PHILOSOPHY 2

FURTHER THROUGH THE SUBJECT

EDITED BY

A. C. GRAYLING

OXFORD UNIVERSITY PRESS

1998

Sober, E., 'Mathematics and Indispensability', *Philosophical Review*, vol. 102 (1993), pp. 35–57.

Neo-Fregeans

A version of platonism based on Frege's context principle is advanced in:

Wright, C., *Frege's Conception of Numbers as Objects* (Aberdeen, 1983).
Field, H., 'Platonism for Cheap? Crispin Wright on Frege's Context Principle', in *Realism, Mathematics and Modality* (Oxford, 1991). An opposed view.
Hale, B., and Wright, C., 'A *Reductio ad Surdum* . . . ?', *Mind*, vol. 103 (1994), pp. 169–83.

Structuralism

Resnick, M., 'Mathematics from the Structural Point of View', *Revue Internationale de Philosophie*, vol. 42 (1988), 400–24. Resnick advocates a form of platonism according to which mathematics is the science of structures.

Modern Anti-Platonism

Feferman, S., 'Why a Little Goes a Long Way: Logical Foundations of Scientifically Applicable Mathematics', in *Proceedings of the Philosophy of Science Association*, vol. 2 (1993), pp. 442–55. (t) Feferman maintains that only a small part of mathematics is needed for science.
Field, H., *Science without Numbers: A Defence of Nominalism* (Oxford, 1980).
Shapiro, S., 'Conservativeness and Incompleteness', *Journal of Philosophy*, vol. 81 (1983), pp. 521–31. (t) A criticism of Field.
Papineau, D., 'Mathematical Fictionalism', in *International Studies in the Philosophy of Science*, vol. 2 (1988), pp. 157–74.

Knowledge of Mathematics

Kitcher, P., *The Nature of Mathematical Knowledge* (Oxford, 1984).
Field, H., 'Tarski's Theory of Truth', *Journal of Philosophy*, vol. 69 (1980), pp. 347–75.
Goldman, A., 'A Causal Theory of Knowing', *Journal of Philosophy*, vol. 64 (1967), pp. 357–72.
Benacerraf, P., 'Mathematical Truth', in B&P. Benacerraf's problem arises if one combines Goldman's view of knowledge with Field's view of truth and reference.
Maddy, P., 'Perception and Mathematical Intuition', *Philosophical Review*, vol. 89 (1980), pp. 163–96. An empiricist account presented as an answer to Benacerraf.
——*Realism in Mathematics* (Oxford, 1990).
Field, H., 'Realism, Mathematics and Modality', in *Realism, Mathematics and Modality* (Oxford, 1991).
Chihara, C., *Constructibility and Mathematical Existence* (Oxford, 1991). Chihara criticizes both Maddy's realism and Resnick's structuralism.
Lakatos, I., *Proofs and Refutations* (Cambridge, 1976). Arguably the most enjoyable of all items listed here: a stimulating exploration of mathematical proof as fallible, and of mathematicians' reactions to proposed counter-examples.

4

PHILOSOPHY AND THE NATURAL SCIENCES

John Worrall

INTRODUCTION

For better or worse (the former in my view), the development of science has had an overwhelming impact on our culture. At the practical level this is obvious enough—space probes, television, personal computers, . . . Add your own favourites to the list. Although some items on the extended list will seem to many mixed blessings at best, even the most technophobic would be hard-pressed to deny that the human condition has been improved by, for example, the increased diagnostic precision afforded by such techniques as computer-assisted tomography or the humbler (only because older) X-ray photograph.

The intellectual impact of the development of science may be less tangible, but is surely of at least equal importance. The modernist, Enlightenment view was that the development of modern science had freed humans from superstition and myth showing them that—amazingly enough—they can discover the innermost secrets of the workings of the universe if they use their intellects appropriately, that is, rationally or scientifically. Everyone knows Alexander Pope's couplet reflecting the eighteenth-century view:

> Nature and Nature's laws lay hid in night,
> God said 'Let Newton be!' and all was light.

But there lies the, more recent, rub. For almost everyone knows Sir John Squire's twentieth-century rejoinder:

> 'Twas not to last, for Devil howling 'Ho,
> Let Einstein be!' restored the status quo.

