logical equivalents are – and nor are counterfactuals. A contingent *T* may share '*E*''s truth and still differ from '*E*' by being no more probable than it would be if '*C*' were false. But then replacing '*E*' by '*T*' in '*E* because *C*' and '$\{x{:}x{=}x\&E\}{=}\{x{:}x{=}x\}$ because *C*' will make them fail condition (*P*). Likewise, a contingent '*S*' may share '*C*''s truth and still differ from it in that '*E*' would be no less probable if '*S*' were false. Again, replacing '*C*' by '*S*' in '*E* because *C*' and '*E* because $\{x{:}x{=}x\&C\}{=}\{x{:}x{=}x\}$' will make them fail condition (*P*). And this, I submit, is why replacing 'Don dies' or 'Don falls' by most other truths in a true 'Don dies because Don falls' only generates falsehoods, like 'Kim lives because Don falls' or 'Don dies because Kim doesn't fall.' The reason is that 'Kim lives' would be no less probable if Don didn't fall, and 'Don dies' no less probable if Kim did.

All these cases are grist to a probabilist's mill. And every argument for causation needing some kind of probability reinforces type (2)'s claim to be the primary form of causal statements. Probability of any kind is probability of truth, which means it can attach to facts, but not to particulars. Bearers of probability, such as a tossed coin landing heads up, are admittedly often called 'events', but they are not particulars, any more than Kim's 'events' in **2** were. They correspond not to singular terms but to sentences ('the coin lands heads up') which when true state facts. If effects are to have probabilities, causation not only may be reported by type (2) statements: it must be.

However, I need not argue that case here. Even if nothing like (*P*) is true, the falsity of (ii), evident in Davidson's own argument, refutes him by showing that a causal '*E* because *C*' need not be the complete truth function we know it cannot be. And then (9) unites with (8) to show, regardless of probability, that facts can be causally related when particular events are not. So not all the relata of singular causation are particular events.

6 But some are. There are truths of type (1), '*c* causes *e*', as well as of type (2), and when the two go together, as in (3) and (4), they are not independent. It is no coincidence that causation links Don's fall to his death just when it links the fact that he falls to the fact that he dies. If (4)'s causation does not reduce to (3)'s, (3)'s must reduce to (4)'s; but how? How can relations between events reduce to relations between facts? We saw in **5** that some type (2) statements are opaque: their truth depends on how they refer to their particulars. But type (1) statements cannot be opaque. A relation between particulars must hold however its relata are described, or it would not relate *them*. So the statement that it holds must be transparent, that is, true whatever

terms are used to refer to its relata. And that is true of type (1) statements: '*c* causes *e*' is not made false by replacing '*c*' or '*e*' by other terms referring to *c* or *e*. If Don's fall causes his death and he is the oldest and fattest climber, then the oldest climber's fall causes the fattest one's death. Thus (3) is transparent, because it reports a relation between particulars.

How then could (3) follow from (4)? The problem here is not the argument disposed of in **5**. That could not apply to (3), which has no constituent statements of whose truth its own could be a function. Transparency is not a problem in type (1) statements – it just makes it hard for them to follow from statements of type (2), and mostly they don't follow: most type (2) truths, like (8) and (9), generate no truths of type (1). However, some do: namely those, like (4), whose constituents '*C*' and '*E*' are existential statements about events. But that makes their opacity harmless, because despite entailing the existence of events of certain sorts, they do not refer to them; and this is what lets transparent truths of type (1), about those events, follow.

To see how this works, consider (4), 'Don dies because Don falls'. 'Don falls' and 'Don dies' do not refer to the events, Don's fall and his death, which (3) says are related as cause to effect. So (4) cannot be made false by replacing terms referring to these events – because there are no such terms. What (4) entails is merely that events of specific sorts – falls by Don [*F*] and deaths of Don [*G*] – exist, and we may take it that there is only one of each sort: one fall by Don and one death of Don. Call these unique *F* and *G* events '*f*' and '*g*'. Then the following are true:

(14) 'There is a *G* event because there is an *F* event';
(15) '*f* = the *F* event';
(16) '*g* = the *G* event'.

And I claim that (14)–(16) entail

(17) '*f* causes *g*'.

But not of course conversely: (17) does not entail (14)–(16). What (17) entails (and is entailed by) is the existence of *some* sorts of events *X* and *Y* for which the following are all true:

(18) '*f* is the *X* event';
(19) '*g* is the *Y* event';
(20) 'There is a *Y* event because there is an *X* event'.

F and G are in fact the sorts of event that make (20) and hence (17) true, but they need not be. Any X and Y that make (18)–(20) true (and are essential to the existence of f and g – see **7** below) will do.

These entailments, as Jeremy Butterfield has pointed out, entail (17)'s transparency. If 'j' and 'k' are any singular terms referring respectively to f and g, (17) entails the existence of a suitable X and Y for which not only (18) and (19) are true, but also:

(18′) 'j is the X event';
(19′) 'k is the Y event';

which with (20) would entail

(17′) 'j causes k'.

This shows how the transparent (3) follows from (4). It also shows how causation here really relates the facts, not the events; f and g only supply the particular relata of causation (by making it true that there are F and G events), not the causation itself. These existential facts need not be related as cause to effect; and if they are, it is not because f and g are. It is the other way round: the events merely inherit the causal relation of the facts they supply. It does not even matter that *they* supply those facts: any F and G events could make (14) true. They need not be f and g, since being F and G does not fix their identity: they will have other properties on which their identity may depend. Don's fall and his death could easily differ sufficiently (from his actual fall and death) to be different events without falsifying (14).

Like any other such sort X or Y, F and G (being a fall by Don and a death of Don) have of course to conjoin general properties with some limitation to get the uniqueness (18) and (19) need. The limitation may be explicit ('*Don* dies') or assumed. We assume, for example, that 'Don dies' suffices, because most people only die once (contrast 'Dracula died *last time* because ...'). Similarly, we assume, Don dies of only one fall: for example, his fall in June 1970, on Mont Blanc.

The real question about X and Y is what general properties they may include: those, I assume, that figure in natural laws. In (17), for example, f and g should at least instantiate statistical laws to satisfy condition (*P*): that is, to ensure that in the circumstances (200 metres up a cliff) the probability of a death would be greater with a fall than without one. In short, I think Davidson (1967a) is right to claim that though statements of type (1), 'c causes e', 'entail no law ... they entail there is a law' (p. 160), if only a statistical one.

7 So much for (1), '*c* causes *e*'; but there are other transparent types of causal statement. There is the neglected

(21) '*c* affects *e*',

and even, as we shall see, a transparent species of type (2), '*E* because *C*'.

The difference between (1) and (21) involves the existence of the events *c* and *e*: for (1) implies, as (21) does not, that the existence of *c* causes *e* to exist as well. And (14)–(16) entail an instance of (1) because, though *F* and *G* do not fix the identity of the *F* and the *G* events, those events still need those properties. If Don did not fall and die, *f* and *g* would not exist. That is why their causal relation involves their existence, so that Don's death is not merely affected, but is caused, by his fall.

Causal properties need not be essential properties, however. Take this statement:

(22) 'Don dies instantly [*I*] because he falls fast [*H*]'.

That is,

(23) 'There is a *G&I* event because there is an *F&H* event'.

Again assuming one *F* and one *G* event, this implies:

(24) 'The *F* event is *H*';
(25) 'The *G* event is *I*'.

And (23)–(25) I say entail

(26) 'The *G* event is *I* because the *F* event is *H*',

which asserts that Don's death is instant because his fall is fast.

Being *H* and *I* are not essential properties of *F* and *G* events. Don's fall and his death could each be less swift and still be the same event. (22)'s causation does not involve their existence, only their being *H* and *I* – what Sanford (1985) calls 'aspects' of them (see **2**). Which is why (22) entails not (3), 'Don's fall causes his death', but only

(27) 'Don's fall affects his death'.

In short, causing is done by essential properties, affecting by inessential ones. When *X* and *Y* in (20) are essential to the *X* and *Y* events *c* and *e*, *c* causes *e*; and when they aren't, it affects it – with something like (26) saying

what the effect is. So in general (20) only entails, of whatever *c* and *e* are the *X* and *Y* events,

(28) '*c* causes *or* affects *e*'.

Whether in a given case (1) or (21) makes (28) true (or there is no fact of that matter) is not a question of causation, and so not relevant to the present point – though, as we shall see in **8**, it does bear on the relation between events and times.

What is to the point is that (27), like other instances of (21), is as transparent as (3). And so is (26), despite being of type (2). Replacing its singular terms by others referring to the same events will not falsify it. It even entails

(29) 'Don's instant death is instant because his fall is fast',

which is not false, as the assertion from **5**,

(10) 'Don's fall is Don's fall because his rope is the weakest',

is false, precisely because instantaneity is inessential to Don's death. 'Don's instant death is instant', unlike 'Don's fall is Don's fall', might have been false. So it can, as (29) claims, be true in fact because Don's fall is fast.

The evident transparency of statements like (26) is what makes many, like Davidson, think that all type (2) statements are transparent. But (26) is a special case; like (17), it is only transparent because the statement, (23), which makes it causal does not refer to its events. Those events are referred to only in (24) and (25), whose transparency lets (26) be transparent as follows. To be transparent, (26) must entail

(30) '*k* is *I* because *j* is *H*',

where '*j*' and '*k*' are any singular terms referring respectively to the *F* and the *G* event. But (26) and (30) are of type (2), '*E* because *C*', which entails '*C*' and '*E*'. So (26) entails (24), 'the *F* event is *H*', and (25), 'the *G* event is *I*', and (30) entails:

(31) '*j* is *H*';
(32) '*k* is *I*'.

To entail (30), (26) must therefore also entail (31) and (32), which it does by entailing (24) and (25), since their transparency makes them in turn respectively entail (31) and (32).

This account of (26)'s transparency should dispel the illusion that it shows a general transparency in type (2) statements. It should also, incidentally, dispel

the illusion mentioned in **5** that the probabilities causation needs can be subjective. They cannot, at least not if condition (*P*) holds, because subjective probabilities make opaque the following:

(33) 'The probability that the *G* event is $I = p$'.

For example, unless I know that the *G* event = *k*, I may believe '*k* is *I*' much less strongly than 'the *G* event is *I*'; so much less that (26) would meet condition (*P*) and (30) would fail it. But that would make (26) opaque, which it is not. For (*P*) to hold, (33) must be as transparent as (26): so its probabilities must be objective.

8 The above, I believe, exhausts the ways in which causation relates events. They all stem from causal relations between facts – and from a mere fraction of them. Much of the world's causation, like Don surviving because he hangs on, or his falling first because he has the weakest rope, relates facts without relating events at all.

But a kindred relation, temporal order, always relates particulars, namely particular times, as well as facts; and Davidson (1967a p. 154) says that this kinship shows that causation always relates events. What it really shows, however, as we shall now see, is that times are not events.

The kinship between temporal and causal order is shown in the temporal analogues of (1)–(4) – which I believe (1)–(4) entail (see Mellor 1981 chs 9–10), though I shall not argue the point here:

(1T) '*c* precedes *e*';
(2T) '*E* after *C*';
(3T) 'Don's fall precedes his death';
(4T) 'Don dies after he falls'.

The statement (4T) and others of the type (2T) report relations between facts, just as type (2) statements do. For (2T), like (2), always entails '*C*' and '*E*' (Don cannot die after he falls unless he both falls and dies). So it too is a partial truth function of '*C*' and '*E*'. But again, not a complete one. '*C*&*E*' no more entails '*E* after *C*' than it entails '*E* because *C*' – if only because, if it did, it would also entail the incompatible '*C* after *E*', that is, '*E* before *C*'.

Yet, as Davidson notes, his argument to show that '*E* because *C*' must be a complete truth function has a parallel for '*E* after *C*'. And the temporal argument's transparency assumption (ii) is more arguable than in the causal original. Take the analogue of (9):

(9T) 'Don's fall is the first fall after Don's rope is the weakest rope'.

This reads oddly, but will be true if (9) is, since if Don falls first because he has the weakest rope, it must be the weakest before he falls, and *a fortiori* before he falls first. But whereas (9) is undeniably opaque, (9T) arguably is not. For one thing, the probability condition (*P*) need not apply, to engender opacity in (9T) as it does in (9). And the analogues of (10)–(13), for example

(10T) 'Don's fall is Don's fall after his rope is the weakest',

are not as self-evidently false as their causal counterparts.

I still think (10T)–(13T) are false: it is hard to credit necessary facts, like the fact that Don's fall is his fall, with temporal location. But even if they are true, Davidson's argument must still fail somewhere. It cannot show that statements of type (2T) are complete truth functions, since we know they are not. Nor can it show that 'after', and its converse 'before', are not really connectives. How could it show that? Since they are used as connectives, that is what they are.

However, we need not settle the question of (2T)'s opacity. Provided '*C*' and '*E*' state existential facts, even opaque truths of type (2T) can yield transparent ones of type (1T) and even (2T), analogous to those that type (2) truths yield. Both (4) and (4T) report relations between the existential facts that there is an *F* event (a fall by Don) and a *G* event (a death of Don). So if *f* = the *F* event and *g* = the *G* event, I say (14)'s temporal analogue,

(14T) 'There is a *G* event after there is an *F* event',

yields a transparent

(17T) '*f* precedes *g*',

just as (14) yields the transparent (17) '*f* causes *g*'. And the temporal analogue of (22),

(22T) 'Don dies instantly [*I*] after he falls fast [*H*]',

likewise yields a transparent

(26T) 'The *G* event is *I* after the *F* event is *H*'

which says that Don's death is instant after his fall is fast. Temporal and causal relations between existential facts about events yield corresponding relations between events in the very same way.

But what about the temporal analogue of (8),

(8T) 'Don survives after he hangs on'?

We saw in **4** that if 'Don falls' and 'Don dies' are existential statements about events, 'Don hangs on' and 'Don survives' are not. They, and hence (8) and (8T) which entail them, are negative existential statements. They only *deny* the existence of events, and *a fortiori* of events related either causally or temporally. The truth of (8T) can yield no temporal relation between events.

But it can yield a relation between times. Whether Don falls or hangs on, dies or survives, there are times at which he does so. 'Don falls', 'Don dies', 'Don hangs on' and 'Don survives' are all existential statements about times. Abbreviating them by '*DF*', '*DD*', '*DH*' and '*DS*', (4T) and (8T) entail:

(4T*t*) 'There is a time when *DD* after there is a time when *DF*';
(8T*t*) 'There is a time when *DS* after there is a time when *DH*'.

So if Don falls or hangs on at t_1 and dies or survives at t_2, (4T) and (8T) can both entail a transparent 't_1 precedes t_2'. These times will not of course be instants of time. The fastest of falls takes time, and so does even an instant death. For t_1 and t_2 to be *the* times when Don falls (or hangs on) and dies (or survives), they must be intervals of time. The sub-intervals (and ultimately the instants) they comprise will be defined in turn by the temporal order of related facts about, for example, Don passing various points on the cliff as he falls. All these times, like events, get their temporal order from that of the facts located at them – not *vice versa*, as Davidson (1967a p. 154) claims.

Times may likewise be defined by all temporally ordered facts. Indeed, I suppose that is what times are. (Given relativity, it is really spacetime zones that are defined, by spatio-temporal relations between facts. But that is no problem, provided spatial relations also hold between facts, which they do: for instance, Don dies *where* he lands, and *200 metres below where* he starts to fall. But since the extension to spacetime is trivial, I will stick to time.) So '*E* after *C*' always entails

(2T*t*) 'There is a time when *E* after there is a time when *C*',

which, if t_C = the time when *C* and t_E = time when *E*, entails

(1T*t*) 't_C precedes t_E'.

That is, '*E* after *C*' turns '*C*' and '*E*' into existential statements about times, whether or not they are also existential statements about events. And since they

often aren't, times cannot always be events: there are not enough events to go round.

And even when '*C*' and '*E*' are existential statements about both times and events, we still cannot equate them as Davidson (1967a p. 154) does. Don's fall *f* and his death *g* have of course the same temporal order as the times, t_F and t_G when Don falls and dies, namely, the temporal order of those two facts. But that does not make those times the same as those events, since they differ causally: *f* causes *g*, but t_F does not cause t_G.

This causal difference between times (or spacetime zones) and events is not just a matter of stipulation, though it could be. If spacetime had to be causally inert, it could be made so by not extending the inference scheme in **7**, from

(20) 'There is a *Y* event [*e*] because there is an *X* event [*c*]'

to

(28) '*c* causes or affects *e*',

to cover times. However, general relativity arguably shows that spacetime is not inert: that its local structure interacts causally with the matter it contains (see Mellor 1980). So this inference scheme may have to cover times; but *f* and *g* will still differ causally from t_F and t_G. For being when Don falls and when he dies are not *essential* properties of t_F and t_G. Even if these times are identified by their places in a temporal order derived from that of such facts, they are the location of too many other facts for their identity to depend on what Don does in them. They would have been the same times even if Don had hung on and survived. So even if causation does *involve* them, it will not involve their *existence*. Thus t_F may *affect* t_G, since t_G is a time when Don dies, because t_F is a time when he falls; but it will not *cause* it, as *f* causes *g*.

9 I have said why and how causation relates facts, and how as a result it can relate events and perhaps times. It remains to say why all this matters, and how it affects what causation is, and what physical basis, if any, it needs.

Some ways in which it matters have emerged already. One is that effects must be facts in order to have the probabilities I think the analysis of causation needs. Another is all the causation which gets left out if it is limited to events: all the causation, for example, in which nothing happens – as when things have properties (colour, shape, temperature, mass, and so on) at a time because they had them earlier and nothing has happened since to change them.

Since causation needs no particulars, it could even act on a *Tractatus* world that really was a 'totality of facts, not of things' (Wittgenstein 1922 1.1) with

no particulars at all. It would indeed make that world more than a 'totality of existent atomic facts' (Wittgenstein 1922 2.04) since '*E* because *C*' is a partial but not complete truth function of '*C*' and '*E*'. But it need not generate events. The laws needed to give effects probabilities may indeed need particulars to distinguish their instances (Ramsey 1925): for without particular constituents, like *a* in *Fa*, facts would not be collectable into laws like '$(x)(Fx \rightarrow Gx)$'. But the particulars need not be things or events. For if, as I argue, causal order entails temporal order, causation would make a *Tractatus* world temporal; and then the entailments in **8** would give it times – and, if its facts were also spatially related, zones and hence points of spacetime. And that ontology, as Carnap (1936–7) and Field (1980) have indicated, might do for its physics.

But not for its metaphysics. Facts in a causal world cannot all consist of particulars having properties or being related to each other: it cannot just be a Tarski world of particulars satisfying sentential functions, for instance such predicates as '...dies' and '...causes...' (Tarski 1944). For, as I remarked in **6**, a relation can hold (that is, particulars can satisfy a predicate ascribing it) only if it does so however the particulars it relates are referred to. Statements that relations hold must be transparent. But we have seen that many causal truths are opaque. They depend on how particulars are described, as well as on their properties and relations. So a causal world must contain more than particulars and universals. It not only can be a 'totality of facts, not of things': in so far as it is causal, it must be.

This is not to deny the reality of things and events. Ours is more than a *Tractatus* world made temporal by causation: its particulars are not just spacetime zones picked out by the facts located at them. They might be if they were all events: extended events, with temporal as well as spatial parts, could be identified with spacetime zones. But there are also things, able to change, and agents, able also to act and to be responsible; and these things and agents must be particulars that extend in time without having temporal parts (see **2**, and Mellor 1980). That, I maintain, makes them – us – irreducible to events, and *a fortiori* to spacetime zones, though I cannot argue that thesis here. However, reducible or not, changeable and interacting things and agents are features of the world which accounts of causation must allow for, as mine does: since particulars of any kind that supply causally related facts can thereby be causally related too.

So much for the relata of causation. What of causation itself? It is not a relation, in the sense of a universal, different in kind from the particular causes and effects which it relates. When Don dies because he falls, that is just another fact, like the fact that he dies and the fact that he falls. It does not

reduce to them: Don can fall and die without dying because he falls. But it is still only a fact, as they are; and it too may have causes and effects. Don may, for example, die because he falls because he wears no helmet – which his dying because he falls may well cause later climbers to wear.

This kind of iterated causation shows moreover how causation can be perceived. Perception is a causal process: we see things, events and facts by their so affecting our senses that we are caused to believe in them. So anything we see must have effects, which causation itself cannot have if all causes are particulars. Hence a specious problem about how it can be detected (Hume 1739 bk 1 pt 3 sect. 2), which then casts doubt on its objectivity, and hence on the objectivity of perceptual knowledge. But if causes are facts this problem goes, and with it goes much Humean scepticism. For now the fact that one fact causes another can itself make us perceive it, by so affecting our senses that we are caused to believe in it. Of course we cannot perceive every kind of causation, any more than we can perceive every kind of cause. Perhaps we cannot see that Don dies because he falls (though I think we can); but there is plenty of causation we can perceive. For example, to see that a thing is red is to see how white light changes colour *because* the thing reflects it as it does.

We rely on causation so much that any account of it must say how we detect it. The ease with which this account does so is yet another reason to believe it, but the very feature that makes causation visible raises a question about its physical basis. If it is not a relation between particulars, and *a fortiori* not a physical one, what is it? What kind of physical fact, beyond the fact that *C* and the fact that *E*, makes statements of type (2) '*E* because *C*', true?