How exactly can the idea of science as the bastion of objectivity and rationality withstand the impact of the great 'scientific revolutions' brought about, for example, by relativity theory and (perhaps more fundamentally still) by quantum theory? Why do such apparently radical changes in theory occur? Is it for reasons dictated by accumulating observational evidence? If so, according to what precise logic of evidence? If not, if some of the 'reasons' are subjective or social, what differentiates science from other bodies of theoretical claims often thought of as altogether less firmly grounded (such as religious or pseudo-scientific claims)? And what exactly do such revolutionary changes in theory imply about the epistemic status of presently accepted theories in science? Do they, in particular, refute the *scientific realist* view of accepted theories as—perhaps approximate—truths? A nineteenth-century realist would have advocated belief in the approximate truth of the theories then accepted in science. Yet presently accepted theories seem—in fundamentals—radically at odds with these earlier ones: for example, Newtonian theory states that the planets

are kept in their (roughly) elliptical paths by action-at-a-distance gravitational forces, yet relativity theory explicitly rejects action-at-a-distance and attributes the planets' orbits to their following geodesics in (curved) spacetime.

Many of the presently central issues in what might be called general philosophy of science concern its 'three Rs': rationality, realism, and revolution. Some of these issues are outlined in Section 1, which includes a brief account of Kuhn's highly influential views, and an outline and critical examination of what is currently the most popular and certainly best-developed *formal* account of rational belief, 'personalist Bayesianism'. (On a number of issues in Section 1, I presuppose—and extend—the treatments given by David Papineau in his chapter 'Methodology: The Elements of the Philosophy of Science' in the earlier companion volume to this book.)

One recent trend in philosophy of science—itself due in part to Kuhn's influence—has been toward the *naturalistic* view that philosophy of science consists simply of descriptions of the way that mature sciences operate. In Section 2, I outline some of these approaches and raise the question whether naturalized philosophy of science can retain normative force or instead must sacrifice the implication that science is genuinely epistemically special.

While there may be reasons to resist the idea that philosophy of science is itself at bottom a science, there can be no doubt that many of its most interesting problems arise in close association with science. Section 3 contains an account of some of these problems: one arising from quantum physics and a couple from evolutionary biology. These problems are more or less randomly selected in an attempt to give some flavour of that important part of philosophy of science concerned, in effect, with analysing and clarifying the logical implications and presuppositions of current scientific theories.

1. RATIONALITY, REVOLUTION, AND REALISM

1.1. Radical Theory Change in Science

Newton's theory (of mechanics plus universal gravitation) is logically inconsistent with Einstein's theory of relativity. The former entails that the universe is infinite; that time is absolute (so that two events simultaneous in one frame of reference are simultaneous in all); that every body acts at a distance on every other body; and that the inertial mass of a given body is a (velocity -independent) constant. Relativity theory contradicts each of these entailments: according to it, the universe may be finite (though unbounded); any two spatially separated events simultaneous relative to one frame of reference are not simultaneous relative to another frame moving with respect to the first; there is

no action-at-a-distance; and the inertial mass of a body increases with its velocity. How far 'down' towards the empirically testable consequences of the theory this inconsistency reaches is an issue to which we shall need to return, but at the level of fundamental theory there is simply outright contradiction.

This is no isolated case. Consider, for example, the history of optics: in the seventeenth and eighteenth centuries the most popular theory of the fundamental nature of light was that it consists of tiny material particles; in the early nineteenth century this was displaced by the theory that light consists not of matter but of motion—of periodic motions (waves) transmitted through an all-pervading elastic medium; in the late nineteenth century this theory was in turn replaced by the claim that light consists of vibrations carried, not by a material medium, but by an immaterial electromagnetic field; and finally (so far!) this theory was replaced by the claim that light consists of photons obeying an entirely new quantum mechanics.

Whether such discontinuities are quite as sharp as they have sometimes been made to seem, and whether they pose as severe a threat as is often imagined to the idea that the development of science is a rational process, are questions that remain to be discussed; but the fact that at the level of fundamental theory in science such changes have occurred is surely undeniable.