Consider the probability condition (*P*): that '*E*' be more probable than it would be in the circumstances if '*C*' were false. I admit (*P*) is debatable, and it certainly does not suffice to entail (with '*C*' and '*E*') '*E* because *C*'. But it does explain, as we saw in **5**, the falsity of many causal claims, and to make it sufficient might only need more of the same: for instance, contiguity or density, to provide causal intermediaries between causally related facts separated in spacetime. If so, causation will neither have nor need the physical basis (or bases) adumbrated in **1**. For (*P*) demands only actual and counterfactual probabilities, which is not a physical basis like energy transfer. Probability – like truth – is no subject matter for a science, since it is always the probability (or truth) *of* something which is the real subject matter. Any contingent proposition, about any subject, may be true – and may also be more probable if some other proposition is true than if that proposition is false. If

this, or anything like it, suffices for causation, causation can no more be the province of one science, such as physics, than truth or probability can be.

This is not to deny physicalism. Perhaps all truths of the form '*E* because *C*' are made true by physical facts. The fact that *C*, the fact that *E*, and the further facts that make '*E* because *C*' true, may all be physical. But these further facts of causation may also still be no less heterogeneous than the facts they link. Indeed they must be, being apt as we have seen to have causes and effects of their own. So physical facts are not linked causally by a distinct causal kind of physical fact, and it is futile to look for any such kind. It is indeed true that physics, *pace* Russell (1912 p. 171), has never 'ceased to look for causes', let alone because 'there are no such things': for there *are* such things. So physics should look for them, and also – case by case – for the facts that link them to their effects: but not for causation itself.

13 *On raising the chances of effects*

1 Introduction There is no doubt that effects need probabilities. Deterministic causes – sufficient and/or necessary conditions – give effects probability 1 with their causes and/or probability 0 without them. And the indeterministic causation of Salmon (1984) and others has probabilities of effects built into its very foundations.

Doubt and debate enter with the questions: what probabilities must effects have, and of what kind? The kind, I agree with Salmon, is what Carnap called 'statistical': objective physical probability, which for brevity I call 'chance'. Causal probabilities cannot be merely subjective or inductive: short-circuits cause fires neither by making people expect them nor by providing evidence for them (effects can be as good evidence for their causes as their causes are for them).

Salmon and I differ on what chance is: he favours a frequency account of it, I favour a propensity one (see chapter 14 and Mellor 1971). But we also differ on what chances effects must have: I and others say they must be greater with their causes than (in the circumstances) they would be without them; whereas Salmon denies this. That is the dispute I wish to settle here. And since it can be settled in the same way whatever chances are, that question can be waived for the time being.

What can't be waived is the question of what the right way to settle the present dispute is. Invoking causal intuitions case by case isn't the way, because they too are unsettled. Take Salmon's (1984 p. 200) excited atom that decays improbably from an excited energy level to a ground state *via* an improbable intermediate level (1). He thinks it is obviously in the ground state *because* it was previously in level 1, whereas I think it is just as obviously in the ground state *despite* that fact, not because of it. (The low *chance* of the atom's second decay may be caused – deterministically – by its first decay; but that's another matter.)

In the end, no doubt, the view should prevail that makes most sense of most plausible cases of causation. But plausibility depends on more than case-by-case intuitions about causation itself. To vary the old (and overrated) adage:

don't ask for the use, ask for the point of the use. Ask, in other words, what we mean to imply when we call a situation causal. That, I believe, is how we should settle our dispute: consider how causation's connotations depend on the chances of effects, and let that settle our disputed cases.

2 The connotations of causation Causation's main connotations are the following:

(1) *Temporal*: Causes precede their effects;
(2) *Evidential*: Causes and effects are evidence for each other;
(3) *Explanatory*: Causes explain their effect;
(4) *Means–end*: If an effect is an end, its causes are means to it.

These connotations clearly don't entail determinism: for (1) causes can precede their effects without being either sufficient or necessary for them; (2) evidence for something need not raise its probability from 0, or to 1; (3) explanations need not be deductive, nor need the falsity of an *explanans* entail the falsity of its *explanandum*; and (4) a means to an end can be worth taking even if it is neither an infallible nor the only possible means.

And what makes paradigm cases of indeterministic causation plausible is precisely that all these connotations apply to them. When we say that an atom was caused to split by being bombarded, we imply that the bombardment preceded the splitting, was evidence for it (and *vice versa*), explains it, and would be a means to it. Similarly, when we say that someone's smoking caused his or her cancer, we imply that the smoking preceded the cancer, was evidence for it (and *vice versa*), explains it, and would have been a means to that – admittedly perverse – end.

Theories of causation and of the other concepts involved should therefore respect these connotations, and say why causation has them. Theories of causation and time should between them say why causes precede their effects, and similarly with our theories of causation and evidence, of causation and explanation, and of causation and what means to ends are. And even though causation's connotations do not entail determinism, they may still set some bounds to the chances of effects, which theories of it should respect. But do they? And if so, which ones do so, and what bounds do they set?

Salmon's theory will not say, since it bases causation on 'the two basic concepts of *propagation* and *production*' (Salmon 1984 p. 139). But in his examples (e.g., 'the electrical discharge produces a fire') 'produces' is just a synonym for 'causes'; while 'propagates' just adds to causation's temporal

connotation the truism that the cause–effect relation is not intransitive. And these concepts will not serve to settle the question at issue.

Most other theories base causation on its evidential or explanatory connotations. Hume (1777 sect. 4 pt 1) in effect defines causation as the basis of non-demonstrative inference. Hempel (1965 sect. 2.1) in effect defines it as providing a certain kind of explanation. I think they are both wrong, for reasons I have given elsewhere (Mellor 1987b): I think the theory of causation should be based on its means–end connotation. Causation is best conceived of as the feature of the world that gives ends means. And this conception, as we shall see, does set bounds to the chances of effects.

There is however an obvious objection to basing causation on its means–end connotation: namely, that to be a means to an end just *is* to be one of its causes – that this is all 'a means' means. And if that were so, the means–end connotation would indeed tell us little about causation. In particular, constraints on the chances of ends would have to follow from independent constraints on the chances of effects, not *vice versa.* But in fact the objection is mistaken. We can say enough about means to limit the chances of ends without invoking causation at all. And the means–end connotation will then limit the chances of effects accordingly.

The trick is to use a principle of non-causal decision theory (Jeffrey 1983) to say what it takes to be a means to an end. The principle is the *expected utility principle*: act to maximise expected utility. Only here (*pace* Jeffrey and others) the expected utility must be objective: in particular, it must incorporate chances, not merely inductive or subjective probabilities (see chapter 16). And then the principle yields the independently plausible condition that an end's chance with a given means must in the circumstances exceed its chance without that means.

I have given the argument for this elsewhere (Mellor 1987a, 1987b), but it may be summarised as follows. Call the end (a desirable fact) E, and the prospective means (a realisable fact) C. The basic idea is that C will be a means to E only when expected utility prescribes it, i.e. when C's expected utility exceeds that of $\sim C$. Of course this won't be true in general, because C may have some intrinsic value or disvalue of its own. But it will be true if that is not the case, i.e. if the C and $\sim C$ rows in the relevant utility matrix are the same:

Utilities	E	$\sim E$
C	u	u'
$\sim C$	u	u'

Now suppose that E's chances with and without C would in the circumstances be:

Chances	E	$\sim E$
C	p	$1\text{-}p$
$\sim C$	p'	$1\text{-}p'$

Then the relevant expected utilities $EU(C)$ and $EU(\sim C)$ of C and $\sim C$ will be:

$$EU(C) = up+u'(1\text{-}p);$$
$$EU(\sim C) = up'+u'(1\text{-}p');$$

so that $EU(C) > EU(\sim C)$ if and only if

$$(u\text{-}u')(p\text{-}p') > 0.$$

But to say that E is the end is to say that u exceeds u', so that expected utility prescribes C just when

$$p > p',$$

i.e. just when the chance of the end would be greater in the circumstances if C were brought about than if it were not. Unless that is so, bringing C about will be no way to bring about the end E: C will be no means to E.

3 Causes as means to ends The means–end connotation now imposes this condition on the cause–effect relation: an effect's chance with a cause must be greater than it would be in the circumstances without that cause. For otherwise, effects could be ends without their causes being means to them.

In deriving this condition, I do not of course assume that all or even any effects *are* ends, nor that their causes could actually be brought about or are only worth bringing about for the sake of their effects. None of this can be true, if only because effects in turn cause other effects. And none of it is entailed by the means–end connotation, which does not say that effects must actually be ends, nor that their causes must be realisable and lack all value or disvalue of their own. All it says is that *if*, while C is related to E as cause to effect, E *were* an end, C would *ipso facto* be a means to it: i.e. that if in those circumstances C had no intrinsic value or disvalue, and could be brought about directly when E could not, the expected utility principle would prescribe bringing C about.

In short, the means–end connotation is conditional. It does not require a causal world to be full of ends and means. It can perfectly well allow causation

in a world without either value or agents, and hence without either ends or means. In such a world, as in ours, causation could still exist as what, if only there were ends, would give them means.

I must emphasise also that using the expected utility principle to define causation's means–end connotation does not require us to endorse all that principle's prescriptions. We needn't for example ignore, as it does, the distinction between action and inaction. Suppose a patient in great and incurable pain so much wants to die (E) that a doctor should let him die even though he shouldn't kill him. Whether a means C to the end E should be adopted will then depend on whether it takes positive action to bring it about. We can admit that, while still insisting that for C to *be* a means to E, it must be such that under the conditions stated above the expected utility principle would prescribe it.

Similarly, we can let expected utility give way to the 'maximin' principle (Luce and Raiffa 1957 p. 278) in some cases. Some possibilities are arguably too bad to be worth running any risk of for any end, however valuable. If so, perhaps C should never be brought about when $C\&{\sim}E$'s utility falls below some minimum value, even if $EU(C)$ exceeds $EU({\sim}C)$. What this threshold value (if any) is, below which expected utility should give way to maximin, is a moot point. But fortunately we needn't settle it. For the maximin principle does not deny that C *is* a means to E, merely that C should be brought about just because its expected utility exceeds ~Cs.

Conflicts between expected utility and the 'dominance' principle (Jeffrey 1983 p. 8) are more serious. If in each column of the C/E utility matrix, the utility in the ~C row exceeds that in the C row, dominance will prescribe ~C, on the grounds that whether E is a fact or not, it will be better if C isn't a fact. Like maximin, dominance can contradict expected utility in such cases because it never lets E's value make C worth bringing about, however much C raises E's chance and thus C's expected utility. But the case for dominance, unlike that for maximin, seems to imply that C is not merely unjustified by E, but is not a means to it. So if expected utility is to define the means–end relation, it must outrank dominance when the two principles conflict.

And so it does when the expected utilities incorporate chances, i.e. statistical as opposed to inductive or subjective probabilities. For then the situation is one of real *risk*, as opposed to mere *uncertainty* (Luce and Raiffa 1957 p. 13), and all parties agree that expected utility outranks dominance then. Doubt arises only when expected utilities are based on inductive or subjective probabilities. And I confess I share the doubt: I doubt if an unpleasant medicine is worth taking just to increase the evidence (inductive probability),

or the degree of my belief (subjective probability), that I will recover from an illness (see chapter 16). But the point need not be argued here. No one takes inductive evidence for a proposition to be *ipso facto* a *means* of making it true. Nor, when subjectivists advocate wishful thinking, do they do so as a means of attaining the wished-for end. The authority or otherwise of inductive or subjective expected utility is therefore irrelevant to the means–end relation. Basing that relation on the expected utility principle only requires the principle to outrank dominance when it uses chances; and no one denies that it does that.

4 Causal dispositions In short, the means–end connotation makes causes raise their effects' chances – by which of course I don't mean that they cause their effects' chances to increase. That condition would be both circular and ineffective, since it both invokes causation and leaves it deterministic (merely replacing the real effects with their chances). What I mean is what I said in **3**: an effect's chance with a cause must be greater than it would be in the circumstances without it.

But the phrase 'in the circumstances' raises well-known problems, which I cannot really tackle here (see Mellor 1987b), but should say something about. First, it at least means 'given the other causes', which may look circular, but isn't. It simply requires each cause C of an effect E to raise E's chance above what it would be without C but with all E's other causes, i.e. with everything else that (*inter alia*) meets this condition.

We need something like this to allow for alternative causes that would occur if C didn't, and would make E's chance as great as or greater than C makes it. And what really makes this hard to allow for is that our condition makes causation depend on what E's chance would be without C, and that too may depend on C. But we want the causal relation to depend on the actual value of what E's chance would be without C – not on what that value would be if C's absence would make a difference to it. The latter is what the phrase 'in the circumstances' is supposed to rule out. The question is what, if it is to do so, must 'the circumstances' be?

Basically, the causal circumstances we need are dispositions of objects and fields. To see why, consider first a deterministic disposition like solubility (in water). The fact that an object a is soluble is what makes putting it in water cause it to dissolve. It does so because while a's chance of dissolving when not in water is 0, its being soluble means (roughly) that its chance of dissolving if it were put in water would be 1.

But this definition of solubility is not quite right. Suppose *a* is not only soluble but valuable, so that we take care to keep it dry. We would only put it in water if it were *in*soluble, so that in fact if *a* were put in water, its chance of dissolving would be 0 – which contradicts our definition of solubility.

Fortunately, however, this true subjunctive is easily falsified by adding to its antecedent the condition that *a* be soluble. If *a* were put in water while it was soluble, its chance of dissolving would be 1, not 0. So the right definition of solubility (*D*) is of a property *S* such that *a*'s chance of dissolving if put in water would be 1 if and only if it were then *S*. But then *a*'s being *S* is the very circumstance in which our condition on causation applies to this case.

(*D*) may also look viciously circular, but again it really isn't. (If it were a quantified material conditional, instead of a subjunctive, it would be what Carnap (1936–7 sects 5–6) called a reduction sentence.) But nor is (*D*) an eliminative definition, since it refers to *S*. We must still say what *S* is: presumably some arrangement of *a*'s molecules that makes them easily separated by water molecules. But that arrangement will no doubt depend on *a*'s chemical composition, so that *S* will vary from one chemical substance to another. So to say that *a* is soluble is not to ascribe a specific property to it, but to make an existential claim: *a* has *some* property *S* such that if and only if *a* were put in water while being *S* would its chance of dissolving be 1. But then whatever *S* is, it will supply all 'the circumstances' we need, since by satisfying (*D*) it will automatically comprehend all the other causes of *a*'s dissolving: e.g. the chemical composition that makes *a*'s molecular structure satisfy (*D*).

And as for deterministic causes, so for indeterministic ones. The relevant circumstances are tacitly defined by the matrix of chances given in **2**. Suppose *E* is my recovery from an illness, *C* my taking some medicine. The matrix does not just state what my chances of recovery would be were I to take, or not to take, the medicine. It tacitly credits me with metabolic properties such that, while I have them, my chances of recovery would be as it states. The matrix of course does not say what these properties are, any more than (*D*) says what *S* is. Their identity is an empirical matter, to be settled by medical research. But whatever the properties may be, my having them is the very circumstance in which my taking the medicine would raise the chance of my recovery and thereby cause it.

In short, causation needs dispositions to make true the subjunctives implied by the condition that causes raise their effects' chances. To identify causes and effects is to give a dispositional specification which some properties of the objects or fields involved must meet; and to identify the causal relation is to find the properties which meet that specification. These of course are tasks for

science, and are clearly interdependent: finding the property that meets a dispositional specification shows how (and hence that) it is met; while failure to find it casts doubt on the disposition, and hence on the causation that entails it. Thus on the one hand, finding the nuclear structure that makes an atom which has it more likely to split when bombarded shows how bombarding such an atom can cause it to split. And on the other, until we identify the properties of cigarette smoke which make those who inhale it more likely to get cancer, the tobacco industry can still continue to dispute the causal connection between smoking and cancer.

5 The efficacy of causes Causation must of course do more than raise its effects' chances, and not all its other features will follow as this one does from its giving ends means. One that does not is the fact that causes are contiguous to their immediate effects – or rather, given the denseness of time, that they are linked to their effects by dense 'ropes of causation', i.e. by the processes which Salmon (1984 ch. 5) takes to be causation's basic entities. But if the means–end connotation does not explain the denseness of causation, nor does it deny it: the two conditions are perfectly compatible.

But the means–end connotation does explain other aspects of causation. In particular, it explains its temporal connotation. For if causes raise their effects' chances, then causal loops, and hence simultaneous and backward causation, are impossible (Dummett 1964, Mellor 1981 ch. 10). The condition that makes causes means when their effects are ends makes even the most underdetermined effects come later than their causes. The means–end connotation not only lets causes precede their effects: it makes them do so.

Another thing the means–end connotation does is give sense to the idea that causal efficacy comes by degrees – which shows incidentally that our condition (that causes raise their effects' chances) is too weak: to be a means to an end E, C must raise E's chance by more than an infinitesimal amount. The reason is that not all means are equally useful, and the more C raises E's chance (from p' without C to p with it) the more useful it is. For when C costs something (i.e. lowers the utility of both E and $\sim E$), the greater p-p' is, the less valuable E needs to be before the expected utility principle will justify incurring that cost. While as p-p' gets less, the less use C becomes: i.e. the more valuable E must be to justify employing C at any cost, however small.

In short, the usefulness of means comes by degrees, of which p-p' is the natural measure; and a means must be *some* use to be a means at all. So a means must raise its end's chances by more than some minimum amount. The amount will no doubt vary from case to case, and even then be hard to specify.

But the amount doesn't matter, any more than it matters how hot something must be to be hot. A thing's temperature tells us all we need to know about how hot it is; and p-p' similarly tells us all we need to know about how useful C is as a means to E.

Nor is causation's means–end connotation the only one that comes by degrees. The temporal connotation doesn't – temporal order, unlike temporal distance, doesn't come by degrees – but the others do. The evidence which a cause C and its effect E provide for each other clearly increases the more C raises E's chance. And so does the extent to which C explains E, for reasons I have given elsewhere (Mellor 1976).

(Jeffrey (1969), Fetzer (1981 ch. 5), Salmon (1984) and others dispute this. They deny that to explain E, C must raise its chance (or make it high), the higher the better. Their objections do not persuade me, but I shall not argue the point here.[1] For Jeffrey *et al.* agree that causes explain their effects by giving them chances. What they deny is not the explanatory connotation, merely the idea that it comes by degrees. And although I think it does, that is not essential to the idea that causation does. That idea gets sense enough from its means–end connotation. Its entailing degrees of evidential and explanatory strength is a real but dispensable bonus.)

The efficacy, p-p', of the causal relation between C and E is thus primarily a measure of C's usefulness as a means of bringing E about; but also of how much evidence C and E provide for each other, and – I maintain – of how far C explains E. So the most effective causes are those that provide the most useful means, the best evidence and the best explanations for their effects. But those are the traditional deterministic causes, which raise their effects' chances from 0 to 1 – and this is no doubt why causes have traditionally been required to determine their effects. Now we know they need not, because although indeterminism progressively weakens causation's connotations, it does not destroy them all at once. But it does destroy them in the end, when causes no longer raise their effects' chances. And because determinism gives causation all its connotations in their highest degree, deterministic causation is still the best. So I confess to having the 'lingering desire for Laplacean determinism, or if worse comes to worst, as close an approximation thereto as possible' of

[1]Except to deny their implausible and unargued assumption that everything, however improbable, must be explainable (e.g. Fetzer 1981 p. 134); and to press the 'pivotal principle' which even Salmon (1984 p. 113) feels he needs 'temerity to ... reject': namely, that what explains E could not also explain $\sim E$ – e.g. that if my smoking explains my *getting* cancer, it could not also explain my *not* getting it. Causes that raise the chances of the effects they explain automatically satisfy this principle. Salmon's and Fetzer's causes do not.

which Salmon (1984 p. 113) accuses me. And so will anyone who understands the implications that make causation worth ascribing in the first place.

14 *Chance and degrees of belief*

1 Are there chances, and if so, what are they? These are still contentious questions. Like Salmon (1979), I believe that there are chances, but I disagree with him about what they are. We do agree on some points, principally that chance is a species of objective probability, namely physical or statistical probability. Some thinkers have claimed to discern another species in this genus, namely the relational probability that inductive logic treats of. I doubt the claim: inductive probabilities are, I suspect, all descended from chances. Certainly most are, so chance is anyway the right species to study first. Whether it exhausts the genus is a question we can afford to leave open.

The first question to ask about chance is what, if anything, is wrong with a frequency account of it. There is clearly an intimate connection between chance and frequency: of heads on coin tosses, of smokers getting cancer, of radium atoms decaying in set times. There need to be good reasons for not identifying chance with frequency. Unlike Salmon, I think there are such reasons, and that sets me the task of providing an alternative account of chance, one which, amongst other things, explains how it does relate to frequency.

Frequency theories of chance have been around a long time (e.g. Venn 1866), and their merits and defects are well known. Their overriding merit used to be that no other theory made sense at all of chance being, as it evidently is, both empirical and objective: frequency theories had the market to themselves. Today they have competitors, a motley collection of so-called 'propensity' theories, which treat chance as something like a weak disposition: of coin tosses to land heads, of smokers to get cancer, of radium atoms to decay. These theories, mine (1971) included, at first sold well enough on the familiar defects of Model T frequentism. But now, perhaps stimulated by the competition, Salmon and others have improved the frequency theories' specification. It is still imperfect, but so are its rivals, as Salmon (1979) has rightly pointed out. They too need improvement and defence if they are to remain on the market. My object here, in response to Salmon's challenge, is to improve the specification of my own propensity model, without, I hope,

adversely affecting its conceptual economy. But since the basic design remains the one I gave in my (1971), I shall draw on that as I go, without elaborating, amending or defending it except where it has been alleged to be defective.