1.2. The Impact of Kuhn's *The Structure of Scientific Revolutions*

Although Karl Popper had been emphasizing the importance of 'revolutionary' changes in science for many years beforehand, it was Thomas Kuhn's 1962 book *The Structure of Scientific Revolutions* that really brought radical theory change to centre-stage in philosophy of science. Kuhn's views seemed more challenging to older conceptions than Popper's. But just what those views are was an immediate cause of controversy and has remained so ever since. Some of those influenced by Kuhn, encouraged by his talk of 'incommensurability' involving (alleged) total breakdown of communication between scientists in different 'paradigms', of total 'revolutionary' revisions of even the evidential basis of science, and even of the *world* changing as the dominant paradigm does, have seen the book as demolishing the idea that theory change in science is a rational process.

I shall not become embroiled in Kuhnian exegesis here: those who wish to follow the twists and turns of the reinterpretations (or clarifications?) of Kuhn's more outlandish-sounding claims can find references in the bibliography. I shall instead accentuate (what I take to be) the positive. I first explain what seems to me valuable and correct in Kuhn's account and then try to clarify the sharpest challenge to the idea of scientific rationality that these correct and valuable views pose.

1.2.1. *Paradigms versus Theories*

Kuhn insisted that science and its development cannot be analysed satisfactorily in terms of single theories: the unit of scientific commitment is not the theory but the 'paradigm'. This is a notoriously unclear term (a fact that has assisted its subsequent widespread use—or abuse). The main significance of Kuhn's idea is best explained through a simplified and somewhat idealized example from the history of science.

The predominant view about the nature of light in the eighteenth century was that it consists of tiny material particles; these are emitted from sources, such as the sun, and follow paths that are, like those of all other particles, governed by Newton's laws. In particular they travel in straight lines (at constant velocity) unless acted upon by a net external force; and conversely, whenever the particles are bent out of their rectilinear trajectories (as they are, for example, when reflected from a mirror or when refracted on entering a transparent medium like water), there must be some force that accounts for this bending.

Two features should be noted. First, this is a very general set of ideas—in order to make detailed contact with empirical phenomena it must be augmented with specific assumptions *both* about the light-particles themselves (what is it, for example, that differentiates those particles that produce the sensation of blue light from those that produce the sensation of red?) *and* about the forces acting upon them in particular circumstances (how exactly do the 'reflecting force' and the 'refractive force' act, *and* how do they *interact*?). Secondly, this process of producing specific theories within the general framework supplied by the basic ideas is not a shot in the dark—the process does not consist of a series of 'bold conjectures'. Instead the sorts of particular assumption that might work are indicated by the general framework itself, in conjunction with 'background knowledge'. The general idea was to 'reduce' optics to particle mechanics. Particle mechanics had, through the work of Newton and his successors, already become a highly developed field. Various sorts of candidate for the differentiating features of the light-particles were already available—blue-making particles might, for example, have a different mass from red-making ones, or perhaps a different velocity; and work in other fields supplied ideas about the sorts of force that might do the job in optics. Moreover, the fact that particle mechanics was already a mathematically highly developed theory meant that any specific theory developed within the framework would be mathematically tractable—that is, scientists would be able to deduce what consequences it has at the empirical level.

What existed, then, in the eighteenth century was not a single theory of light, but a general underlying idea (light is *some sort* of particle affected by *some sorts* of force) together with a set of ideas for identifying the particular sorts of

particle and the particular sorts of force these might be—a set of ideas provided by previous scientific successes. The problem was to construct specific theories which would account for the phenomena within the general framework supplied by the corpuscular optics 'paradigm'. Paradigms, then, are characterized chiefly by a general theory and a set of ideas for developing that general theory into specific theories that will capture and explain the relevant phenomena. Kuhn refers to this set of ideas as underwriting a 'puzzle-solving tradition', previous successes often supplying 'exemplars' for later developments. Although Kuhn's detailed development of this view—especially his emphasis on inarticulable skills, 'disciplinary matrices', and the like—can be challenged (and certainly stands in need of clarification), he was surely pointing in the direction of an important and then relatively neglected aspect of mature science. Imre Lakatos, with his notion of a research programme complete with 'positive heuristic', and Larry Laudan, with his notion of a research tradition, both later underlined this same point in slightly different (and considerably sharper) ways.