2 I start with subjective probability. This is a measure of the strength of beliefs that satisfies the standard axioms of mathematical probability. Thus the strongest possible belief has probability 1, the strongest disbelief has probability 0, and the various degrees of doubt have probability values in between; if the probabilistic degree of belief is p, that of its negation is $1\text{-}p$; and so on. One way of applying a measure of my belief's strength is for me to think what odds I would choose for compulsory bets on its truth in a betting situation so specified as to prevent anything else influencing the choice of odds. A so-called 'Dutch book' argument then shows that a simple function of these odds, the betting quotient, is constrained to satisfy the probability axioms: otherwise, in the situation specified, I would lose on some outcomes and win on none, which is no bet at all. When a betting quotient is so constrained, it is called coherent; I shall refer to it as a CBQ for short.

I take the Dutch book and other arguments to show that the strength of most of our beliefs has a probability measure, and this is what I mean by subjective probability. Not everyone agrees: Kyburg (1978), for example, does not. But I have tried elsewhere (see chapter 3 pt II) to meet his objections to subjective probability, and here I shall take my riposte as read, since using probability to measure degrees of belief is not the most contentious aspect of my propensity theory. My main task will not be to defend subjective probability but to say how and why chances should constrain it. The constraint is clearly to be imposed somehow by what we know; but not just by our knowledge of non-chance facts limiting the degree of our belief through being more or less good evidence for its truth. That relation (of being more or less good evidence) is an *a priori* inductive probability if it is a probability at all, and inductive probability is not our concern. Chances, if there are any, are in the world, not just in our inductive logic: knowledge of them is part of our evidence, not just a measure of how good the evidence is. The question is, how chances are in the world, i.e. what knowledge of chance is knowledge of, and why and how such knowledge should constrain the degree of our beliefs.

On the first part of this question I have little to add to what is in chapter 6 and Mellor (1971). Most of our knowledge of chances I take to derive from knowledge of things having a kind of dispositional property which, taking the term from Popper (1957), I call 'propensity'. For example, the bias – or fairness – of a coin or of a coin-tossing gadget is a propensity, and so is the

half-life of a radium atom. To know a propensity is not, however, immediately to know a chance. It is only to know what chance there would be, for example, of heads *if* the coin were tossed, or of the atom's decay in a set time *if* it remained unbombarded for that long: i.e. what degree of belief in those outcomes could be objectively justified in those circumstances. Propensities are what make true what Lewis (1980) calls 'history-to-chance conditionals'; only, since history does not act directly at a temporal distance, they are properties of things at the very times to which the consequents of the conditionals refer. Propensities, in other words, are not chances, but dispositions whose displays are chances; just as fragility is a disposition whose display is the fragile thing breaking. In both cases, the disposition is a property of a thing – a coin, an atom, a fragile glass – whereas the display is a property of an event – the coin being tossed, the atom being unbombarded for a stretch of time, the glass being dropped.

Few advocates of propensity distinguish it from chance in this way, unfortunately. Most just call chance 'propensity' because they take chance itself to be a disposition, namely a disposition to yield a 'long run' frequency, e.g. of heads if the coin toss were endlessly repeated. But this does little more than a good frequency theory can do. In particular, it does not explain how knowledge of such a disposition should tell us how strongly to believe that heads will be the outcome of a single toss. Propensity theories of this kind mostly deserve Kneale's charge (Körner 1957 p. 80) against Popper's account: they think calling chance 'propensity' is enough to solve the conceptual problems it presents.

Such theories prostitute the good name of propensity – they are no better than they should be, and I desire not to be associated with them. What I do desire to be associated with are theories that treat of dispositions, like colour and temperature, to affect us in ways which the dispositions explain and by which they can in turn be characterised: colours by the visual sensations we get when we look at them, temperatures by the feelings of warmth they generate in us. The best way to understand propensity is to compare it with such dispositional properties, in order to see just how and why it differs from them. Any such property would do; I take colour because it is among the most familiar and least technical.

3 A thing's colour may be specified by the colour of the light reflected from it, and this in turn by the subjective colour judgment which perceiving the thing by this light should produce. Similarly, a propensity may be specified by the chance which displays it, and the chance in turn by the degree of belief

which perceiving it – and hence the propensity – should produce. In both cases we are specifying a disposition by its display, and the display by the mental state which should result from perceiving it and through it the disposition. The question is whether this is an intelligible and useful way of specifying an objective property and, in particular, how it can be done for propensities.

To specify a disposition in this way we must do two things. First we must identify the mental state involved independently of the disposition it will be used to specify. 'Looks green' must not just mean 'is the colour green things look'; 'seems probable' must not just mean 'is how probable probable things seem'. Otherwise the specification would be unenlighteningly circular. Secondly, we must say when being in this mental state amounts to knowing that something has the disposition in question. When is seeing green having knowledge of green light and thence of a green thing? When is a degree of belief knowledge of a chance and thence of a propensity?

I shall undertake these two tasks in turn, confining myself mostly to the relation between the display and the mental state, i.e. between chance and degree of belief. Once we get that right, the relation of chance to propensity will, I believe, pose no problems not adequately dealt with in chapter 6 and Mellor (1971).

Take colour first. The mental state to be identified when I see a colour is the having of a belief with a certain content, e.g. that some particular thing is green. Now we need not identify this state completely. For instance, we need not distinguish it from believing that something else is green: it is only the colour content we are after. What we must distinguish it from is believing that something is red, or hot, or anything else but green.

We start by appealing to the peculiar visual experience people typically have when something looks green to them. There certainly is such an experience, which differs from that of seeing something red or feeling it hot. The problem is to say what it is. For we apply our colour words primarily to things, not to their looks; and it is tempting to think that 'looks green' does just mean 'looks the colour green things look'. Suppose, for example, that fresh grass is a paradigm of a green thing. Does 'green' not mean in part 'the colour of fresh grass', and 'looks green' therefore 'looks the colour fresh grass looks'? If so, specifying colours by their looks will be going in a futile circle.

Fortunately it is not so. The look of a paradigm is what we baptise, not the paradigm itself. Suppose a foreigner wants to know what colour 'green' stands for. It is no use my telling him that fresh grass is green unless he sees some. And having seen some, he can perfectly well conceive that the grass might

have looked quite different, i.e. have been a quite different colour. The paradigm only has to be green in the actual world, we might say, not in all possible worlds. In other words, 'fresh grass is green' does not have to be analytic for fresh grass to be a paradigm of green; 'looks green' therefore does not in that sense mean 'looks the colour fresh grass looks' (cf. Kripke 1971).

'Looks green' applies to whatever, in any possible world, looks the same colour as our paradigms look in this world. And whether two things look the same colour to someone is nothing more than the question whether he can tell them apart when their other properties are hidden from him. That is how we detect colour-blindness, where different colours look the same, and similarly, how we detect the acuity that can sort out colours indistinguishable to the rest of us.

So far so good. We can, I believe, identify the mental state characteristic of a colour perception without disabling reliance on the actual colour of the thing perceived. But whether or not that is really so for colour, it is certainly so for propensity and chance. Here it is the strength, not the content, of a belief which has to be identified. The content can be anything which we could think to have a chance: a coin landing heads, a man getting cancer, an atom decaying within the year. In each case what distinguishes the perceived chance is only the degree of the belief. So we need not trouble to distinguish beliefs themselves, only the different degrees of any one of them.

I have alluded already to the standard method of distinguishing degrees of belief as subjective probabilities. I admitted that this has been jibbed at; but not for appealing to chances. No one thinks we need to know the real chance of an event in order to measure the strength of someone's belief in it. On the contrary, many subjectivists are positively motivated by determinism: it is just because they suspect that there are no objective chances that they invoke subjective probabilities to provide a surrogate subject matter for statistical science. The surrogate is indeed inadequate; but their view is neither inconsistent nor epistemically problematic: everyone agrees that we can detect degrees of belief just as well in a deterministic world as in a chancy one.

The problem with chance is not identifying the mental state involved; it is saying when being in that state amounts to having knowledge. When is having a degree of belief knowing a chance? Here the case of colour presents the easier problem; and since its solution will set us up to tackle chance, I will again take colour first.

4 When is seeing something green knowing that it is? The stock answer is: when it really is green and our seeing it that way is somehow justified. The second part of this answer I shall follow Ramsey (1929b) in taking to be a matter of reliable causation. In a good light, a thing's being green will reliably cause me to see it green when I look at it, and its not being so would reliably cause me to see it otherwise. In those circumstances, my seeing it green amounts to my knowing that it is.

The problem this answer presents here is that it appeals to the thing's really being green. This fact is needed both to make my belief true and to justify it by being a reliable cause of it. But I am trying to specify this very fact in terms of the thing looking green; and it has to be shown how that can be done without trivialising this answer. There must be more to being green than looking green, or the answer will put no objective constraint at all on our knowledge of colour. And what more there is must be statable without appeal to the colour of things.

We can start by saying that a thing is green when it looks green to all or to most people. It should be clear from **3** that there is no vicious circularity in saying this and, although it is not quite right, it is a good first approximation; and seeing what it entails will enable us to improve on it. What makes something seen by reflected light look green to most people is that it reflects light predominantly of certain wavelengths, and these make most people see things green. Indeed light of these wavelengths makes everyone see things green who is not colour-blind nor has something else wrong with his eyes. In other words, there are laws linking the properties of things, of light, and of our eyes and brains, to the colours which things look to us to have. Those laws are what make certain things such that all or most people would see them green; more precisely, such that all normally-sighted people would see them green under normal white light. And that is what it is for something really to be green.

Note that this specification of greenness is not circular. Apart from how things look, it mentions no colours, only such properties as reflecting light of certain wavelengths. Nor is it as unduly democratic as it seems. Normal sight is specified physiologically, not by Gallup poll; and normal white light is specified as a certain spread of wavelengths. As with our paradigmatically green grass, while we naturally choose a specification which fits our actual lighting and eyesight, that fact is not built into the meaning of the word 'green'. It makes perfect sense to imagine radiation so affecting our eyes that we would all come to see green things red.

Nevertheless, being green is specified in terms of how things actually look to us. There remains an essential sense in which a blind man does not know what green is, because he does not know what it looks like.

Still, in order to be green, a thing must satisfy certain laws. Colour is therefore not a purely phenomenal property; indeed, like Hesse (1974 ch. 1), I suppose that nothing is. It is, in fact, obvious that even our everyday observation of colour is not just a matter of recording visual sensations. We have digested enough of the relevant laws to enable us to see the same objective colour through very varied visual experiences by allowing, quite unconsciously, for other factors like shadows and reflections. In the same way, we can likewise see varied colours *via* identical visual experiences. A dimly lit white surface, for example, will give us the same visual sensation as a brightly lit grey one; yet we can tell the colour difference, by seeing how differently they are lit. And even with all this unconscious correction, we know our colour judgment is still not infallible. Even when well-informed, it can still be overruled by evidence that the lighting is odd or our eyesight defective.

We can use our laws to anticipate our visual sensations, as well as to check and to correct them. Thus we can know – or even, in paint manufacture, make – something to be green which no one has ever looked at. Our blind man may not, in one sense, know what green is; but by knowing its laws, he can say what it is, case by case. He can say, for example, that what makes monochromatic light green is its wavelength lying between such-and-such Angstrom units. In other words, these laws supply truth conditions for statements about the colour of things: 'monochromatic light is green' is true just in case its wavelength lies in the specified range.

Now we who can see well enough to discover and test the laws that govern colours need not be able to formulate them all in order to judge the colours of things. But the existence of some such laws, and of the truth conditions they entail, is nonetheless our warrant for admitting colour to be an objective as well as a phenomenal feature of the world. It is what entitles us to propose seriously the seemingly trite truth condition: '*X* is green' is true if and only if *X* is green. The warrant assures us that there really are such facts as things being green, which can therefore determine the truth of statements and beliefs, whether about colour or about other matters. But then we can in particular use these facts to give the stock answer to our original question. Seeing something green amounts to knowing that it's green just when it really is green and our seeing it so has been reliably caused, e.g. by the fact that it is green.

5 In saying when it is that seeing something green is knowing that it's green, I have appealed to the content of the belief, namely that something is green. The content is what specifies the fact which, if it obtains, both makes the belief true and provides in perception a reliable cause of it. But this sort of account appears not to work for knowledge of chance, because chance is supposed to constrain the strength of a belief, not its content. Knowing the chance of an event is having a justified degree of belief in its occurrence. The chance does not figure in the content of this belief; nor does its content specify the fact we need in order to justify its degree. Neither the event's occurrence nor its non-occurrence justifies believing in it to degree 0.5 say, rather than to degree 0.3 or 0.6. So what, if anything, does?

Frequency is the obvious candidate: how often such an event occurs in suitably similar circumstances. At first sight this seems to enable us after all to account for knowledge of chance on the standard model: knowledge of chance is true belief, causally justified, in the relevant frequency. In fact, this answer will not do. Suppose there is only one such event: e.g. a coin is tossed once only. The frequency of heads will be 1 or 0, but this clearly does not make 1 or 0 the chance of heads on that toss. And even supposing we do have a credible frequency, say fifty heads in a hundred tosses, still this would not justify that degree of the relevant belief. For if this frequency is a chance at all, it must be the chance of heads on each of those hundred tosses, i.e. it must prescribe the degree of belief we should have in heads being the result of any one of them. Now the whole point of such a chance is to be knowable in advance. We must be able to justify belief of degree 0.5 before knowledge of the actual result enables us to justify belief of degree I to 0. But the frequency depends on the actual results; it is not causally available in advance. The frequency of heads in a hundred tosses could not be a justifying cause of my degree of belief in heads on the first toss in the way a thing's being green can be a justifying cause of my seeing it green.

So actual frequency will not serve. The right one usually does not exist, and is causally useless when it does exist. The frequentist's next resort is hypothetical long-run frequencies: how often the coin *would* land heads if only it were endlessly tossed in the same way. But this will not do either. Coins are never endlessly tossed: hypothetical frequencies are fictions, not facts, and fictions can neither cause beliefs nor make them true. In particular, there is no one frequency of heads among merely possible tosses of a coin. Tosses with other frequencies of heads are just as possible as those with the frequency 0.5 which is supposed to justify that degree of belief.

It must, if anything, be a fact about the actual toss that, were it to be endlessly repeated, 0.5 would be the long-run frequency of heads. This fact is the disposition postulated by the propensity theories I disowned in **2**. The disposition might indeed exist, and it might even cause the degree of belief it is supposed to justify. But how is it supposed to justify it? What has a disposition to produce a long-run frequency of heads on other, mostly non-existent, tosses to do with my prospects of getting heads on this actual toss? It is no use doing what Hacking (1965 p. 135) does with his 'frequency principle', namely in effect defining the concept of justification or support to be such that this disposition supplies it. That just provokes the question: why should I adopt for this toss the degree of belief that is justified in that sense? And to that question I know of no sufficient answer. We cannot get to justified degrees of belief starting from frequencies, actual or hypothetical, or from dispositions to produce them. The only way to start, as with colour, is by specifying the fact by the degree of belief it is supposed to justify. And if we do that, the frequencies, as we shall see, will take care of themselves.

6 So now I must say how chance can be specified as a fact justifying degrees of belief, if not by making beliefs about frequencies true and justifying them. I also have to show that chance, when it is so specified, will relate to frequency in the way we know it does; and it is, in fact, simpler to tackle this problem first.

I start therefore by assuming that there are chances, waiving for the moment the question of what they are and how they work. The immediate problem is to relate them to frequencies. Now I believe I solved this problem in chapter 8 of my (1971), but Salmon (1979) thinks not. So, without repeating the solution in every detail, I need to sketch its salient features and defend it briefly against Salmon's attack.

Suppose a degree p of belief in a coin landing heads is justified by knowing that p is its chance of doing so. To get what follows from this supposition about the frequency of heads on repeated tosses, I appeal to the laws of large numbers. By these it follows in particular that as high a degree p^* of belief as we like, short of 1, is justified in any frequency proposition F of the following form: in enough such tosses, the frequency of heads would be within ε of p, where ε is any positive real number, however small.

This very high degree p^* of belief in the proposition F looks as if it needs another chance to justify it; and if that were so, there would be an arguably vicious regress. But p^* does not need a chance to justify it, as the following argument shows.

In chapters 1 and 2 of Mellor (1971), I argued that a degree of belief increasing towards 1 must turn into full belief before it gets there. Salmon himself remarks (1979 p. 187) that no one's beliefs would make him risk unlimited loss for a penny gain, i.e. that none at least of our contingent beliefs is actually of degree 1.[1] For a belief to be justified, therefore, its justified degree need not be 1; it need only be a degree sufficiently close to 1. This being so, it follows that as a justified degree of belief tends to 1, it will turn into justified belief somewhere before it gets there. Where it does so will depend on context, but that is immaterial here, since the laws of large numbers can get the justified degree of our belief in any F as close as any context could conceivably require. So whatever the context, our assumption justifies us in simply believing every F, and hence in believing that, in a sufficiently long run, the frequency of heads would come indefinitely close to the chance p. And what makes that belief true is just that the toss has the disposition, to produce long-run frequencies, to which I alluded in the last section.

In short, specifying chance by the degree of belief it justifies does not in fact lead to a vicious regress; nor will it sever any of chance's proper frequency connotations. But what exactly is the specification, and how does it work?

7 I must first say what it is for a degree of belief to be justified by a fact about an event. The sense of justification we require has to be factual rather than, say, moral: an outcome of the coin toss is to some degree to be expected, not to be approved or deplored. Yet justification here is not a matter of making a belief true, since truth applies to the content of a belief, not to its strength. There needs to be some other mode of justification by facts which relates specifically to the degree of a belief rather than to its content.

To see what this mode of justification is, consider again the connection between betting and degree of belief mentioned in **2**. I argued in Mellor (1971) that people's choices of coherent betting quotients (CBQs) show how

---

[1] But what, Salmon then asks, if I believe a chance to be 1, as I clearly can and often do? Does that not destroy a correspondence between chance and degree of belief which my theory requires? Not at all. If none of my contingent beliefs is of degree 1, then in particular my belief that the chance of an event e is 1 will not have that degree, nor therefore will my belief that e will occur. For example, suppose I am convinced (to a degree very much closer to 1) that the chance of e is either 1 or 0.5, my degrees of belief in these chances being very close respectively to 0.99 and 0.01. Suppose also that in the circumstances this belief of degree 0.99 is strong enough to count as belief, i.e. no increase in its strength would affect any of my actions. Then I do believe the chance of e is 1. But it does not follow that my degree of belief in e occurring is 1. On the contrary, given these suppositions, it follows that my degree of belief in e is very close to 0.995: enough therefore to constitute believing that e occurs, but not believing it to degree 1.

strong they think their beliefs are, provided they suppose the betting situation to be restricted in specified ways in order to exclude any effect on their choice of attitudes, other than the belief whose strength is to be measured. I then used the entailment just established to show that only at a CBQ = p can I know that I would break even in a long enough run of bets on coins landing heads when their chance of doing so is always p. Now under the restrictions needed to make my CBQs measure my belief, I must suppose myself compelled to bet without control either of the stake or of the direction of the bet. In that situation breaking even is the best result I could possibly hope to know of. So I have a plain gambling rationale for choosing this CBQ in that situation and, hence, for having the degree of belief which, in that situation, this CBQ measures.

We need something like this argument to show in what sense a chance can be said to justify a degree of belief. I have therefore to meet an objection which Salmon (1979 pp. 188–92) has brought against it. The objection is to my definition of breaking even. In Mellor (1971), I defined breaking even on a run of bets as having a net loss less than the least unit of the currency, on a fixed stake divided equally among however many bets there are in the run. Salmon, on the other hand, supposes a fixed stake on each bet, and defines breaking even as being no worse off after all the bets in the run. Betting Salmon's way, there is no CBQ at which I can know I would break even: the net loss is as unlimited as the number of bets – not surprisingly, since the total stake is likewise unlimited. But that is not just putting one's shirt on, as Salmon has it: it is putting on an indefinitely extensive wardrobe that no one could ever have. In real life, we have no more than a limited amount of goods, however large, to wager with; and breaking even is, I still maintain, losing less than whatever fraction of that equals the smallest unit of the currency.

However, the core of the argument would survive even in a casino lucky enough to attract Salmon's unlimited patronage. Its patron could indeed never know that he would break even; but he could know how to lose his wardrobe as slowly as possible, namely at a CBQ = p. And as that is the best knowledge he can have in the circumstances, p is still the uniquely justified degree of his belief.

8 For me to think there is a chance of heads is thus for me to have a degree of belief in heads and to think it justified in the above sense by some fact about the coin toss. If I am right, the fact will be that the toss has some property, such that at a CBQ contained in the shortest interval which measures my

belief, I can know I would eventually break even on bets on heads resulting from tosses all of which have this property.

The measure of the property, namely the chance of heads, is the degree of belief which it justifies in this fashion. Like colour, chance is being specified by how it seems. A man incapable of doubt, i.e. whose beliefs are all of degree very close to 1 or 0, does not know what chance is, in precisely the sense in which a blind man does not know what green is. He is never in the mental state characteristic of one who perceives that property.

But a man without doubts can still say what chances are, case by case, just as a blind man can say what colours are. Neither man can say it *a priori*: nothing *a priori* prescribes how alike in other ways things or events have to be to share the same colour or chance. Laws of nature are what prescribe that, as we have already seen in the case of colour. In the case of chance, the laws in question are statistical laws. A statistical law is one which says that having some other specified property suffices to fix a chance. The laws of radioactive decay, for example, specify the nuclear structure of radium to be such a property. The law says that all atoms with that structure are such that, if I were to bet often enough at a CBQ = 0.5 on them decaying in the next 1622 years, I can know I would break even; so that they justify my having that degree of belief in any one of them doing so.

In other words, laws supply truth conditions for statements about chance, just as they do for statements about colour. 'The chance of this atom decaying within 1622 years is 0.5' is true just in case the atom's nuclear structure is that of radium. Now those of us who are capable of doubt, and thus of perceiving chances, need not know these detailed truth conditions in order to detect a chance. We can detect a chance behind sufficiently suggestive frequency data, for example, or from evidence of symmetry in a coin, without knowing the particular laws which govern it. But the existence of such laws, and of the truth conditions they entail, is, as with colour, our warrant for taking chance to be an objective feature of the world. To say there is a chance is to claim such a warrant. We may not know the statistical law governing a particular coin toss, but unless there is one, the toss has no objective chance of landing heads.

9 'The chance of X is p' is true if and only if the chance of X is p. This is not as trite a truth condition as it seems. It claims there are such facts as X having a chance, and this will be so only if X is subject to statistical law. However, although this truth condition is not trite, it can be misleading in a way the corresponding one for 'X is green' is not. It can easily suggest that the rôle of

chance facts is to make true a belief with a particular content, namely a belief about a chance. But that, we have seen, is not so. Its rôle is objectively to justify a particular degree of a belief about something else. There is an important sense in which statistical science has no subject matter of its own. The subject matter of a statistical science is just that of its non-statistical counterpart.

This is why it has proved so hard to frame an acceptable account of objective chance. People naturally feel that, if chance is objective, it must make true beliefs with some characteristic content. The frequency theory's appeal is precisely that it offers such a content, namely facts about frequencies. Conversely, those who have seen that making beliefs true is not what chance is for, are naturally disposed to deny its objectivity: hence the popularity of subjectivist accounts of it. Where both parties err is in supposing that the only objectifying job facts can have to do is making beliefs true.

But this is not so, as we may see in ethics. An objectivist there will naturally give truth conditions, such as: '*X* is good' is true if and only if *X* is good. This truth condition, like that for chance, is not trite: to give it, the objectivist has to think there are such facts as *X* being good, and that *X*'s having some other properties would make that fact obtain. But he is not thereby committed to thinking that this fact's rôle is to make true a belief whose content is that *X* is good. Obviously not: its rôle is objectively to justify a mental state quite different from believing, namely approving of *X*. The content of all the relevant *beliefs* about *X* is entirely non-moral. Morality, like chance, has no subject matter of its own. It may be objective nonetheless: beliefs are not the only mental states capable of objective justification. Failure to see this underlies two views in ethics which correspond closely to frequency and subjective views of chance. One view tries to provide a distinctive content for the belief that *X* is good; e.g. that *X* promotes human happiness. The other view sees that this misses the whole point of morality (since it remains an open question whether one should approve of promoting human happiness or anything else), and concludes that there is no objective goodness at all.

Properly to recognise both objective goodness and objective chance, we need to extend our conception of facts as suppliers only of true beliefs. The extension is less for chance than it would need to be for goodness, since chance at least has to do with belief, just as truth does. Indeed I suspect that in the end truth will prove to be no more than an extreme case of chance. To call a belief true is, after all, to say that objective justification exists for believing it to a degree very close to 1. An adequate and comprehensive account of how

degrees of belief can be justified should be able to make a separate theory of truth redundant.

10 I am not, however, trying yet to sack all truth-making facts, only to employ some chance facts as well. Up to now, though, truth-makers have run a virtually closed shop in the facts business, and I foresee considerable opposition to these chancy new recruits. Are they really necessary? Is there really objective work to be done which only they can do?

The work chance facts do is justifying degrees of belief. Now, according to Bayesian decision theory, degrees of belief combine in a familiar way with degrees of desire to justify choices of action. For example: I have to go out, don't like carrying an umbrella needlessly, but also don't like getting wet. So whether I should take one depends on the degree of my belief that it will rain. Pure subjectivists would maintain that there is no need to justify the degrees of belief which in turn justify courses of action, and so there is no need for chances. But a subjectivism which is that pure is not really credible. A degree 0.99 of belief in imminent heavy snow might well justify travelling with a snowplough; but not in the Sahara in July, because there that degree of that belief would be objectively absurd.

Perhaps, however, there does not need to be a chance of snow, only an inductive probability of it relative to truth-making facts about temperature and humidity? Not so. Bayesian decision theory does need chances, as Jeffrey (1980b) has remarked. Suppose I am a smoker deciding whether to give up the habit because of my fear of cancer. For me, smoking 'dominates' not smoking, i.e. I shall prefer to smoke whether I have cancer or not. But I also know I should very much prefer not to have cancer, and I think cancer much more probable if I smoke. In short, the degrees of my relevant beliefs and desires make the theory tell me to give up smoking. And so I should, but only if what justifies the degrees of my conditional beliefs are propensities, not merely inductive probabilities. Since I prefer to smoke in any case, I would be a fool to quit if my smoking were merely better evidence than my not smoking for the hypothesis that, whether I smoke or not, I shall get cancer. Quitting can only be justified if it is an action which will cause a change in my prospects, namely a reduction in the probability of my getting the disease. But a probability which has causes is a part of the physical world, not merely part of an inductive logic: that is, as I remarked in **2**, a chance. In other words, inductive probabilities are not enough to make sense of the prescriptions of Bayesian decision theory. The work objective probability has to do there can only be done by chance.

Chances of course are not only effects, as in the case of smoking and cancer. They are causes too, and are invoked by statistical theories to explain phenomena. Statistical mechanics, for example, invokes a chance distribution over the positions and momenta of gas particles in order to explain the gas laws. The exploding of an atomic bomb is the effect of increasing the chance of it recapturing its own fission products. Chance facts figure as prominently in the causal history of the world as do the facts which make beliefs true.

Yet the need to employ chance facts might still be resisted. We have seen that statistical laws give apparently non-chance truth conditions for chance statements. To every chance there seems to correspond a truth-making fact, e.g. that a radium atom has such-and-such a nuclear structure. Do not these non-chance facts suffice to justify degrees of belief, and to fill the rôle of chances as causes and effects, as well as providing inductive probabilities? Perhaps the chance of heads is not itself a property of a coin toss, but merely a way of referring to whatever non-chance property actually justifies a particular degree of belief in that outcome of it.

This view is the statistical analogue of a longstanding view of dispositions. Fragility, it has been held, is not itself a property of a fragile glass. Rather it is a way of referring to whatever non-dispositional property, such as the glass's molecular structure, is causally responsible for making the glass break when dropped. It will then be a law of nature that whatever has this property is fragile; just as it is a law that whatever has the nuclear structure of the radium atom has a fifty-fifty chance of decaying in 1622 years. In each case it is the law which supplies the truth conditions; but given the law, the only facts we need admit are that the glass has such-and-such a molecular structure, and the atom such-and-such a nuclear one. Neither dispositions nor chances need to figure as constituents of facts in their own right.

I have attacked this view of dispositions elsewhere (see chapter 6), basically because the real properties which are supposed to provide truth conditions for fragility are as dispositional as it is. There are simply not enough non-dispositional properties of things to go round. It takes other dispositional facts to account for glasses being fragile, and for changes in fragility being causes and effects. The other facts may be microscopic, but that does not make them non-dispositional. And so it is with chance. There are not enough non-chance properties of events to go round. Nuclear physics is irreducibly statistical; the nuclear structure of radium is not in reality a non-chance fact at all. For an atom to have that structure is for there to be certain chances of nuclear events occurring within it. It takes chance facts, not just truth-making facts, to

account for radium's objective radioactivity. The facts may be submicroscopic; but they are no less chancy for that.

11 I conclude that chances are indispensable ingredients in any world governed by statistical laws. Such a world need not, however, be indeterministic, as I thought when I wrote my (1971). That was an error, as Salmon (1979) and others have convinced me, and I am pleased to take this occasion to recant. But I need then to show how my chances can be made compatible with determinism, and to rebut my earlier arguments to the contrary.

Consider first a coin landing heads with chance 0.5. The chance justifies that degree of belief in heads; but the fact of heads also justifies a degree close to 1 in the same belief. How can two different degrees of the same belief both be justified? The answer is that the two justified degrees are not actually degrees of the same belief. There are two beliefs: the past tense belief that the coin has already landed heads, and the future tense one that it will do so. The two beliefs are both made true, if they are, by the same fact, namely the coin landing heads, but they are not the same mental state. For one thing, they relate differently to action (see Perry 1979): there are circumstances in which believing a coin will land heads would make me behave quite differently from the way I would behave if I believed it had already done so. For another, although the same fact makes both true, it makes them true at different times. The past tense belief can only be true after the coin has landed, the future one only true beforehand. But since causes always precede their effects, the coin landing heads can only be a justifying cause of the past tense belief while it is true. A future tense belief of which it could be a cause would have to be about another, later event. I cannot see that a coin *will* land heads, only that it *has* done so.

So only the past tense belief is justified to a degree close to 1 by the coin landing heads; the future tense one is not. The coin landing heads does not therefore prevent the future tense belief being justified merely to degree 0.5 by facts, about the propensities of the coin and tossing device, which are causally available before the toss. I can see beforehand that a coin is fair, i.e. is such that, if tossed in a standard way, the chance of it landing heads would be 0.5. And I know it will be tossed that way, because I know I can toss it that way, and I intend to. This is how I know that there is a chance 0.5 of the coin landing heads on this future toss, and hence that the future tense belief that it will land heads is justified to degree 0.5. My knowing also that the corresponding past tense belief will then be justified to a degree close to either 1 or 0 does not detract from this knowledge at all. On the contrary, without

knowing that, I could not envisage settling a bet on the coin landing heads; and I can make no sense of justifying any degree of any belief without being able to settle bets.

Generally therefore, it is only future tense beliefs which are justified to degrees remote from 1 and 0. (Betting on past events can of course occur, but it always depends on the future disclosure of crucial evidence not yet available to us.) But it is the degrees of future tense beliefs which matter to decision theory, as the example of **10** shows. If smoking makes it probable that I will get cancer, that can make it sensible for me to stop smoking. But not if it only makes it probable that I already have cancer, i.e. that it justifies only a high degree of the past tense belief. Just because causes always precede their effects, such a past tense probability could not be an effect of my present smoking. It could only be an inductive probability, not a chance, and my decision therefore should not turn on it.

Similarly with coin tossing. If it matters to me whether a coin lands heads, then all I can hope to affect is the probability that it will do so, not the probability that it has already done so. If I toss it one way, the chance of heads is p_1, if I toss it another, the chance of heads is p_2: these are propensities of tossing devices. Knowing them is what justifies the degrees of the future conditional beliefs which I need in order to decide which way to toss the coin. So long as propensities and the chances that display them can justify intermediate degrees of future tense beliefs, they do all that decision theory needs. That, so far as past tense beliefs are concerned, they are overridden by the actual results of the tosses is neither here nor there.

But now suppose the result of the toss is, in fact, determined by earlier facts, albeit ones not readily observed. Do they not justify a degree close to 1 or 0 in the future tense belief? And if so, how can a degree 0.5 be justified in the same future tense belief? For recall the betting situation I appealed to in giving a relevant sense to justification: in it, the believer's opponent, not himself, fixes both the stake and the direction of the bet: i.e. fixes who wins if the coin lands heads. If the result is determined, and thus perceptible, in advance, anyone betting at a CBQ of 0.5 would surely be cleaned out; and wherein then lies the justification of that degree of belief?

These were the considerations which persuaded me in my (1971) that chance was incompatible with determinism; since I no longer believe that, I have to say what is wrong with them.

The fact is that our theory of knowledge needs a causal account of justification, whether of beliefs about colour or about chance, only because our conceptual and perceptual abilities are limited. Were that not so, we should not

need to distinguish knowledge from true belief in terms of justification at all. In the case of God, for example, we need no such distinction. What God knows, He knows immediately, not through the mediation of senses such as we have. He does not acquire His knowledge by sensory mechanisms of possibly variable reliability, of whose actual reliability we must therefore be satisfied before conceding that a belief is more than coincidentally true.

But we are not gods. Our knowledge comes to us by fallible means, and we need to distinguish it from coincidentally true belief. That is why we need a theory of justification for belief, a theory therefore of our limited perceptual and conceptual abilities. Since it is our theory, our conceptual limitations will be built into it; what it will deal with explicitly, and what anyway matter here, are our perceptual limitations. These need not be specified precisely, but they must be presumed. In particular, if our gambler cannot tell in advance that a coin will land heads, we must suppose that his opponent cannot tell that either. We are concerned with bets amongst ourselves, not bets with the All Seeing – a self-evidently foolish pastime at any CBQ.

So intermediate degrees of future tense belief can be justified, and chances can exist, even in a deterministic world. Determinism therefore I now take merely (but still falsely) to assert that every event instantiates the consequent of a non-statistical causal law, not that it does not also instantiate a statistical one.

12 By the same token, I can make sense of an event instantiating more than one statistical law, attaching different chances to the same result. Most events do not, but some do. The famous example is Laplace's biased coin, of which the bias is unknown. It seems that on the first toss, the right CBQ for a bet on heads is 0.5, even though the chance is known to be something else; how can that be on my account? The fact may be that there are two chances of heads: 0.5 as well as the chance that displays the coin's bias. There is nothing paradoxical in one outcome of an event having two chances, even if it is uncommon. All it means is that two degrees of the same belief could both be objectively justified; and that can be so even if no one could actually have them both at once. It can be so, because the CBQs that measure them enable knowledge of breaking even in two different betting situations, depending on what counts as repeating the toss. In one, the same biased coin is repeatedly tossed, and it is at some CBQ other than 0.5 that one can know one would eventually break even. In the other, the other bets would be on other biased coins, of which we suppose half would in the long run be biased towards heads. The fact we suppose, then, is that biasing mechanisms have equal

propensities to bias coins towards heads and tails; and it is our supposed knowledge of this propensity which justifies the degree 0.5 of our belief in heads on the first toss. (For the detailed proof of this, see Mellor 1971 pp. 129–36.)

But which degree of belief should we actually have on the toss of such a coin? Well, either 0.5 or the degree that reflects the bias of the coin would do. But the value 0.5 is known, insofar as we know that biasing processes do not distinguish heads from tails. Whereas the other value is unknown, because our perceptual limitations, Laplace supposes, prevent us perceiving the actual bias of the coin. So it is no more practicable to recommend that other value than it is to recommend full belief in whichever side will actually land heads. And that is why, *faute de mieux*, one should bet at evens in this case.

13 Chances may thus be said to be relative to our perceptual limitations. This does not make them relative to all descriptions of events, nor to the actual frequencies which those descriptions determine. A coin toss may be one of seventy made within the shadow of St Paul's, of which twenty-eight land heads. It does not follow that there is this or any other chance of landing heads that is common to all these tosses. There is no such chance: no statistical law applies to tosses under the description 'thrown within the shadow of St Paul's', and there is no CBQ at which we could know we would eventually break even on repeated bets on such tosses landing heads. Yet any one of these actual tosses may, under some other description, be governed by one or more statistical laws; and if so, one or more degrees of belief in its outcome being heads could be justified in beings of suitably limited perception.

I have specified chances by the degrees of belief they can justify in beings of our causally limited perceptual ability. In the same way, I have specified the colour of electromagnetic radiation of various wavelengths by colour judgments they can justify in beings of our causally limited eyesight. That is a perfectly good way of specifying objective properties: ultimately indeed the only way, since science must deal in phenomena, not noumena, if its epistemology is not to be merely magical (see Mellor 1973). But whether there are beings like us or not, the properties so specified remain. The effects of radiation and of chance are, we have seen, not limited to their effects on us. Maybe the All-Seeing should have no degrees of belief other than 1 and 0; but he can still know the world to be such that we should. That is as much a fact about it as any other; and so therefore is chance.

15 *The warrant of induction*

1 Introduction This lecture will last less than twenty-four hours. I know that, and so do you. And you knew it before I said so. How? Because you knew that lectures don't last twenty-four hours. How do you know that? You haven't heard this one, and 'for all you know' (as the saying is) I could go on all night. But you know I won't. And the 'all you know' which tells you that, without entailing it, is the fact that none, or almost none, of the many lectures, on all subjects, which you've heard or heard of, have lasted that long. If many of them had, you wouldn't have known that this one won't; but as it is, you do know that.

That's a piece of induction. We believe that something (a lecture) has an observable but as yet unobserved property (a short duration), and this belief is warranted by the fact that many otherwise diverse things of the same kind (lectures) have all – or nearly all – been observed to have that property. The problem, set for us by David Hume in 1739, is to say why such observations warrant such a belief.

The greatest of all Cambridge philosophers, Frank Ramsey, said in 1926 that 'we are all convinced by inductive arguments, and our conviction is reasonable because the world is so constituted that inductive arguments lead on the whole to true opinions' (Ramsey 1926 p. 93). As usual, he was right, but too brisk for most philosophers. Most are still unpersuaded, despite the subsequent work of Professor Braithwaite (1953 chs 6–8) and others. Tonight I will try to say, less briskly, but still within twenty-four hours, why I think Ramsey was right.[1]

First I must say what it is to warrant a belief. Or rather, what it is about a belief that is to be warranted. What's to be warranted is its truth, that is the truth of what's believed: for example, that this lecture will last less than twenty-four hours. That's why observations which warrant a belief can't also warrant a contradictory one: contradictory beliefs can't both be true. Yet our

[1]What follows of course also owes much to other writings on induction and knowledge, notably Ramsey (1929b), Armstrong (1973), Goldman (1980), Grandy (1980), Hacking (1980), Levi (1980b), Nozick (1981 ch. 3) and van Cleve (1984).

inductive warrant doesn't entail that the belief it warrants is true. The belief that this lecture will stop tonight would be warranted now even if I went on to make that belief false; and the belief I would then make true would still not be warranted now. But how can a warrant be valid and yet fail? That's part of the problem of induction.

The problem isn't just that your inductive evidence fails to entail the belief it warrants. It's worse than that. For that belief is a prediction, about when this lecture will end. And its success as a prediction will be settled in due course by an observation: for example, by your consulting a watch when the lecture ends. Some such observation will settle that prediction regardless of its present inductive warrant. If I keep you here till dawn, your present inductive evidence that I won't will count for nothing against the evidence of your senses that I have done. And if I stop sooner, your senses will need no help from induction to tell you so. You know that your inductive warrant for what you now believe is negligible compared with the warrant that some future observation will give it – or deny it.

Yet we trust such predictive beliefs incessantly, often with our lives. Take the belief that this building will stay up while I speak two more sentences. We have, I hope, strong inductive warrant for that belief, and thus for staying here at least that long. Yet the belief is still only a prediction (comma), because we know that its present warrant will be quite superseded by that of our senses ... now. But how can a warrant that is now so weak ever have been strong enough to trust our lives to? And as in this case, so in general: how can a prediction ever be better than a guess? That's another way of putting the problem of induction.

2 Induction and observation I'm taking induction to be something that can warrant anticipating an observation, the problem being to say how. This reading of induction sets limits to the problem, and makes assumptions about observation, which I must briefly defend. It limits the problem by excluding other cases in which evidence might warrant a belief that it doesn't entail. Take mathematics. All even numbers so far checked are the sum of two prime numbers. This doesn't entail that they all are. Does it warrant that belief? I don't know; and if it does, I don't know how: but not, I suppose, in the way that beliefs about how long this lecture will last are warranted. So no one should object to my account of that, just because it won't also explain the force (if any) of inconclusive evidence in mathematics.

More seriously, my account won't explain everything that has been called induction in science. It won't explain how observations give rise to new

theories: because new theories typically predict new kinds of entities (new subatomic particles, say), which what I call induction doesn't. Nor will it always explain how observations warrant existing theories. A theory's entities may not (yet) be observable; and until they are, induction can only warrant what the theory implies about other things that are observable. How that can warrant believing the theory's unobservable implications – if it can – must be quite another matter.

Induction may not be enough to warrant a theory in the sense of warranting belief in it or in something like it. But it is necessary. Since warranting a belief means warranting its truth, it also means warranting anything it entails. And theories entail generalisations which if true state laws about the observable phenomena they explain. To warrant a theory we must warrant those laws. But we can't warrant them by observation directly, since we can only ever observe instances of a law, never the law itself. For laws may have future instances as well as past ones, and since we can't observe the future, we can never observe that a generalisation will never be refuted by some future instance: some interminable lecture, or some wayward subatomic particle. So we can warrant a law by observation only by observing some of its past instances, and then only if this warrants predicting, of any other instance, that observing it would give the same result. But this is what I call induction. Induction may not be the whole story about how observation can warrant believing laws and theories; but it must be a part of it.

Now to observation. I've said that observation is what settles predictions, including theoretical predictions. This doesn't mean that I think observation is independent of theory. We must hold some microphysical theory in order to see a bubble chamber track as the track of a subatomic particle. But a prediction about that particle's track may still be warranted inductively by observations of previous tracks, and be refuted by observing this one. The fact that observers need theories doesn't stop theoretical predictions being warranted inductively by past theoretical observations, or overthrown by future ones.