1.2.2. *Anomalies versus Experimental Refutations*

Some phenomena were recognized as presenting special problems for the corpuscular optics paradigm—as especially difficult for any specific theory developed within that paradigm to capture. One such phenomenon was that of partial reflection: if a beam of light in air is incident on the surface of some transparent material, like glass or water, then in general only part of the light is refracted into the material, while the rest is reflected back into the air. The corpuscular optics approach dictated a (repulsive) force of reflection and an (attractive) force of refraction; but why, as partial reflection seemed to indicate, were some of the particles acted on by the reflective force and others by the refractive? In Kuhn's terms this was an experimental *anomaly* for the corpuscular optics paradigm. As Kuhn claims is generally true, the anomaly was regarded, at least initially, as a problem to be solved within the paradigm rather than as a 'falsification' or reason to reject (or even seriously to question) the paradigm.

Once it is recognized that specific theories are generated within general frameworks by the addition of detailed assumptions, the whole debate about anomalies and 'falsifications' (see Lakatos and Musgrave 1970) is remarkably easy to resolve. Sometimes there is a set of detailed assumptions that is privileged in some way—perhaps because the specific theory based on them has been successful with some other phenomena—and that specific theory is actually inconsistent with the anomalous phenomenon (or rather with its description). Sometimes (perhaps more often) specific theories for a certain range of phenomena are in the process of construction—that is, no particular

set of more detailed assumptions recommends itself; but it is clear that the anomalous phenomenon is going to be a special difficulty for that construction because any of the obvious or straightforward detailed assumptions suggested by the framework would produce overall theories inconsistent with the phenomenon at issue. It is into this second class that, for example, partial reflection falls.

Neither kind of anomaly, of course, yields any direct reason to give up the general framework theory. When, for example, the initial corpuscular account of what happens when light passes the edge of an opaque object (namely that it proceeds undisturbed in its rectilinear path) turned out to be inconsistent with the phenomenon of 'straight-edge diffraction', corpuscularists simply assumed that there must be a diffracting force (in addition to the already assumed reflective and refractive forces) and proceeded to try to pin down the features of that diffracting force. Kuhn's point about anomalies of this first kind amounts in effect just to the one made long ago by Duhem (1906): deductive logic does not require giving up the general theory underlying the framework in such a case, but only giving up *either* that theory *or* one of the erstwhile preferred specific assumptions. If some observational statement O follows not from T (the 'general framework' theory) alone but only from $T \& S$ (where S is some set of more detailed assumptions couched within that same framework but not essential to it) then, should O turn out to be false empirically, it follows only that so also must be either T or at least one of the particular assumptions in S.

There is no general reason why a scientist should not hold onto T in such a situation and therefore identify the problem as that of finding and replacing the 'faulty' detailed assumption. On the contrary, there is good reason to retain T, at least as the initial move. The general theory T underpins a whole approach to optics—it sustains in Kuhn's terms a 'puzzle-solving tradition'; hence retaining T means retaining a framework that at least to some degree guides the construction of replacement specific theories. The scientist who sees straight-edge diffraction as a 'puzzle' to be solved within the corpuscular optics approach rather than as a reason to reject the whole approach is simply obliged to give up the idea that no force affects the light-particles as they pass the edge of opaque bodies, and is pointed towards postulating a 'diffracting force'—one that may be attractive at some distances and repulsive at others. Once some assumption about the forces is made, the existing mathematics of particle mechanics permits the deduction of logical consequences. A scientist who might instead be inclined to reject T on account of this phenomenon would be left—at any rate initially—in an intellectual vacuum.

The second type of anomaly (exemplified in the historical case under consideration by partial reflection) serves, albeit more indirectly, to underline this same methodological lesson. Such an anomaly provides a potential 'Duhem problem' rather than a real one: if any set of available detailed assumptions were added to the general corpuscular view, this would produce a specific theory inconsistent with the facts about partial reflection, but there was in this case no independent reason to prefer any particular set of such detailed assumptions. The whole problem right from the start in such a case was to articulate a specific theory (add a set of detailed assumptions to the core view) that would correctly yield the phenomenon at issue. Partial reflection was an 'anomaly' because it was clear that no set of straightforward assumptions was going to succeed in capturing it. Again it is no wonder that the option of retaining T was attractive: taking that option constrains the problem and makes it manageable. One fairly obvious suggestion (investigated by Newton himself) was to give up the idea (always clearly an idealization) that the light consists of point particles and try the idea that they might, as bodies of finite extension, have different properties on different 'sides' (so that the particles arriving at the interface with one side or 'pole' uppermost might be, say, reflected, while those arriving there with the other 'pole' uppermost might be refracted into the glass).