Nor need observations be infallible to settle predictions. Of course observations can yield unwarranted beliefs (in mirages, for instance); they can need inductive support; and their warrant can sometimes be outweighed by inductive evidence. A mere glimpse may not warrant the belief that there are sparrows in the park, unless sparrows have been seen there many times before. It may take a long look by a good birdwatcher to warrant believing in the presence of a bird that's both nondescript and rare. But all this shows is that inductive evidence can outweigh a glimpse, not that it can outweigh

observation altogether. It can't, since inductive evidence consists of observations of the very kind that it lets us anticipate: observations of birds, subatomic particles, lectures. Seeing that some birds are sparrows can't warrant predicting that a new one is as strongly as it warrants the belief (that the old ones are) on which this prediction is based. But then, by the same token, an equally good look at the new bird must warrant a belief about it more strongly than the observations of those other birds could: so we must be able to settle the prediction that it's a sparrow by looking at it. In short, observation can always outweigh an inductive warrant for any prediction about anything observable – as everyone knows who's ever been unpredictably stood up.

But observations are still fallible, a fact we obscure when we use words like 'observe' so that by mere definition we can't observe what isn't there. So I will not use 'observe', 'see', 'hear', etc. in that misleading way. When my looking to see if there are sparrows makes me believe that there are, I shall say that I see that there are, whether there are or not. For there might not be: our senses always can give us false beliefs and sometimes do. Observation no more entails the truth of the beliefs it gives us than induction does. How then can it warrant them?

Well, not of course by induction, so answering this question is no part of my present task. But answering it will help, by enabling me to reset a standard of warranted belief which has hitherto made the problem of induction seem insoluble. For as we have just seen, observation must be able to warrant any beliefs that induction can warrant. Yet by the standard that induction is generally condemned for failing, observation couldn't warrant them. But that's incredible: no one really doubts that observation can and does warrant most of the particular everyday and scientific beliefs it gives us. So we need a more credible standard of warranted belief anyway, one that observation can meet. If we can find such a standard, and then show that, and how, induction can meet it too, the problem of induction will be solved. And that's how I propose to solve it.

3 Observation and warranted belief How should a standard of warranted belief be set? A standard implies some goal or object whose achievement meeting the standard should ensure. Take the safety of goods, which is the object of many British Standards. Meeting those Standards should ensure that goods are safe: that is the test of a good safety standard. Here the object is not safety but truth. To warrant a belief is to warrant its truth. So the test of a

good standard of warranted belief is this: would the truth of a belief whose warrant meets that standard be thereby ensured?

How can a warrant, like that of our senses, ensure the truth of a belief it doesn't entail? Well, suppose I see that there are sparrows in the park. This belief will be true if and only if there really are sparrows in the park. So my observation should warrant my belief if it's such that, if I'm seeing sparrows there, then they're there.

What would ensure the truth of this conditional proposition? We know an entailment would. If my seeing sparrows entailed that they were there, the proposition that they're there if I'm seeing them would be not merely true but necessarily true. And if that were so, then seeing them would certainly warrant believing they were there. But it isn't so. Seeing sparrows doesn't entail that they're there. Our conditional proposition is not a necessary truth.

Still, propositions can be true without being necessarily true. Suppose you have a mass of M units. This fact makes true a whole raft of conditional propositions of the form 'if a force of F units is applied to you, you will accelerate at F/M units'. And since your mass is contingent – we could all be more or less massive than we are – those conditionals are all only contingently true.

Now suppose some similar contingency makes it true that if I'm seeing sparrows then they're there. Why can't such a contingency make this observation warrant that belief? After all, it still can't also warrant the contradictory belief that sparrows aren't there. The conditionals 'they're there if I'm seeing them' and 'they're not there if I'm seeing them' can't both be true while I'm seeing them, however contingent they are. Nor could the warrant both hold and fail. The sparrows can't fail to be there when I'm seeing them while the conditional 'they're there if I'm seeing them' is even contingently true.

So far so good. But what might our contingent warrant be? Not just that sparrows happen to be there while I'm seeing them. That does arguably make it true then that they're there if I'm seeing them. But if so, that conditional is then only made true by the truth of the very belief it's meant to warrant: namely, the belief that the sparrows are there. So if that were the warrant, the truth of that belief would in effect be warranting itself; and it can't do that.

For our purposes we need something more, to make a conditional of the form 'if P, then Q' true, than the mere fact that P and Q are true: something more like mass. Your mass doesn't only entail an infinity of conditionals, about how you would accelerate under various forces. It guarantees that, while you keep it, the same force would always give you the same acceleration. Mass is a deterministic disposition: if a force F is applied to a mass M, the chance –

the physical probability – that it will accelerate at a rate F/M is 1. And that makes the conditionals which a mass entails non-trivially true: since the mere fact that M does accelerate at a certain rate doesn't entail that it had no chance of doing otherwise.

Your mass in fact embodies a causal link between forces and accelerations: it's what makes forces cause you to accelerate at a certain rate. And just such causal links, between us and what we observe, are what enable our observations to warrant the beliefs they give us. To see, for example, is to be affected by light which something emits or reflects into our eyes, the relevant effect being that we get a belief about the source of that light. I see a sparrow when the light from something causes me to believe that it's a sparrow. My belief is true if it is a sparrow, and warranted if the causal link between me and it makes it non-trivially true that, if I believe it's a sparrow, then it is.

How does the causal link make this conditional proposition true? By making sparrows cause me to believe that there are sparrows there, and other things cause me to believe that there aren't. If the causation is deterministic, the chances of both these effects will be 1, from which it follows non-trivially that if I do see sparrows, then they're there. That's what makes my seeing them warrant my belief that they're there.

The causal links, and hence the chances, which this warrant depends on are of course contingent on many facts about me and the sparrows: that I have the concept of a sparrow, decent eyesight and my eyes open; that the sparrows are not only there but well lit and in view; and so on. Those facts might have been otherwise: so my belief might not have been warranted. But as things are it is warranted, since as things are there's no chance of my observation giving me a false belief.

But causation needn't be deterministic. An effect's chance must indeed be greater if a cause occurs than (in the circumstances) it would be if it didn't occur (see chapter 13). But its chance with its cause may still be less than 1, and its chance without it may be greater than 0. Smoking may for example cause a man to get cancer even if he isn't certain to get it if he does smoke, and even if he might get it without smoking.

And even an indeterministic causal link can make my looking for sparrows warrant my beliefs about them, by making their presence nearly always cause me to see them, and their absence nearly always cause me not to see them. The warrant is worse, of course, when it makes my chance of getting a true belief less than 1. But a 99% or 95% chance may still be good enough: it will still make my seeing sparrows warrant the belief that they're there far more strongly than the belief that they're not.

This kind of chancy causation moreover enables observation to be fallible, since it lets me sometimes see sparrows that aren't there. Now when that happens, it isn't true that the sparrows are there if I'm seeing them. So that conditional needn't be true after all. Observation can warrant a false belief, provided it gives the belief a small enough chance of being false.

This, I maintain, is how the greater chances that effects have when their causes occur than when they don't enables observation to warrant beliefs: contingently and fallibly, but quite well enough. And chances which can do that can enable induction to do so too. But before showing how, I must briefly dispose of two objections to the standard of warranted belief that I've so far been appealing to.

4 Chance and causation The first objection is to my concepts of chance and causation. To serve our turn, chance must be more than a merely epistemic probability: that is, more than a mere measure of how strongly beliefs are warranted. To say that a belief is warranted when it's epistemically very probable is just to say that it's warranted when it's warranted. This doesn't tell us what the warrant is, nor why it is one. Talk of chance can tell us that only if chance can be specified and detected independently of how it sometimes warrants our beliefs, but yet in a way which shows how and when it does warrant them. And so it can be.

We specify and detect chances mainly as indeterministic causes of statistical facts. Suppose a particular coin tossed in a certain way W has a chance C (between 1 and 0) of landing heads. This entails the existence of a statistical law, namely that if enough sufficiently similar coins were tossed this way, there'd be a very high chance that a fraction close to C would land heads. Suppose many such coins are so tossed, and the fraction that land heads is in fact close to 0.6. The chance of that happening is greater if C is 0.6 than if it isn't. And if C is 0.6, that is the cause: the fact that the chance of each of these many W-type tosses landing heads is 0.6 is what causes – and thereby explains – the fact that the fraction which do land heads is close to 0.6.

Many statistical facts about distributions of physical and biological properties are similarly caused by chances. Take a pure lump of the most common isotope of radium. In any given time a closely constant fraction of its myriad atoms will decay: in 1622 years, about 50%. That's because – literally because – each of those atoms has this chance of decaying in that time. Or take the fact that about half the foetuses conceived in large populations are female. This too is an indeterministic effect, of each of those many conceptions having a 50% chance of producing a foetus with only X-chromosomes. And so on.

Chances also have causes. Bending a coin affects its chance of landing heads when tossed. Bombarding a radium atom raises its chance of decaying. Starting to smoke causes people's chances of getting cancer (and hence the fraction of people who do get cancer) to increase.

Chances, in short, don't only help to link other things as cause to effect. They too are causes and effects: because they have causes and effects, by which we can specify and detect them, quite independently of how they enable observation – and, as we shall see, induction – to warrant our beliefs. Most chances actually warrant no beliefs at all. They only warrant them in observation because there the effects of some fact include someone's getting a belief: ideally the very belief which that fact makes true, like the belief that there are (or that there aren't) sparrows in the park. This is what turns the chance C of those effects into a probability that whichever belief is thereby caused is true. And what then makes C an epistemic probability is its entailing that many such observations would give us a very high chance of getting a fraction of true beliefs very close to C. So if C is 1, such observations would always give us true beliefs. And if it's very close to 1, they would give us a very high chance of almost always getting true beliefs. What better measure could there be of the prospects of truth which observation gives the beliefs it produces, and hence of how strongly it warrants them?

5 Knowing and knowing you know Well, one answer to that rhetorical question is: a self-intimating measure. Surely a warrant for a belief of mine is only worth having if I know I have it, so that I know which of my beliefs are warranted and therefore worth trusting? But this standard for warranted beliefs doesn't make belief-warrants self-intimating. This is the second objection to it.

Now I could just add to my standard the rule that belief-warrants must be self-intimating. But I won't: because that rule in fact makes beliefs almost impossible to warrant, while doing nothing to ensure either that warranted beliefs are true or that their warrants are useful to those who have them.

This all follows from a conceptual link between warranted belief and knowledge. The link is that knowing about something is having a true and warranted belief about it. I will call this the 'true warranted belief' thesis. I assumed it at the start when I said that you knew this lecture would last less than twenty-four hours, and inferred that you had a true and warranted belief to that effect. And although the thesis is disputed, I shall go on assuming it, just because it causes trouble and is at least plausible. So I'd better assume that it holds.

The trouble it causes is this. Suppose, to vary the example, you have perfect eyesight and look in good daylight at a green frog on a red leaf. This causes you to get a belief about the frog's colour which in these circumstances has a very high chance of being true. Suppose it is true: you get the belief that the frog is green. I say your observation warrants this belief (belief 1). So, by the true warranted belief thesis, you know the frog is green.

Now try to make this warrant self-intimating: if you have it, you know you have it. That is, by the true warranted belief thesis, you believe you have it, and this belief (belief 2) is true and warranted. So, since this warrant too must be self-intimating, you believe belief 2 is warranted; and *that* belief (belief 3) is warranted. So you also believe that, that belief (belief 4) is warranted and so on, rapidly *ad infinitum.*

If you can believe all that, you can believe anything. In fact, I doubt if you have any of these extra beliefs, and you certainly won't have all of them. But you don't need any of them. Consider for example that part of the warrant for your belief that the frog is green must be that you aren't red–green colour-blind. But you may well not believe this, because you may never have heard of colour-blindness. That can't stop you knowing that the frog you see is green, or no one could have known anything about colours before the phenomenon of colour-blindness was discovered. So your warrant for believing that the frog is green can't have to be self-intimating: you must be able to know that something is green without knowing that you know that.

If belief-warrants, and therefore knowledge, had to be self-intimating, you could know almost nothing that observation tells you. Suppose you do believe you aren't colour-blind. To be knowledge, this belief must be warranted. How? By your having passed a colour-blindness test? But for this warrant to be self-intimating, you must believe you've passed a colour-blindness test, and that belief must be warranted: say by the test having been tested to see if it really can pick out people who are colour-blind. But how could that have been done if, until it is done, no one can know whether anyone is colour-blind?

The fact is that if, to know something, you had to know you knew it, knowledge would entail a series of warranted beliefs which observation does not, and generally could not, give us. If observation is to warrant even a few of the beliefs it gives us, this thesis, that you always know what you know, will have to go. And so it should: because the thesis sets a quite unwarranted standard for warranted belief. That is, it does nothing to ensure that warranted beliefs are true, which is the whole point of warranting them. My knowing that if I make an observation my chance of getting a true belief is high won't in general make the chance any higher. Most of the beliefs my senses give me

are warranted no more strongly when I know they're warranted than when I don't.

Nor is it true that belief-warrants are worth having only when we know we have them and therefore know which of our beliefs to trust. We can't only trust – that is, be disposed to act on – some of our beliefs and not others. Not to trust a belief is not to have it. Suppose you want a frog, and pick up that green thing on that leaf because you believe it is a frog. If you did stop believing that this belief of yours was warranted, you might indeed no longer trust it in this way – because you would no longer have it: you would no longer really believe that the green thing was a frog.

This doesn't mean that it isn't useful for beliefs to be warranted in my sense. That is useful, because it's useful for beliefs to be true. Truth is the attribute of beliefs which ensures that the actions they cause will succeed.[2] For example, if your belief, that what you're picking up is a frog, is true, your action will succeed: it will get you what you want, namely a frog. And if this belief is warranted, in my sense, it will have a high chance of being true: so your action will have a high chance of success. This is what makes it worth having warranted beliefs. Your knowing that they're warranted would only make them more worth having if it raised their chances of being true. And generally it doesn't.

In short, the standard for warranted belief set by the 'we know what we know' thesis is both unattainable and ineffective. It's no objection to observation or induction that they can't meet it. Yet, as we shall see, this is the objection that has made the problem of induction seem insoluble. For the 'we know what we know' thesis is still oddly attractive, despite its absurd consequences. Why it's so attractive, I'm not sure. I suspect it's because we still think of ourselves as essentially conscious, and especially self-conscious: so we like to think that if we know something, we know we know it. But the fact is that we needn't, and we usually don't, know what we know. And once we recognise this fact, we can solve the problem of induction.

6 Induction Let me call events which last less than twenty-four hours 'terse'. This event, as you see, is a lecture. It's also terse, but that you can't yet see. You predict it, by inferring it: this is a lecture, so it's terse. And this inference displays a general disposition or habit: you would have drawn it of any lecture, not just this one. And this inferential disposition, like your mass (your inertial disposition), embodies a causal link: just as your mass makes forces

[2]See chapter 2.4 and Whyte (1990).

cause you to accelerate, so your inferential disposition makes your coming to believe that an event is a lecture cause you to predict that it's terse.

We've already seen how causal links can enable observations to warrant beliefs: namely, when the fact that makes a belief true is what causes you to get that belief. The rôle of causation in an inference is different. What causes you to believe the conclusion of an inference isn't the fact that makes it true, but the fact that you've come to believe the premise of the inference. How can this warrant the conclusion?

Suppose, to simplify the discussion (it's not essential), that your inferential habits are deterministic, like your mass: this habit would always make you infer that a lecture was terse, never that it wasn't. And suppose that every lecture has some chance of being terse. Then whenever your premise ('this is a lecture') is true, your conclusion ('this is terse') has some chance of being true. And if this chance is high enough, your prediction is warranted.

I shall call an inferential habit warranted (or good) if, whenever the premise is true, the conclusion has a high enough chance of being true. For then there's a very high chance that all or nearly all the conclusions of many such inferences will be true when their premises are. A particular inference may still of course fail to warrant its conclusion, because its premise may be false.[3] But that's no fault of the inference. All an inference can do is give a conclusion as high a chance of being warranted as its premise has of being true. That's good enough to warrant calling it warranted, or good.

We have of course many habits of inference. Some perhaps we're born with; most we acquire. And, as Hume remarked, we acquire most of them by induction. Mostly, the more we see that otherwise diverse things with one property F (being a lecture) all or nearly all have another property G (being terse), the more we tend to predict that other things which are F are also G.

This isn't a universal tendency: it doesn't apply to all observable properties.[4] Take ageing. The more years we adults do survive, the more we tend to predict that we won't survive another one. We recognise many such 'counter-inductive' phenomena: metal fatigue and caterpillars turning into

---

[3]This stops counter-examples to the 'true warranted belief' thesis which include inferences from premises that are warranted but untrue, as in Gettier (1963). But it doesn't stop such an inference being part of a two-stage causal process of observation and inference that still lets the observation warrant the conclusion by giving it a high chance of being true.

[4]I don't count Goodman's (1965) 'grue' and its many progeny. Green, like blue, is an observable property: so what 'grue' denotes isn't. When I see truly that a frog is green (and hence grue) my belief is caused by the fact that it's green, and not at all by the fact that I see it less than so many years after the birth of Christ (see Blackburn 1969, Jackson 1975). How we know green is a property is another question, which I needn't answer: it suffices here that it is, because belief-warrants needn't be self-intimating. Disposing of the 'we know what we know' thesis also disposes of Goodman's 'new riddle of induction'.

butterflies are two more obvious examples. But we recognise them inductively. Our mortality statistics show that the observed chances of adults surviving another year have so far always or nearly always decreased with age. That's why we tend to predict that ours will. And similarly in the other cases. Our counter-inductive tendencies always have an inductive basis.

The question is: what warrants the inductive basis? We certainly think it's warranted. Counter-inductivists aren't just odd: they're mad. Imagine one. He won't eat bread: he thinks it would poison him, because it's never poisoned anyone before. He would eat cyanide, which he also expects to freeze in the oven and bake in the fridge; but not by swallowing it. He won't use any language people have so far understood, or breathe air, or drink water. And throughout his (brief) life he consistently defends his wholesale counter-inductivism by predicting that as it's almost never worked yet, it will now.

And so it could. He could be right. He could outlive us all. But he won't; and we know he won't. How? What makes induction a better basic tendency than counter-induction?

Take any pair of basic properties which we've seen to be correlated, like being a frog and being green. Suppose all or nearly all the many frogs so far seen have been green. The more, and the more diverse, those frogs and their surroundings, the more we tend to predict that other frogs will be green. Exactly how many, and how diverse, doesn't matter: what matters is what warrants this general inductive tendency.

The tendency increases both with the number of frogs seen, and with their diversity. Consider first the way it increases with the number. So suppose the frogs (and their surroundings) are all of a kind: tree frogs, for example. The more tree frogs are all seen to be green, the more we tend to predict that other tree frogs are green – and the more counter-inductivists tend to predict that they're not. What makes ours the better tendency?

That depends on what chance a tree frog actually has of being green. It may be a law of nature that all tree frogs are green: that is, the chance of any tree frog being green may be 1. Suppose it is. Then all observed tree frogs will be green, so induction will always make us predict that other tree frogs are green. And this inference couldn't be better: whenever its premise ('this is a tree frog') is true, the chance of its conclusion ('this is green') being true is 1. Whereas counter-inductivists, seeing only green tree frogs, will always predict that other tree frogs aren't green: an inference which couldn't be worse.

So much the worse for counter-induction. And it's no better off if the law is that no tree frogs are green. For then no observed tree frogs will be green,

induction will always yield the good habit of inferring that others aren't either, and counter-induction the bad habit of inferring that they are. And as for frogs and colour, so for all basic observable properties. Whenever they're linked by a deterministic law, induction will always yield good habits of inference, and counter-induction will always yield bad ones. In all such cases our inductive tendency is warranted, and the counter-inductive one isn't.

So far so good – provided these warrants needn't be self-intimating. I say the law that all tree frogs are green warrants my inferring that something is green from the fact that it's a tree frog. But suppose I must know that I have this warrant. Then I must know this law. So I must believe it, and this belief must be warranted. But the law entails the very inference which it's meant to warrant: tree frogs can't all be green unless this one is. So unless my inference is warranted already, my belief in the law won't be warranted. Thus to claim that the law is what warrants this particular application of it simply begs the question of whether it's warranted at all.

This is the stock objection to contingent solutions to the problem of induction: they beg the question. And so they would if belief-warrants had to be self-intimating. But as we've seen, they don't. The law that all tree frogs are green can warrant the habit of inference which induction will then give me, just because I needn't know that it does. I can know by induction that a frog is green without knowing that law, just as I can know that it's green by looking at it without knowing I'm not colour-blind. I may know the law, just as I may know that I'm not colour-blind; but I needn't. So my saying that the law is what warrants this induction doesn't beg the question.

Deterministic laws warrant induction. And so do statistical laws, and hence the chances that entail them. For suppose it's a law that all tree frogs have a certain chance of being green. The greater this chance, the better the inference that a tree frog is green. But also the greater the chance that many tree frogs will all or nearly all be green. In particular, the greater the chance that all or nearly all observed tree frogs will be green – and hence that we inductivists will infer that others are too. In short, the better the inference, the more likely we inductivists are to make it. And the less likely counter-inductivists are (since the better it is, the less chance there is of all or nearly all observed tree frogs not being green). So far still so good.[5]

---

[5]If tree frogs' chances of being green are all 50%, the inferences, that this one is green, and that it's not, are equally unwarranted. So therefore is induction, which has the same (minute) chance of making us make either inference as counter-induction has. So induction fails here, but only because there's no success to be had. There will of course be a good inductive inference to the fact that this tree frog's chance of being green is about 50%, but that's another matter.