1.2.3. *Revolutions and Reason*

A particular field is, on Kuhn's account, practising *normal science* when there is only one remotely plausible general framework theory available and anomalies are routinely treated as problems for the paradigm to solve rather than as bringing that paradigm into question. As the terminology suggests, this is—in his view—the standard state of a scientific field (at any rate once that field has achieved 'maturity'). Kuhn seems now to have accepted that he initially overdid the 'paradigm monopoly' view—although most eighteenth-century scientists preferred the corpuscular approach to optics there were always significant dissenting voices (one belonged to Euler). Moreover, there are at least some fields of apparently 'mature' science—for example, various fields in contemporary biological sciences—where there are two or more rival general views vying for acceptance. It none the less has sometimes happened, especially in physics and chemistry, that for long periods one general view, one paradigm, has dominated. But there are also periods when the general view is challenged and replaced by one inconsistent with it. Why do these changes occur? It was always Kuhn's views about the process of paradigm *change* or about 'extraordinary' or 'revolutionary' science that provided the chief target for critical fire (Lakatos, for example, claimed that Kuhn's views made theory change in science a matter of 'mob psychology').

The corpuscular approach may not have monopolized optics in the eighteenth century but it does seem to have been easily the most widely accepted approach. In the early decades of the nineteenth century the theory that light consists of periodic disturbances transmitted through a mechanical, elastic

medium became dominant. Did this replacement (Kuhn, of course, likes the term 'revolution') occur for good, objective reasons? The account in *The Structure of Scientific Revolutions*—with its talk of a *crisis* of confidence affecting the scientific community, leading some, usually younger scientists, to undergo 'something akin to a religious conversion' to an alternative approach, leading in turn to a 'bandwagon effect' that sees a new community consensus form around the new alternative—seemed to many commentators to amount quite unambiguously to an irrationalist (or arationalist) view of theory change. Although a state of 'crisis' is, on Kuhn's account, always produced by an increase in the number of anomalies and/or by the persistent intractability of certain anomalies, he seemed quite explicit that there were, and could be, no general rules for counting and weighing anomalies and so no threshold beyond which 'crisis' was justified. And Kuhn was similarly explicit that there are no rules for when the shift to a new paradigm becomes rationally dictated and hence continued adherence to the old paradigm irrational: either a new consensus will form around the new basic idea or it will not, that is all there is to the matter.

Most of what his critics found objectionable in Kuhn's account of theory change is reflected in his remarks about 'hold-outs' to 'scientific revolutions'. He claimed that if we look back at any case of a change in fundamental theory in science we shall always find eminent scientists who resisted the switch to the new 'paradigm' long after most of their colleagues shifted. These 'hold-outs'—Priestley's defence of phlogiston against Lavoisierian chemistry is a celebrated example—are often (though not invariably) by elderly scientists who have made significant contributions to the older paradigm. Kuhn added to this interesting but relatively uncontroversial descriptive claim the challenging normative claim that these 'elderly hold-outs' were no less justified than their more fickle contemporaries: not only did they, as a matter of fact, stick to the older paradigm, they were, moreover, not wrong to do so. On Kuhn's view, 'neither proof *nor error* is at issue' in these cases, there being, as he added in a later attempt to clarify his views, 'always some good reasons for every possible choice'—that is, both for switching to the revolutionary new paradigm and for sticking to the old. Hence the hold-outs cannot, on Kuhn's view, be condemned as 'illogical or unscientific'. But neither of course can those, like Lavoisier, who switch to the new paradigm be so condemned. It is this alleged absence of a single 'correct' course of action or single correct set of beliefs for a scientist to adopt that seemed the most threatening aspect of Kuhn's position.