Furthermore, the more tree frogs we see, the less risk we inductivists run of making inferences that aren't good. For when a tree frog's chance of being green isn't close to 1, the more tree frogs we see, the less the chance that they'll all or nearly all be green and so induce us to infer – badly – that others are. That's what warrants the way in which our tendency to infer that tree frogs are green gets stronger as we see more and more tree frogs which are all or nearly all green.

That's how induction is warranted (and counter-induction unwarranted) by simple laws linking a pair of observable properties. But laws are rarely so simple. A frog's chance of being green may depend on many things. Hence the other factor in induction: the way in which our tendency to infer that frogs are green also increases with the diversity of the frogs which we've seen to be all or nearly all green.

Suppose you and I both see many frogs, but I only see tree frogs, while you also see many frogs of other kinds. How will this help you to make better inductive inferences about frogs? It won't if all frogs have the same chance of being green (or whatever): then we're on a par. But it will if they don't. For suppose tree frogs have a high chance of being green, and others don't. Then the inference 'it's a frog, so it's green' is good when it's a tree frog and bad when it's not. And because my chance of seeing only green frogs is now greater than yours (since I only see tree frogs), my chance of being induced to make this inference when it's bad is greater than yours. Whereas, provided you notice that tree frogs differ from other frogs, your chance of seeing only green tree frogs, and hence of being induced to make the good inference 'it's a tree frog, so it's green', is just as great as mine.

In short, whenever diversity in our inductive data matters, it's better to have it than not, and better to notice it than not. Of course we won't know if it matters at the time, since that will depend on laws we don't yet know. So the only warrantable tendency we can have is to be always readier to make inductive inferences when we know our data are diverse than when we know they aren't. And this is the tendency we do have.

But what if our unseen frog has no chance, high or low, of being green? In other words, no law, deterministic or statistical, simple or complex, links being a frog of whatever kind this one is with being green. Then there's no good inference from this being a frog of any such kind to its being green, or to its not being green. Neither of the inferential habits which induction or counter-induction might give us is good. But equally there's no specific chance, high or low, of induction or counter-induction giving us either habit, since there's no specific chance of many frogs of such a kind being all or

nearly all green, or all or nearly all not green. In this lawless situation, neither induction nor counter-induction is warranted.

How do we know this isn't our situation? Maybe we don't know. But we don't need to know. Induction only fails here because here there's no success – no good habit of inference – to be had. When there is a good habit to be had, induction will always give us our best chance of getting it; and the better the inference, the better the chance. And that's enough to warrant induction as a general basic tendency.

But is it enough for induction to warrant the specific inference that this lecture is terse? Well, it is if lectures (or at least Inaugurals in English) have a high chance of being terse. For this will both warrant the habit of inferring that such lectures are terse and give induction a high chance of inducing this warranted habit, as it has done.

But do lectures like this have a high chance of being terse? Of course they do. You may not know this (though I dare say you do), but that doesn't matter. All that matters is that they do. And this fact (which, since you do at least believe it, you can't honestly deny) entails that your belief that this lecture is terse is both warranted, and warranted by induction. And so you do know already that it's terse – that it will last less than twenty-four hours – because, as you'll see at the end of this sentence, this belief of yours is not only warranted, it's true.

16 *Objective decision making*

1 Doctors often prescribe a drug for a pain on the grounds that taking the drug will relieve the pain, a reason I shall call 'objective'. But many modern decision theorists in effect reject this reason for taking the drug. They say rather that the patient should take it if he thinks it will relieve the pain (Jeffrey 1983 ch. 1, Eells 1982 ch. 1), or if it is rational for him to think so (Skyrms 1980a sect. IIC), reasons I shall call respectively 'subjective' and 'rationalist'. ('Rationalist' because calling the reason 'rational' would commend it, which I don't.) The burden of this paper is that the doctors are right and these decision theorists wrong. The patient's action needs an objective reason. The rationalist reason suffices only when derived from an objective one, and the subjective one suffices not at all.

My main target will be subjectivism, for two reasons. First, it is now almost orthodox, and its defects matter in both theory and practice. Taken subjectively, decision theory's technicalities give makers of public decisions an all too congenial and seemingly scientific rationale for declining to justify the hunches they act on. But secondly, subjectivism is at least a clear alternative to objectivism, which rationalism isn't. Appeals to rationality are mostly bluff. There is no good theory of what it is nor of how to recognise it. How, for instance, can we know it is rational to think that a pain-killing drug will work if we can't know that it *will* work? But if we can know that, why not act on that knowledge? Because it is harder to come by? And here the rationalist faces a dilemma. If we can be wrong about rationalist reasons, they are apt to become as objective and hard to know as objective ones; if we can't, because they are whatever we think they are, they become subjective. Rationalism in short occupies a no man's land, which I hope to clear out *en passant* in attacking subjectivism. But first we must draw up the lines of battle.

2 I start by conceding the descriptive ground to subjectivism. Of course someone in pain will only take a drug if he thinks it will work, whether it will or not. But the dispute about reasons is about what makes his taking it right or sensible, not about what makes him take it. I shall argue that he is right to take

rather than to decline it only if it will relieve the pain or at least make relief more probable, and sensible only if his action will probably be in this sense right. His faith in the drug does not in itself make his taking it either right or sensible; and his faith's being rational makes his action no better unless it makes the action more probably right.

I confine myself here to simple actions done at least partly to get something else done that is not directly 'do-able' – the intended 'outcome' of the action. The drug, for instance, is taken in order to relieve a pain that one cannot just will away, the relief being the intended outcome of taking the drug. Since I suppose one would not take the drug for (for example) its taste, the action here is done almost wholly for the sake of its outcome. But decision theory can also cope with actions done partly for their own sake, so I consider them as well. (It can also assess actions chosen from a range of alternatives, but I shall stick to one: the nature of the rationale for doing something does not depend on how many materially different ways there are of not doing it.)

I follow Jeffrey's (1983 ch. 4) account in terms of propositions: the propositions A that the action is done and O that the outcome occurs. Doing the action, which I shall call 'doing A', is making A true, the intended outcome being that O is true. Decision theory then assesses the doing of A in terms of how valuable in the circumstances the truth of $A\&O$, $A\&{\sim}O$, ${\sim}A\&O$ and ${\sim}A\&{\sim}O$ would be, the measures of these values being their so-called 'utilities' (Jeffrey's 'desirabilities').

Some decision-making principles, notably 'Maximise expected utility', also take account of how probable O's truth would in the circumstances be were A done and were it not done. Other principles – for example, 'dominance' and 'maximin' – do not, and it is often debatable which principle should apply (Nozick 1969, Levi 1980a appendix). I shall stay out of these debates, since all these principles pose the same question: should whatever utilities and probabilities they invoke be taken to be objective, subjective or rationalist? I will argue the objective answer for expected utility in order to be definite and to take in probability; but the argument for objective utilities will apply to other principles as well.

Decision theory doesn't of course cover all assessments of action, let alone of agents. It often matters whether A was made true by action or inaction: for example, whether I killed someone or let him die, a difference not captured in the utilities of A's truth and falsity. Nor does decision theory cover interactions for which we need game theory, nor the assessment of character, nor of people's motives for acting. But its utilities and probabilities do give reasons for acting; and if they are objective, so may such reasons be. Yet

writers on reasons for action (e.g. Bond 1980) and on decision theory rarely discuss each other, which is a pity. I hope that what follows may provoke more general discussion than space allows here.

But decision theory is more than a simple-minded model of reasons for action. In particular it can easily allow that many diverse features of an agent and his surroundings affect whether he should do *A*. It claims only that their effects are summed up in the utilities and probabilities *O*'s and ~*O*'s truth would have in the circumstances were *A* true and were it false; and so they often are. How such effects are to be summed up is for others to say, decision theory's job being to say how the sum determines the sense or rightness of making *A* true. It complements theories that say how valuable and probable things are in various circumstances just as Newton's laws of motion complement the sciences that say what Newtonian forces are. Both sum up diverse phenomena in simple measures (force and mass, utility and probability) for specific purposes – predicting accelerations, assessing actions – without at all denying their diversity in other respects.

O's and ~*O*'s utilities and probabilities thus depend only *inter alia* on *A*'s truth value: *how* they depend on it will itself depend on circumstances. So we need not suppose, implausibly, that the utility and probability of a drug's relieving a pain are the same in all cases, merely that often they exist and settle whether one should take the drug. That much is undeniable; but what then must these utilities and probabilities be?

3 Arguments against objective decision theory are chiefly arguments against objective utilities and probabilities; but there are others. Decision theory's reasons for doing *A* or ~*A* depend on how valuable *O*'s and ~*O*'s truth would be were *A* done and were ~*A* done. So whichever is done, and whatever the outcome, the reason for the action depends on more than actual utilities and probabilities. Of the four utilities appealed to, of the truth of *A&O*, *A&*~*O*, ~*A&O*, ~*A&*~*O*, only one will be realised; of the four probabilities, that *O* and ~*O* would have were *A* done and were it not done, only two. Objective decision theory therefore needs more than objective utilities and probabilities. It needs facts about what these would be were things to happen and be done that in fact do not happen and are not done. Such 'counterfactual' facts are contentious, quite apart from their being about utilities and probabilities, and that contention had better be settled first.

The contention is actually about subjunctive facts, not just about counterfactual ones. Suppose I drop a glass, and it breaks. Everyone will admit those two facts, but some would deny factual status to the further claim that the glass

was fragile (would have broken if dropped) both at the time and beforehand. Beforehand the subjunctive 'If the glass were now dropped it would break' was indeed counterfactual, but not when the glass was dropped: yet if the subjunctive was true before, it was certainly true then. The subjunctive I take to be the same whether or not the glass is currently being dropped and breaking, and in neither case does it follow from those admitted facts. I have argued elsewhere (see chapter 6) that it states a further fact, namely that the glass is fragile, which is a real fact because (*inter alia*) changes in fragility are real events with real causes and effects. The rest of that argument I will not repeat here, except to say that there will be very few facts if there are no subjunctive ones, since ascriptions of all properties, not merely of standard dispositions like fragility, entail just such subjunctives, and not much else. Stating a thing's inertial mass, for instance, says only which way and how fast it would accelerate were any given force applied to it; that in turn says only how different a velocity it would have were the acceleration to continue for any length of time; that in turn ... (and so forth). Yet if saying what things' masses and accelerations are isn't stating facts, what is?

Subjunctive facts are thus by no means peculiar to objective decision theory, and their problems are not its problems. If there can be objective utilities and probabilities, there can be facts about what they would be were certain actions done, in whatever sense there can be facts about masses and accelerations. For probabilities, the facts will mostly be that things have what I call 'propensities' (Mellor 1971 chs 4–5): for example, the bias of a coin, which is how unequal the probabilities of heads and tails would be were it to be fairly tossed – a property it has even when it has no chance of being tossed and *a fortiori* equal because zero actual probabilities of landing heads and tails. Indeed deterministic properties like fragility and mass are really only extreme cases of propensities, in which the consequent probabilities are 1 or 0. Massive objects don't just happen to accelerate as they do: given any net applied force, the probability of the corresponding acceleration will be 1.

Whatever general objections there may be to propensities, subjectivists and rationalists cannot object to them, for they need them too. My subjective probabilities, which measure how strongly I believe propositions, are defined by how I would (most probably) behave if their truth mattered to me: for example, by the odds I would accept if made to bet on their truth. But these are behavioural propensities of mine, subjunctive facts about me: so if there are no such facts, there are no subjective probabilities. And the same goes for 'rational' or 'epistemic' probabilities, which are what my subjective probabilities would be if only I were rational and heeded the available evidence.

Subjectivists have tried to replace propensities with conditional probabilities, but without success. Certainly objective conditional probabilities will not do. Suppose for instance our coin were so biased that were it to be fairly tossed the probability of heads would be 0.8. The probability of heads conditional on a fair toss is by definition the probability of the coin being fairly tossed and landing heads divided by the probability of it being fairly tossed; and we may grant that this, if it exists, will be 0.8. But it may not exist. For as we have seen, the probability of the coin being tossed may be zero, in which case this conditional probability will be undefined. Or it may be undefined because there is no objective probability of the coin being tossed. Yet the coin will still have its bias. And it is no use defining it as what the conditional probability would be if only the coin had a non-zero probability of being fairly tossed; for that is itself a propensity, whose value clearly derives from the original fact that were the coin fairly tossed the probability of heads would be 0.8.

Subjectivists of course dispense with objective probability: subjective conditional probability is meant to replace not propensity itself but belief in it, construed as a subjunctive fact about, for example, the unconditional probability (I believe) I would attach to a coin landing heads if I believed it was being fairly tossed. But this doesn't work either. For I might believe that the coin is biased while attaching zero probability to its being tossed, so that my subjective probability of its landing heads conditional on its being tossed is undefined. Or it might be undefined because I attach neither zero nor any other probability to the coin being tossed: for example, because I am trying to decide whether to toss it and can attach no definite probability in advance to that decision. And again it is no use invoking the subjective conditional probability I would have if only I did attach a definite non-zero probability to the coin being fairly tossed. That is as much a subjunctive fact about me as the one we are trying to replace, and indeed, as in the objective case, clearly derives from it.

The question about subjunctive facts therefore is not whether there are any, only which there are. In particular, do we need subjunctive facts about objective utilities and probabilities to make sense of decision theory; and, if so, how can we have and know them? But the problems here concern the utilities and probabilities, not the subjunctives they occur in. From now on, therefore, I shall take subjunctives for granted, and ask only whether we need, and how we can know, objective utilities and probabilities.

4 It is usually right to take a pain-killing drug because pain is usually undesirable: life without pain is usually better than life with it. Perhaps not always: pain may sometimes be necessary to a greater good, and then it would not be right to take the drug. Decision theory accommodates all this by taking pain to lower the utility of most situations, which is why it mostly prescribes the drug; while allowing the relief of a pain sometimes to lower the utility of a situation, in which therefore it would not be prescribed. Which situations are which is for others to say, since, as we saw in **2**, this dependence on outside data no more vitiates decision theory than it vitiates Newtonian mechanics.

So much I can take for granted, since it leaves the status of utilities open. I say they are objective, but not in the sense of there being facts about them, or about probabilities, knowledge of which demands a true belief (see chapter 14). Probability in decision theory relates to the strength of a belief, as utility does to the strength of a desire. What objectivity demands here is not truth but the objective rightness of some degree of desire or belief, a degree whose measure will be some restricted interval of utility of probability values (see Levi 1980a ch. 9).[1]

The point is best seen in the failure of the frequency theory of probability. Suppose a coin toss has a fifty-fifty chance of landing heads. This I claim is therefore the objectively right degree of the belief that the coin will land heads, even when no one has this belief to this degree, and even when (for example) a freak run of heads or tails makes quite a different degree of this belief more rational. Now the rightness of this degree of this belief doesn't reduce to the truth of any belief. The only *prima facie* candidate, true belief in some frequency of heads on many real or imaginary tosses of the coin, will not do. For the laws of large numbers relate the probability of heads on this toss to no more than a high probability of a similar frequency of heads on many such tosses: the rightness of this degree of belief in heads this time doesn't correspond to the truth of any belief in such a frequency lying in any seriously restricted interval.

But even if the rightness of a degree of one belief did correspond to the truth of another one, the correspondence would not follow from the truth. This is the real objection to any theory that makes objective probability a 'truth-maker' for beliefs of some special probabilistic kind. They cannot explain why my having such a belief should give me the corresponding degree

[1]To simplify what follows, I shall credit both subjective and objective utilities and probabilities with having precise values. They clearly don't, any more than felt or measured temperatures do, but the assumption is innocuous here, as it is in thermometry.

of any non-probabilistic belief, for example in this toss landing heads, and thus make me act accordingly, for example by choosing odds for a bet on that outcome. And the same goes for similar theories of value (Bond 1980 chs 4–5). They cannot explain why my having a belief in an objective value should make me desire it and so act to achieve it.

I can accept all these objections to objectivism, because they pass me by. My utilities and probabilities are indeed meant to make degrees of desire and belief objectively right, but not *via* making beliefs true. Why should they? Why must the rightness of mental states always follow from the truth of a belief? Not because facts by definition correspond to true beliefs and only facts can make mental states objectively right: the correspondence theory of truth is in no condition to corner the market in objectivity. Indeed, as a pragmatist, 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. But however that may be, no account of truth can give it a monopoly of objectivity. For we only accept objective truth because we think certain full beliefs objectively right, and that can hardly prevent us from thinking the same thing about other degrees of belief and degrees of desire.

We may of course believe a degree of desire or belief to be right; and this belief may be true (and its truth may be more or less valuable or probable). But the rightness of the degree of that desire or belief is what makes this belief true, not the other way round. So far I am with the rationalist, who will also deny 'truth-making' facts about utilities and probabilities while jibbing at subjectivism. But he will read rightness here as rationality: that is, as justification by other desires and beliefs one is similarly justified in having to some degree. I will not. For that concept of justification faces familiar and formidable problems even in accounting for knowledge of truth-making facts, where there is a concept of objective rightness (truth) for justification to aim at. And here justification is supposed to supplant, not supplement, the concept of objective rightness in degrees of desire and belief, which makes it even harder to make sense of.

No such problems face the alternative conception of knowledge as true belief got by reliable means (Ramsey 1929b). That escapes stock objections to knowledge conceived as justified true belief (Grandy 1980) and extends immediately to knowledge of objective utilities and probabilities. To be knowledge, our degrees of desire and belief must be right and be got by reliable means: that is, in a way that makes their acquisition more probable if they are right than if they aren't. But that means letting whatever makes them

right be an objective factor affecting an objective probability of getting them, which only objectivism can admit. So although knowing good prospects when you see them isn't acquiring justified true beliefs about their value, but regularly and reliably becoming disposed to go to as much trouble to realise them as they're worth, the prospects still have to be objective.

So objective utilities and probabilities will not idle in accounts of our knowledge of them, as Platonic numbers, for example, do in accounts of mathematics. Indeed we need them to account for knowledge even of truth-making facts. Take perceptual knowledge, such as my seeing that something is red. I know it's red because it is red, I believe it's red and I got that belief by reliable means, namely by looking at it with good eyes in a good light. Following Dummett (1964), I have argued elsewhere that causation in general, and therefore the causal processes of perception in particular, themselves depend on the rightness of a conceivable action, namely bringing about a cause to bring about one of its effects (see chapter 13). The merit of a belief acquired by perception thus rests on the utility of its truth, which entails (*via* the expected utility principle) that its acquisition should be more probable if it is true than if it is false, that is, that perception must be a more or less reliable means of acquiring true beliefs. The objectivity of causation itself, and thus of perception and of the knowledge it delivers, depends on the objectivity of utilities and probabilities.

5 A utility is the utility of a proposition's truth, for example, of the proposition that I take a drug and my pain is relieved (see Jeffrey 1983 p. 60), just as probability is probability of truth; and almost any empirical truth may be in the circumstances more or less valuable, more or less probable. Utilities and probabilities may thus attach to almost any of the prospective truths that they at once require and transcend – including truths about utilities and probabilities, despite these reducing in turn to the rightness of degrees of desire and belief. The case for the objectivity of utilities and probabilities is the case for there sometimes being objective reasons for bringing such prospective truths about. What I have to show is that there are such reasons. But to do that I need no exhaustive list or theory of what's valuable and what's probable, any more than a believer in truth-making facts has to list or account for all those. Nor need I prove the mere possibility of objective reasons for making propositions true, that being as obvious as the possibility of truth itself. The question is whether in fact there are such reasons.

My case for objective reasons will be that we need objective utilities and probabilities to make prescriptive sense of decision-making principles such as

expected utility. The case against them is that we don't. To rebut it I need show only that we need them in a typical instance, for example to explain the rightness and sense of taking a drug to relieve a pain. But I will show that by showing how little sense subjective and rationalist utilities and probabilities make of the decision-making principles involved; and that in turn will recommend an objective reading of other instances. Only I shall not here apply the argument to other instances, nor try to prove that right action always needs objective reasons. Here I mean only to put the burden of proof, for too long put upon objectivists, back on their opponents, where it belongs.

What sense can subjectivists and rationalists make of our pain-killing example? I will grant them at once their subjective utilities and probabilities. These are just the degrees of desire and belief I used above to characterise knowledge of their objective counterparts. To know an objective utility or probability is usually to have got by reliable means an objectively right subjective utility or probability. What I deny in subjectivism is not the existence of these mental states but their adequacy to explain the rightness and sense of the actions they cause. And as the idea that they are adequate has several sources (their number being their most notable feature), I will try to dam these sources *seriatim.*

The first source is an equivocation between theoretical and prescriptive senses of 'ideal'. The fact is that subjectivist psychology, which says how degrees of desire and belief determine actions, at best approximates the truth. People don't, for instance, always maximise their subjective expected utilities: that is, sometimes no assignment of subjective utilities and probabilities will make that assumption fit what they do. Still, decision theory's explanations of action are often usefully simple and accurate enough, fitting 'ideal' people much as the gas laws fit 'ideal' gases. But 'ideal' here doesn't mean that people who falsify the theory are wrong or irrational, any more than non-ideal gases are. Rationality would only require ideal behaviour in this sense if it required one to maximise a merely subjective expected utility, and that is the question at issue. The theory's being an idealisation in the sense of being often false is no argument for taking it as a prescription.

But some arguments for the theory's measures of desire and belief do appeal to rationality in action, which is then imputed to the measures thus derived. Take the so-called 'Dutch book' argument for a probability measure of belief. This measures my degrees of belief by making me choose betting odds, and hence betting quotients, in circumstances designed to let only my degrees of belief affect my choice (for example, I am forced to bet, my opponent choosing the bets' directions and stakes after I choose the odds: see

Mellor 1971 ch. 2); and unless my betting quotients are probabilities I could then be forced into a combined bet that I would lose whatever happened, which I should clearly be irrational to allow. So whenever my betting quotients and thus my degrees of belief fail to be probabilities, as critics (e.g. Kyburg 1978) claim they often do, I am supposed to be irrational. Thus the theory admits its falsity, but blames it on the subjective irrationality of betting when one can only lose.

This is the usual reading of the Dutch book argument, and it is wrong on both counts (see chapter 3, part II). First, there is nothing subjective about the irrationality of exposing oneself to a Dutch book, even when it is irrational, which it usually isn't. A Dutch book is after all only a method of giving money or other goods away, which I trust no one will say is usually wrong or irrational. (To say it is irrational if I don't want to give money away begs the question, since it is not self-evidently wrong or irrational to give money reluctantly to a good cause.) And why should a reluctant philanthropist with a taste for gambling not do good by betting at incoherent quotients: why else buy tickets for church lotteries? But even when it is wrong or irrational of me to choose incoherent betting quotients, that doesn't show my degrees of belief to be irrational because not probabilities: for insofar as my quotients are incoherent, they simply fail to measure my degrees of belief. The fact is that the Dutch book argument doesn't show it to be irrational of us not to fit subjective probability theory, and its critics haven't shown that we don't fit it.

All the Dutch book and other arguments do is recommend a probability measure for degrees of belief, and a measure for degrees of desire that can combine with it in the calculation of subjective expected utilities. The theory that we maximise these may or may not be true, but degrees of desire and belief measured by it can no more fail to fit its measures of them than temperatures measured on the Celsius scale can fail to fit that. The measures might indeed be called ideal measures prescribed by the theory, but not in the sense of ideals that we should try (but might fail) to live up to.

6 Subjectivism's next ploy, after palming off its descriptive defects as prescriptive virtues, is to appeal to a division of labour. Let others say what subjective utilities and probabilities we should have, subjective decision theory will tell us what it's rational to do with those we do have (see e.g. Eells 1982 p. 5). That sounds fine, but its appeal in fact derives from the division of objective labour noted in **2**, in which objective utilities and probabilities warranted by others are used to assess the rightness and sense of actions. When my subjective utilities and probabilities are objectively right, then certainly I

should act on them as subjective decision theory says, since my action will then be objectively right. But the rightness of my action then derives directly from the objective utilities and probabilities. I needn't first have subjective probabilities and utilities and then get them right, since I needn't have them at all: I might be one of those whose psychology subjective decision theory fails to fit. That wouldn't prevent me from rightly taking a drug to relieve pain, nor from doing so sensibly by acting on my doctor's reliable judgment of when a drug would relieve a pain that ought to be relieved.

Granted, I define objective utilities and probabilities as objectively right subjective utilities and probabilities. But that is a conditional definition: what my subjective utilities and probabilities must be for my having them to amount to my knowing their objective counterparts. No doubt this is a common way of knowing objective utilities and probabilities, a way exploited in measuring them by measures of the corresponding degrees of desire and belief. But the definition doesn't make this the only way of knowing objective utilities and probabilities. Consider the analogous case of objective colours, which we might also characterise by the mental states by which people commonly know them (see chapter 14). That is, we might distinguish different colours by the different visual sensations normally sighted people have when they see them. But that still allows people to be wrong about objective colours, and to know them in other ways. In a sense, no doubt, a blind person doesn't know what colours are, just as someone who never doubts doesn't know what probabilities are. But a red–green colour-blind person may still come to know by other reliable means that something is red. And a person who lacks subjective utilities and probabilities as precise as their objective counterparts may likewise develop other reliable means of acting rightly: that is, as a person would act who did have them and got them right.

To act rightly or sensibly according to expected utility, one need not have subjective utilities and probabilities. Nor does having them suffice. Suppose I do have them, only they're so wrong that I act wrongly when I maximise my subjective expected utility. Is my action really at all to be commended for conforming to that principle? Of course the subjective theory is descriptively 'as applicable to the deliberations of a monster as it is to that of a saint' (Eells 1982 p. 5), but that doesn't mean we should at all commend monsters for being at least subjectively rational. But calling behaviour 'rational' without commending it just abuses the term, and concedes the subjective theory's prescriptive bankruptcy.

But might one not still commend a general policy of maximising subjective expected utility for producing right action when applied to right subjective

utilities and probabilities? After all, the policy can hardly be blamed for sometimes producing wrong action when applied to wrong ones. One might similarly commend a general policy of making deductions from one's beliefs, because deduction preserves truth and can hardly be blamed for sometimes producing false conclusions from false premises. As David Lewis (to whom I owe the analogy) put it: 'Garbage in, garbage out – so what?' But this defence of subjective expected utility concedes the existence of objective utilities and probabilities, just as this defence of subjective deduction concedes objective truth. Why follow a subjective policy because it generates right outputs from right inputs if there are no right inputs?

And if there are right inputs, why follow subjective policies at all? We have seen that right action needn't be derived from subjective utilities and probabilities; and that when it is, their objective counterparts, not the subjective derivations, are what make the action right. Similarly for subjective deduction: a true belief that is deducible from others needn't in fact be so deduced; and when it is, what matters is that the propositions it is deduced from are true, not that they are believed. And while false beliefs may well be produced by deduction from other false beliefs, just as wrong action may be produced by maximising wrong subjective expected utilities, that pedigree no more commends a false belief than it commends a wrong action. The right prescription in each case is surely objective: deduce your beliefs from truths; act to maximise objective expected utility.

7 But if we never know what is true, only what we believe, is it not idle to prescribe deduction from the truth? Surely the best we can do is deduce new beliefs from existing ones, since then at least the new beliefs will be true if the existing ones are. Wholesale subjective deduction seems to be the only sensible and usable prescription. And I think the same idea is the main source of subjectivism in decision theory. 'Maximise objective expected utility' is thought to be an idle prescription because we no more know what is really valuable or probable than we know what is really true. We only know what we desire and what we believe, and how strongly we desire and believe it. So the only usable and sensible prescription seems to be to maximise subjective expected utility, which will at least make us act rightly when our subjective utilities and probabilities happen to be right.

Some of the confusions in this idea have been noted already. Of course people in fact act only on their subjective utilities and probabilities, right or wrong, just as they make deductions only from what they believe, true or false. The question however is what they should do, not what they actually do.

And to take decision theory prescriptively is not to prescribe taking it as an ingredient in actual decision making. The most ardent advocate of Jeffrey's (1983) classic book on decision theory will hardly prescribe consulting it instead of a doctor when trying to decide what to do about pain. That is not how decision theory is meant to be applied, and it would be no argument for subjectivism that, so applied, it could be applied only to subjective utilities and probabilities.

In any case, that isn't true. The contents and strengths of our desires and beliefs are defined by subjectivists (and I agree) by how they would make us act in specified situations; and our consciousness of that, and therefore of them, though generally reliable, is by no means infallible. We keep few of our desires and beliefs consciously in mind at any time, and about those we do have in mind we may always be self-deceived: that is, our beliefs about what they are may be false (see chapter 3). As subjectivists understand belief, therefore, we can easily come to think that a belief we have is wrong or unfounded without thereby ceasing to have it; and everyone will admit this possibility for desires. So even as we act we may think we are acting wrongly or foolishly because we think the subjective utilities and probabilities we are acting on are wrong or ill-advised; and this is no mere theoretical possibility but a depressingly familiar experience.

What makes the experience depressing, of course, is that recognising error in our desires and beliefs rarely suffices to correct them. Undesirable desires and beliefs can no more be willed away than pains can; and a wrong action may indeed be excused by the fact that the agent could not help having the desires and beliefs that made him do it. And this I think is yet another source of subjectivism: if we cannot help our subjective utilities and probabilities, actions that are at least correctly derived from them seem thereby to be sufficiently excused, and there seems no point in criticising them further for error or folly. But to excuse an action is not to commend it: on the contrary, it is to admit that the action is wrong and therefore needs excusing. These excuses rather require than replace an objective assessment of what they excuse. And decision theory itself shows us how to alter indirectly what we cannot alter at will. Maybe I cannot will away a desire or a belief (in a quack nostrum, for example) that I recognise to be wrong or foolish; but I can go to a doctor to cure my addiction, and decide to acquire such desires and beliefs in future by more reliable means. Seeing that what I am doing is wrong may not stop me doing it now, but it may well make me stop myself doing it again.

Such assessments of our own desires, beliefs and actions are of course themselves caused by what we actually think of them, not by what we ought to

think. But no one I trust will infer from this that the content of these assessments is subjective, or that they could not themselves be wrong or unfounded. Even if we emulate the redundancy theory of truth and decline to say what it is for degrees of desire and belief to be right, only what it is to judge them to be right, that will still help neither subjectivists nor rationalists. For suppose that judging a belief (that p) to be true is believing myself to have that very belief (or – if p is indexical – the belief got from it by allowing for relevant differences in the believer's circumstances). This still means that I believe the belief (that p) to be true because (I believe) it corresponds to the fact that p, not because anyone believes that p or has reason to believe it. The redundancy theory of truth explains rather than dispenses with a correspondence conception of objective truth. It does not claim that believing one has the belief that p is all there is to p's truth, only that it is all there is to thinking p true; and that this, given an independent account of belief, suffices for saying what truth is. Similarly, if my judging a degree of belief that p to be right is believing myself to have that degree of that belief (allowing for context if p is indexical). Also, if judging a degree of desire that p to be right is desiring to get (or to go on having) that degree of that desire, again allowing for context if need be. I don't say these are sufficient accounts of the rightness of degrees of desire and belief; merely that, even if they are, it is objectivism they support, not subjectivism or rationalism.

So we are not confined to applying decision theory subjectively even to decisions we are presently making. And we are obviously not so confined when using it to assess the rightness and sense of our past actions and the actions of others. The truism that we act only on what we actually desire and believe no more excuses subjectivism in decision theory than it excuses decision-theoretic epistemologies (like that of Levi 1980a) which obliterate all distinctions between belief, true belief and knowledge.

8 Decision theory is for assessing decisions, not for making them. But even in making them, we not only need not prescribe it subjectively or rationalistically, we often could not. Consider the subjective prescription: 'Maximise your subjective expected utility'. To follow it we must know what our subjective utilities and probabilities are. And far from these being all we know when we act, we often know nothing about them at all. Knowledge here being true belief got by reliable means, we only know these states of mind when we have beliefs about them. But we only have beliefs about our own subjective probabilities when we have them consciously in mind; and the same goes for subjective utilities. But we often act without conscious deliberation or

awareness of what makes us act. Since then we know nothing of the desires and beliefs we are acting on, we cannot possibly be applying the subjective prescription. Nor can we be applying the rationalist one, which tells us to check first that our desires and beliefs are rational. Yet we might still be applying the objective prescription. For the subjective utilities and probabilities we act on might be right, and reliably acquired, and so constitute knowledge of the relevant objective utilities and probabilities (see **4** above).

But how then do we know we are acting rightly: that is, know the subjective utilities and probabilities we are acting on to be right (since we don't even know what they are)? The answer is that in such cases we don't know that they are right, and we don't need to. We know the objective utilities and probabilities, and know therefore how to act (that is, how to act rightly), and that is good enough. Here, as for knowledge constituted by true belief got by reliable means, knowing something can't entail knowing you know it (see chapter 15). That entailment, natural as it seems, only generates the notorious regresses of knowledge conceived as true belief justified by justified beliefs, and it must share the fate of that discredited conception. And it is a clear corollary of **4** that knowing the rightness of one's action is knowing how to act rightly, not knowing that one's action is right in the sense of having even a reliably acquired true belief on the matter. I know all I need to know about right action when I know how to act rightly, whether or not I know I have that knowledge.

But perhaps prescriptions can only be applied consciously. If so, then whenever I could apply the objective prescription, I could apply the subjective or rationalist one instead. But should I? Suppose I do act with conscious deliberation, thus having not only the desires and beliefs I act on but also beliefs about what they are. Should I really be told to maximise my subjective expected utility instead of its objective counterpart? Should I, that is, be told not to worry about the rightness or reliability of my subjective utilities and probabilities, but just to act on them regardless: to take whatever drug I feel like taking, instead of first adopting my doctor's more reliable subjective utilities and probabilities and acting on those? Obviously not.

But, the rationalist will say, even a reliable doctor may be wrong. The drug he prescribes may not relieve my pain or even make relief more probable. Yet I should still take the doctor's advice, because it gives me reason to think the drug effective. So it does, but only because the doctor is at least objectively more reliable than I am: that is, his subjective utilities and probabilities are more probably right than mine are. But suppose I don't know whether he is reliable: should I not consult him if I think he is? Only if I know

at least that he's a doctor, that his training makes him probably more reliable than I am. Being sensible here depends at every stage on increasing an objective probability: if not of relief then of relief's being probable, or at least of its probably being probable, or ... And this regress is no use to rationalists: it at best postpones the dilemma alluded to in **1**. Eventually my reason for acting must reduce either to my actually desiring or believing something (thus becoming subjective), or to objective utilities and probabilities I needn't know anything of. And once subjectivism is rejected, the original utilities and probabilities – of actually relieving the pain – clearly provide my best reason for action. They are what I should act on if I know them, and all the others (for example, those that define the doctor's reliability) can only be defined and discovered by means of them. It can only be sensible to consult a doctor because he's reliable if it's probably right to take the drug he prescribes because that is reliable.

In short, although there is really no point in applying decision theory at all while actually making decisions, if you must apply it, do so objectively if you have the choice. If you have no choice, it will probably not be because you can't apply the theory objectively, but because you can't apply it any other way. We needn't prescribe subjective or rationalist decision theory to agents *faute de mieux*: if anything, we can prescribe the objective theory to them *faute de pis.*

9 My last source of subjectivism is scepticism, specifically about objective utilities. Modern physics, if nothing else, makes objective physical or epistemic probabilities undeniable. There is more resistance to objective utilities, and hence to accounts of decision theory that invoke them. But neither subjectivists nor rationalists can deny objective utilities in the sense of **4**. For to prescribe maximising subjective or rational expected utility is to claim greater objective utility for actions that do maximise it than for actions that don't. And this claim is often true. That people actually want something often makes it better that they get it, whether or not they have any (other) reason to want it. (Not always, naturally: people sometimes want to inflict gratuitous suffering.)

But even when I should get what I want just because I want it, that doesn't support the subjective prescription. For a start, I may not know what I want. I noted in **7** that behaviourally defined subjective utilities and probabilities are not infallibly introspectible. We get our beliefs about what we desire and believe by an 'inner sense' (see chapter 3) that is no more reliable than our outer senses. I can easily believe I desire something when I don't, or when I

desire it more or less than I think I do. So if I should act on my desire, I should only act on what I believe I desire if that belief is true (as indeed it usually is, especially when the desire is to be rid of a pain). But then even the prescription to maximise my subjective expected utility is objective. It only seems subjective because I know what I desire and my having the desire happens in this case to make it objectively right.

But it concedes too much even here to say that I should get what I want. I am not such an authority on the utility of relieving my pain as subjectivists suppose. I may be the authority on my present pain and how bad that is. But I am taking the drug to relieve my future pain, and I can hardly tell now by introspection how good that relief will be, still less how bad the pain would have been had it not been relieved. Of the four utilities – of the truth of $A\&O$, $A\&\sim O$, $\sim A\&O$ and $\sim A\&\sim O$ – which decision theory needs to assess my action, I will only ever experience one, and that only after I have acted. So even when the utilities depend only on how I will or would feel in those circumstances, even I can only make objectively falsifiable predictions of them. And although I may be the most reliable predictor of these possible future feelings, I need not be. Obvious examples are drugs, injections, and local anaesthetics I have never had before. I may fear them mightily, and take them only because my doctor assures me that they won't be as bad as I expect. His experience makes him a more reliable judge of these utilities than I am, even though my subsequent experience is what will or would settle the truth of his prediction and thus the rightness of my action. Clearly I should take his advice, whether I think I should or not. Not even in this case, the most favourable to subjectivism, should I be commended for an action that needlessly perpetuated my pain, just because I thought it was a good idea at the time.

But doesn't this case favour rationalism rather than objectivism? There are several theses about time, modality and induction that appear to deny us objective knowledge of future and merely possible utilities, and allow us at best to predict them more or less reasonably. One such thesis is that future tense statements lack truth value: if so, there are no objective facts about my future pain to be known. But this thesis I have disputed at length elsewhere (see chapter 11 and Mellor 1981), and here I will just assume that predictions can be true or false. And then the case for subjunctive facts made in 3 will apply equally to future facts, for example, about how bad my pain would have been tomorrow had it not been relieved. Objections to utilities apart, that counterfactual is no more problematic than counterfactuals saying how fast something would have accelerated tomorrow under various forces. What makes them true is the mass the thing has the day after they are uttered; what

makes the counterfactual about my pain true is some metabolic property I have on that day.

But even granted future subjunctive facts about utilities, can they be known in more than a rationalist sense? The causal processes of perception are the objectively reliable means that make the true beliefs it gives us knowledge. But there being no backward causation, there is no perceiving future facts. Facts about how bad my pain will be or would have been tomorrow cannot reliably generate true predictions about them in me today. Such predictions can be backed up only by perceptions of past and present facts. So they seem only to be justified in the traditional rationalist sense by otherwise justified beliefs; and Hume's critique of induction makes even that justification problematic.

Hume indeed shows the rationalist's conception of knowledge of the future to be as hopeless as his conception of other knowledge is. But that does not prevent us knowing the future in the sense of reliably acquiring true beliefs about it. Of course we cannot now perceive the future: if we could we could not affect it, and all prescriptions for action would be pointless (see Mellor 1987a). But perception is not the only reliable means of acquiring beliefs, if only because causation doesn't just make effects reliably succeed their causes. It also correlates effects with side-effects which we can then use to predict the effects. Thus for instance we use a falling barometer to predict the ensuing rain of which it is a side-effect. If the objective epistemic probability of rain is sufficiently greater than it would have been had the barometer not fallen, the reliability of the prediction makes it, if true, knowledge. Similarly with the pain-killing drug. The data the doctor relies on in prescribing it – e.g. the entry in the pharmacopoeia – are made to be side-effects of its pain-killing power. They give the desired relief an objectively greater epistemic probability that makes it objectively sensible to take the drug. That they do so is of course a matter of fact: there is no question of proving it, nor therefore of rebutting Hume's proof that it cannot be proved. Nor need we know the pharmacopoeia to be reliable for it to be so and thus to make sense of prescribing from it (though doctors generally do know this because they know how drugs are tested). Here, as in **8**, knowing something doesn't entail knowing one knows it.

But then we could rely on the drug itself instead of the doctor or the pharmacopoeia. Their knowledge is only a side-effect of its efficacy, and I need only consult them when I don't myself know if it works. And then I should consult the doctor because he does know. By taking his advice I reliably acquire right subjective utilities and probabilities that will make me act rightly in taking the drug. So when he tells me I should take the drug

because it will (probably) work, he means exactly what he says. *Pace* our decision theorists, he does not (and should not) mean that I should take it because either he or I believes, or has reason to believe, it will (probably) work.

References

Ackrill, J. L., transl. (1963) *Aristotle: De Interpretatione*, Oxford: Oxford University Press.

Altham, J. E. J. (1973) 'The causal theory of names II', *Aristotelian Society Supplementary Volume* **47**, 209–25.

Anscombe, G. E. M. (1975) 'The first person', *Mind and Language*, ed. S. Guttenplan, Oxford: Clarendon Press, 45–65.

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

(1973) *Belief, Truth and Knowledge*, Cambridge: Cambridge University Press.

(1978) *Universals and Scientific Realism*, 2 volumes, Cambridge: Cambridge University Press.

(1983) *What is a Law of Nature?*, Cambridge: Cambridge University Press.

(1989) *Universals: An Opinionated Introduction*, Boulder, Colorado: Westview Press.

Austin, J. L. (1939) 'Are there *a priori* concepts?', *Philosophical Papers*, 1970, Oxford: Oxford University Press, 32–54.

Ayer, A. J. (1956) 'What is a law of nature?', *The Concept of a Person*, 1963, London: Macmillan, 209–34.

(1972) *Probability and Evidence*, London: Macmillan.

Berkeley, G. (1710) *The Principles of Human Knowledge*, ed. G. J. Warnock, 1962, Glasgow: William Collins.

Blackburn, S. (1969) 'Goodman's paradox', *Studies in Philosophy of Science*, ed. N. Rescher, Oxford: Blackwell, 128–42.

(1980) 'Opinions and chances', *Prospects for Pragmatism*, ed. D. H. Mellor, Cambridge: Cambridge University Press, 175–96.

Block, N., ed. (1980) *Readings in the Philosophy of Psychology Volume 1*, London: Methuen.

Boden, M. A. (1984) 'What is computational psychology? I', *Aristotelian Society Supplementary Volume* **58**, 17–35.

Bond, E. J. (1980) *Reason and Value*, Cambridge: Cambridge University Press.

Bradley, F. H. (1897) *Appearance and Reality*, 2nd edn, Oxford: Clarendon Press.

Braithwaite, R. B. (1926) 'Universals and the "method of analysis"', *Aristotelian Society Supplementary Volume* **6**, 27–38.

(1933) 'The nature of believing', *Knowledge and Belief*, ed. A. P. Griffiths, 1967, Oxford: Oxford University Press, 28–40.

(1953) *Scientific Explanation*, Cambridge: Cambridge University Press.

Broad, C. D. (1923) *Scientific Thought*, London: Kegan Paul.

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

Brown, R. (1970) 'Comment [on J. O. Wisdom's "Situational individualism and the emergent group properties"]', *Explanation in the Behavioural Sciences*, ed. R. Borger and F. Cioffi, Cambridge: Cambridge University Press, 297–305.

Burge, T. (1986) 'Individualism and psychology', *Philosophical Review* **95**, 3–46.

Cargile, J. (1969) 'The Sorites paradox', *British Journal for the Philosophy of Science* **19**, 193–202.

Carnap, R. (1936–7) 'Testability and meaning', *Readings in the Philosophy of Science*, ed. H. Feigl and M. Brodbeck, 1953, New York: Appleton-Century-Crofts, 47–92.

(1966) *Philosophical Foundations of Physics*, New York: Basic Books.

Carroll, L. (1887) *Through the Looking-Glass*, 1962, London: Folio Society.

Cartwright, N. (1979) 'Causal laws and effective strategies', *Noûs* **13**, 419–37.

(1983) *How the Laws of Physics Lie*, Oxford: Clarendon Press.

Castañeda, H. N. (1984) 'Causes, causality and energy', *Midwest Studies in Philosophy* **9**, 17–27.

Chomsky, N. (1969) 'Quine's empirical assumptions', *Words and Objections*, ed. D. Davidson and J. Hintikka, Dordrecht: Reidel, 53–68.

Churchland, P. M. (1988) *Matter and Consciousness*, revised edn, Cambridge, Mass.: MIT Press.

Davidson, D. (1967a) 'Causal relations', *Essays on Actions and Events*, 1980, Oxford: Clarendon Press, 149–62.

(1967b) 'The logical form of action sentences', *Essays on Actions and Events*, 1980, Oxford: Clarendon Press, 105–22.

(1969) 'The individuation of events', *Essays in Honor of Carl G. Hempel*, ed. N. Rescher, Dordrecht: Reidel, 216–34.

(1970a) 'Events as particulars', *Essays on Actions and Events*, 1980, Oxford: Clarendon Press, 181–7.

(1970b) 'Mental events', *Essays on Actions and Events*, 1980, Oxford: Clarendon Press, 207–25.

(1985a) 'Replies to essays', *Essays on Davidson: Actions and Events*, ed. B. Vermazen and M. Hintikka, Oxford: Clarendon Press, 195–254.

(1985b) 'Reply to Quine on events', *Actions and Events*, ed. E. LePore and B. P. McLaughlin, Oxford: Blackwell, 172–6.

de Finetti, B. (1931) 'Probabilism', *Erkenntnis* **31**, 1989, English transl., 169–223.

Delaney, P. C. (1962) 'Radioactive disintegration theory', *Encyclopaedic Dictionary of Physics Vol. 6*, ed. J. Thewlis, Oxford: Oxford University Press, 45–8.

Denbigh, K. G. (1955) *The Principles of Chemical Equilibrium*, Cambridge: Cambridge University Press.

Dennett, D. C. (1978) 'Where am I?', *Brainstorms*, Montgomery, Vermont: Bradford Books, 310–23.

Diamond, C. and Teichman, J., ed. (1979) *Intention and Intentionality*, London: Harvester.

Dretske, F. I. (1977) 'Laws of nature', *Philosophy of Science* **44**, 246–68.

(1981) *Knowledge and the Flow of Information*, Oxford: Blackwell.

Dummett, M. A. E. (1964) 'Bringing about the past', *Truth and Other Enigmas*, 1978, London: Duckworth, 333–50.

(1973) *Frege: Philosophy of Language*, London: Duckworth.

Dummett, M. A. E. (1974) 'Postscript', *Synthese* **27**, 523–34.

Eddington, A. S. (1929) *The Nature of the Physical World*, Cambridge: Cambridge University Press.

Eells, E. (1982) *Rational Decision and Causality*, Cambridge: Cambridge University Press.

Enç, B. (1982) 'Intentional states of mechanical devices', *Mind* **91**, 161–82.

Evans, G. (1973) 'The causal theory of names I', *Aristotelian Society Supplementary Volume* **47**, 187–208.

(1982) *The Varieties of Reference*, Oxford: Clarendon Press.

Feigl, H. (1963) 'Physicalism, unity of science and the foundations of psychology', *Inquiries and Provocations*, 1981, Dordrecht: Reidel, 302–41.

Fetzer, J. H. (1971) 'Dispositional probabilities', *Dispositions*, ed. R. Tuomela, 1978, Dordrecht: Reidel, 267–76.

(1981) *Scientific Knowledge: Causation, Explanation and Corroboration*, Dordrecht: Reidel.

Feyerabend, P. K. (1962) 'Explanation, reduction and empiricism', *Minnesota Studies in the Philosophy of Science Volume 3*, ed. H. Feigl and G. Maxwell, Minneapolis: University of Minnesota Press, 28–97.

(1970) 'Consolations for the specialist', *Criticism and the Growth of Knowledge*, ed. I. Lakatos and A. Musgrave, Cambridge: Cambridge University Press, 197–230.

Field, H. H. (1977) 'Logic, meaning and conceptual rôle', *Journal of Philosophy* **74**, 379-409.

(1978) 'Mental representation', *Erkenntnis* **13**, 9–61.

(1980) *Science Without Numbers*, Oxford: Blackwell.

Fodor, J. A. (1974) 'Special sciences', *Synthese* **28**, 97–115.

(1975) *The Language of Thought*, New York: Crowell.

(1981) *Representations*, Brighton: Harvester.

(1987) *Psychosemantics*, Cambridge, Mass.: MIT Press.

Fodor, J. A. and Pylyshyn, Z. W. (1988) 'Connectionism and cognitive architecture: a critical analysis', *Cognition* **28**, 3–71.

Friedman, M. (1981) 'Theoretical explanation', *Reduction, Time and Reality*, ed. R. Healey, Cambridge: Cambridge University Press, 1–16.

Geach, P. T. (1979) *Truth, Love and Immortality*, London: Hutchinson.

Gettier, E. (1963) 'Is justified true belief knowledge?', *Analysis* **25**, 121–3.

Goodman, N. (1965) *Fact, Fiction, and Forecast*, 2nd edn, New York: Bobbs-Merrill.

Goldman, A. (1980) 'The internalist conception of justification', *Midwest Studies in Philosophy* **5**, 27–51.

Grandy, R. E. (1980) 'Ramsey, reliability and knowledge', *Prospects for Pragmatism*, ed. D. H. Mellor, Cambridge: Cambridge University Press, 197–210.

Hacking, I. (1965) *Logic of Statistical Inference*, Cambridge: Cambridge University Press.

Hacking, I. (1980) 'The theory of probable inference: Neyman, Peirce and Braithwaite', *Science, Belief and Behaviour*, ed. D. H. Mellor, Cambridge: Cambridge University Press, 141–60.

Hellman, G. P. and Thompson, F. W. (1975) 'Physicalism: ontology, determination and reduction', *Journal of Philosophy* **72**, 551–64.

Hempel, C. G. (1965) 'Aspects of scientific explanation', *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*, New York: The Free Press, 331–496.

(1966) *Philosophy of Natural Science*, Englewood Cliffs, NJ: Prentice-Hall.

Hesse, M. B. (1974) *The Structure of Scientific Inference*, London: Macmillan.

(1980) 'A revised regularity view of scientific laws', *Science, Belief and Behaviour*, ed. D. H. Mellor, Cambridge: Cambridge University Press, 87–103.

Hughes, G. E. and Cresswell, M. J. (1968) *An Introduction to Modal Logic*, London: Methuen.

Hume, D. (1739) *A Treatise of Human Nature*, ed. L. A. Selby-Bigge, 1888, Oxford: Oxford University Press.

(1777) 'An enquiry concerning human understanding', *Enquiries concerning the Human Understanding and concerning the Principles of Morals*, ed. L. A. Selby-Bigge, 2nd edn, 1902, Oxford: Clarendon Press, 5–165.

Jackson, F. (1975) 'Grue', *Journal of Philosophy* **72**, 113–31.

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

Jeffrey, R. C. (1969) 'Statistical explanation *vs* statistical inference', *Essays in Honor of Carl G. Hempel*, ed. N. Rescher, Dordrecht: Reidel, 104–13.

(1980a) 'Coming true', *Intention and Intentionality*, ed. C. Diamond and J. Teichman, London: Harvester, 251–60.

(1980b) 'How is it reasonable to base preferences on estimates of chance?', *Science, Belief and Behaviour*, ed. D. H. Mellor, Cambridge: Cambridge University Press, 179–87.

(1983) *The Logic of Decision*, 2nd edn, Chicago: Chicago University Press.

Johnson-Laird, P. N. (1983) *Mental Models*, Cambridge: Cambridge University Press.

Kaplan, D. (1979) 'On the logic of demonstratives', *Journal of Philosophical Logic* **8**, 81–98,

Kim, J. (1973) 'Causation, nomic subsumption, and the concept of event', *Journal of Philosophy* **70**, 217–36.

Kingman, J. F. C. and Taylor, S. J. (1966) *Introduction to Measure and Probability*, Cambridge: Cambridge University Press.

Kirk, R. (1986) *Translation Determined*, Oxford: Clarendon Press.

Kneale, W. C. (1949) *Probability and Induction*, Oxford: Clarendon Press.

Kolmogorov, A. N. (1933) *Foundations of the Theory of Probability*, English transl., 1950, New York: Chelsea.

Körner, S., ed. (1957) *Observation and Interpretation*, London: Butterworths.

Kripke, S. A. (1971) 'Identity and necessity', *Identity and Individuation*, ed. M. K. Munitz, New York: New York University Press, 135–64.

(1972) 'Naming and necessity', *Semantics of Natural Language*, ed. D. Davidson and G. Harman, Dordrecht: Reidel, 763–9.

Kuhn, T. S. (1962) *The Structure of Scientific Revolutions*, Chicago: University of Chicago Press.

(1970) 'Reflections on my critics', *Criticism and the Growth of Knowledge*, ed. I. Lakatos and A. Musgrave, Cambridge: Cambridge University Press, 231–78.

Kyburg, H. (1978) 'Subjective probability: criticisms, reflections, and problems', *Journal of Philosophical Logic* **7**, 157–80.

Levi, I. (1974) 'On indeterminate probabilities', *Journal of Philosophy* **71**, 391–418.

(1980a) *The Enterprise of Knowledge*, Cambridge, Mass.: MIT Press.

Levi, I. (1980b) 'Induction as self correcting according to Peirce', *Science, Belief and Behaviour*, ed. D. H. Mellor, Cambridge: Cambridge University Press, 127–40.

Levi, I. and Morgenbesser, S. (1964) 'Belief and disposition', *American Philosophical Quarterly* **1**, 221–32.

Lewis, D. K. (1969) *Convention*, Oxford: Blackwell.

(1973) *Counterfactuals*, Oxford: Blackwell.

(1980) 'A subjectivist's guide to objective chance', *Philosophical Papers Volume II*, 1986, Oxford: Oxford University Press, 83–113.

(1986) 'Postscripts to "Causation"', *Philosophical Papers Volume II*, Oxford: Oxford University Press, 172–213.

(1989) 'What experience teaches', *Mind and Cognition: a Reader*, ed. W. Lycan, Oxford: Blackwell, 499–519.

Longuet-Higgins, C. (1972) 'The failure of reductionism', *The Nature of Mind*, ed. C. Longuet-Higgins et al., Edinburgh: Edinburgh University Press, 16–32.

Luce, R. D. and Raiffa, H. (1957) *Games and Decisions*, New York: Wiley.

Lycan, W. (1981) 'Psychological laws', *Philosophical Topics* **12**, 9–38.

Mackie, J. L. (1962) 'Counterfactuals and causal laws', *Analytical Philosophy (First Series)*, ed. R. J. Butler, Oxford: Blackwell, 66–80.

(1972) *Truth, Probability and Paradox*, Oxford: Oxford University Press.

(1974) *The Cement of the Universe*, Oxford: Clarendon Press.

Marr, D. (1982) *Vision: A Computational Investigation in the Human Representation of Visual Information*, San Francisco: Freeman.

McArthur, R. P. (1976) *Tense Logic*, Dordrecht: Reidel.

McGinn, C. (1978) 'Mental states, natural kinds and psychophysical laws I', *Aristotelian Society Supplementary Volume* **52**, 195–220.

(1980) 'Philosophical materialism', *Synthese* **44**, 173–206.

(1983) *The Subjective View*, Oxford: Clarendon Press.

McTaggart, J. M. E. (1908) 'The unreality of time', *Mind* **18**, 457–84.

(1927) *The Nature of Existence*, Cambridge: Cambridge University Press.

Mellor, D. H. (1967) 'Imprecision and explanation', *Philosophy of Science* **34**, 1–9.

(1971) *The Matter of Chance*, Cambridge: Cambridge University Press.

(1973) 'Materialism and phenomenal qualities II', *Aristotelian Society Supplementary Volume* **47**, 107–19.

(1974) 'Special relativity and present truth', *Analysis* **34**, 74–7.

(1976) 'Probable explanation', *Australasian Journal of Philosophy* **54**, 231–41.

(1978) 'Conscious belief', *Proceedings of the Aristotelian Society* **78**, 87–101.

(1980) 'Things and causes in spacetime', *British Journal for the Philosophy of Science* **31**, 282–8.

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

(1982) 'The reduction of society', *Philosophy* **57**, 51–63.

(1987a) 'Fixed past, unfixed future', *Michael Dummett: Contributions to Philosophy*, ed. B. Taylor, Dordrecht: Nijhoff, 166–86.

(1987b) 'Indeterministic causation', *African Philosophical Inquiry* **1**, 157–75.

Moore, G. E. (1942) 'A reply to my critics', *The Philosophy of G. E. Moore*, ed. P. A. Schilpp, Chicago: Northwestern University Press, 553–677.

Nagel, E. (1969) *The Structure of Science*, London: Routledge.

Nagel, T. (1983) 'The objective self', *Knowledge and Mind*, ed. C. Ginet and S. Shoemaker, Oxford: Oxford University Press, 211–32.

(1986) *The View from Nowhere*, Oxford: Oxford University Press.

Nozick, R. (1969) 'Newcomb's problem and two principles of choice', *Essays in Honor of Carl G. Hempel*, ed. N. Rescher, Dordrecht: Reidel, 114–46.

(1981) *Philosophical Explanations*, Oxford: Oxford University Press.

Pap, A. (1962) *An Introduction to the Philosophy of Science*, New York: The Free Press.

Papineau, D. (1986) 'Causal factors, causal inference, causal explanation II', *Aristotelian Society Supplementary Volume* **60**, 116–36.

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

Popper, K. R. (1957) 'The propensity interpretation of the calculus of probabilities, and the quantum theory', *Observation and Interpretation*, ed. S. Körner, London: Butterworth, 65–70.

(1990) *A World of Propensities*, Bristol: Thoemmes.

Price, H. H. (1969) *Belief*, London: Allen & Unwin.

Prior, A. N. (1959) 'Thank goodness that's over', *Philosophy* **34**, 12–17.

(1962) 'Changes in events and changes in things', *Papers on Time and Tense*, 1968, Oxford: Oxford University Press, 1–14.

Putnam, H. (1975) 'The meaning of "meaning"', *Mind, Language and Reality*, Cambridge University Press, 215–71.

Quine, W. v. O. (1953) 'On what there is', *From a Logical Point of View*, Cambridge, Mass.: Harvard University Press, 1–19.

(1954) 'The scope and language of science', *The Ways of Paradox*, 1966, New York: Random House, 215–32.

(1960) *Word and Object*, Cambridge, Mass.: MIT Press.

(1969) 'Natural kinds', *Essays in Honor of Carl G. Hempel*, ed. N. Rescher, Dordrecht: Reidel, 5–23.

(1985) 'Events and reification', *Actions and Events*, ed. E. LePore and B. P. McLaughlin, Oxford: Blackwell, 162–71.

Ramsey, F. P. (1925) 'Universals', *Philosophical Papers*, 1990, Cambridge: Cambridge University Press, 8–30.

(1926) 'Truth and probability', *Philosophical Papers*, 1990, Cambridge: Cambridge University Press, 52–109.

(1927) 'Facts and propositions', *Philosophical Papers*, 1990, Cambridge: Cambridge University Press, 34–51.

(1928) 'Universals of law and of fact', *Philosophical Papers*, 1990, Cambridge: Cambridge University Press, 140–4.

(1929a) 'General propositions and causality', *Philosophical Papers*, 1990, Cambridge: Cambridge University Press, 145–63.

(1929b) 'Knowledge', *Philosophical Papers*, 1990, Cambridge: Cambridge University Press, 110–11.

(1929c) 'Philosophy', *Philosophical Papers*, 1990, Cambridge: Cambridge University Press, 1–7.

Ramsey, F. P. (1929d) 'Theories', *Philosophical Papers*, 1990, Cambridge: Cambridge University Press, 112–36.

Russell, B. (1903) *The Principles of Mathematics*, Cambridge: Cambridge University Press.

(1912) 'On the notion of cause', *Mysticism and Logic*, 1953, London: Penguin Books, 171–96.

(1940) 'Egocentric particulars', *An Enquiry into Meaning and Truth*, London: Allen & Unwin, 102–9.

Ryle, G. (1949) *The Concept of Mind*, London: Hutchinson.

Salmon, W. C. (1979) 'Propensities: a discussion-review', *Erkenntnis* **14**, 183–216.

(1984) *Scientific Explanation and the Causal Structure of the World*, Princeton: Princeton University Press.

Sanford, D. H. (1985) 'Causal relata', *Actions and Events*, ed. E. LePore and B. P. McLaughlin, Oxford: Blackwell, 282–93.

Schlesinger, G. (1963) *Method in the Physical Sciences*, London: Routledge & Kegan Paul.

Searle, J. R. (1958) 'Proper names', *Mind* **67**, 166–73.

Shepard, R. N. and Cooper, L. A. (1982) *Mental Images and Their Transformations*, Cambridge, Mass.: MIT Press.

Skyrms, B. (1980a) *Causal Necessity*, New Haven: Yale University Press.

(1980b) 'Higher order degrees of belief', *Prospects for Pragmatism*, ed. D. H. Mellor, Cambridge: Cambridge University Press, 109–37.

Sober, E. (1986) 'Causal factors, causal inference, causal explanation I', *Aristotelian Society Supplementary Volume* **60**, 97–115.

Stich, S. P. (1983) *From Folk Psychology to Cognitive Science*, Cambridge, Mass.: MIT Press.

Storer, T. (1950–1) 'On defining "soluble"', *Analysis* **11**, 134–7.

Stroud, B. (1987) 'The physical world', *Proceedings of the Aristotelian Society* **87**, 263–77.

Suppes, P. (1984) *Probabilistic Metaphysics*, Oxford: Blackwell.

Swoyer, C. (1982) 'The nature of natural laws', *Australasian Journal of Philosophy* **60**, 203-23.

Tarski, A. (1944) 'The semantic conception of truth', *Readings in Philosophical Analysis*, ed. H. Feigl and W. Sellars, 1949, New York: Appleton-Century-Crofts, 52–84.

Taylor, B. (1985) *Modes of Occurrence*, Oxford: Blackwell.

Teller, P. (1979) 'Quantum mechanics and the nature of continuous physical quantities', *Journal of Philosophy* **76**, 345–61.

Tooley, M. (1977) 'The nature of laws', *Canadian Journal of Philosophy* **7**, 667–98.

Turner, J. (1955) 'Maxwell on the method of physical analogy', *British Journal for the Philosophy of Science* **6**, 226–38.

Urbach, P. (1988) 'What is a law of nature? A Humean answer', *British Journal for the Philosophy of Science* **39**, 193–210.

van Cleve, J. (1984) 'Reliability, justification and the problem of induction', *Midwest Studies in Philosophy* **9**, 555–67.

Vendler, Z. (1962) 'Effects, results and consequences', *Analytical Philosophy: First Series*, ed. R. J. Butler, Oxford: Blackwell, 1–15.

Venn, J. (1866) *The Logic of Chance*, 4th edn, 1962, New York: Chelsea.

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

Whyte, J. T. (1991) 'The normal rewards of success', *Analysis* **51**, 65–74.

Williams, D. C. (1951) 'The myth of passage', *Journal of Philosophy* **48**, 457–72.

Wilson, M. (1982) 'Predicate meets property', *Philosophical Review* **91**, 549–89.

(1985) 'What is this thing called "pain"? – the philosophy of science behind the contemporary debate', *Pacific Philosophical Quarterly* **66**, 227–67.

Wittgenstein, L. (1922) *Tractatus Logico-Philosophicus*, London: Routledge.

Woolhouse, R. S. (1973) 'Tensed modalities', *Journal of Philosophical Logic* **2**, 393–415.

Zemach, E. M. (1976) 'Putnam's theory on the reference of substance terms', *Journal of Philosophy* **73**, 116–27.