What argument did Kuhn have for this claim? Why, that is, did he hold that those who resist the new paradigm are no less rational than their more mobile colleagues? According to his account in *The Structure of Scientific Revolutions*: 'The source of resistance is the assurance that the older paradigm will ultimately solve all its problems, that nature can be shoved into the box the paradigm supplies' (1962, 151–2). Not only does this feeling of assurance fail to be 'illogical or unscientific', it is that same feeling of assurance that 'makes normal or puzzle-solving science possible'.

But this argument is disappointing. It is of course true that the hold-outs cannot be faulted as 'illogical', if illogicality would require flying in the face of *deductive* logic. Echoing Duhem's point of long ago, the general theories that form the basis of a paradigm have no directly testable deductive consequences of their own, so there must always be *some* auxiliary assumptions that when added to those general theories will entail any given set of evidential statements. To take an example from optics again, although diffraction and interference phenomena are often nowadays taken as direct refutations of the corpuscular approach, this is quite wrong and is certainly not how those phenomena were viewed at the time. Just as Kuhn suggests, the defenders of the corpuscular theory did indeed insist that these phenomena could be shoved into the emissionist 'box'. More interestingly, they actually constructed explanations—at any rate in outline—of how the phenomena could be so 'shoved': by postulating a complicated force of diffraction, for example, or by making interference a physiological phenomenon (the fringes being produced by interference of waves produced in the eye by the light-particles). Such explanations are bound to exist *if all that we require of them is that they deductively yield correct descriptions of the phenomena at issue.* (After all, if this were the only requirement, then the following 'theory' would suffice: light consists of particles subjected to entirely unknown forces emanating from the edges of opaque objects that happen to result in the particles moving along unspecified paths that produce the following patterns of illumination . . . (where we simply fill in the dots on the basis of the known experimental results).)

But are we bound to say that constructing such an explanation automatically balances the evidential scales (at any rate with respect to the phenomena at issue)? Only if we hold that if some evidential statement e is entailed by each of two theories T_1 and T_2 then e confirms both theories to the same extent and so can supply no reason to prefer one of the two theories. But surely no sensible account of confirmation can endorse such a view? All sensible accounts must come to terms with elusive, but clearly important notions like simplicity, unity, and absence of *ad hoc*-ness. No doubt corpuscularists could cobble together a theory that had a correct description of the phenomenon of straight-edge diffraction as a deductive consequence, but this phenomenon may still yield an objective reason to prefer the wave theory if (as is actually the case) it falls out in an entirely natural way from that theory but—so far as we can tell—can only be accounted for in an *ad hoc* way by the corpuscular theory. (The intuitive difference here is often expressed as the difference between explaining a phenomenon (as the wave theory explains diffraction) and merely capturing it *post hoc* (which was the best the corpuscular theory could do).)

Suppose we take it that the principal aim of a theory of scientific rationality is to supply a natural and defensible account of when evidence objectively supports some scientific theory. Suppose further that the claim that the development of science has been by and large a rational process amounts simply to the claim that theory change has always been from theories that are less well supported empirically to theories that are better supported empirically. It follows that Kuhn's points about hold-outs present no threat to the idea of scientific rationality, provided that an account of confirmation or empirical support can be produced that has the following consequence: the support lent to a general theory (or paradigm) by some phenomenon that has been 'shoved' into its 'box' is, in general, less than that lent to that same theory by some phenomenon that 'falls naturally out of its box' (that is, one that is given a 'natural', straightforward explanation within that general theory).

We shall see in Section 1.3.6 whether the approach to empirical support that is currently most popular—personalist Bayesianism—can indeed underwrite such a distinction.

1.2.4. *Kuhn's Later Account of 'Theory Choice'*

In chapter 13 of his (1977) book *The Essential Tension*, Kuhn develops a much more explicit (and, I believe, more challenging) account of the factors underlying what he calls 'theory choice'. He there insists that he never denied that 'reason' in the form of the 'objective factors' from the philosopher's 'traditional list' (including such factors as empirical accuracy and scope, consistency, simplicity, and 'fruitfulness') plays a crucially important role in theory change: 'I agree entirely with the traditional view that [these objective factors] play a vital role when scientists must choose between an established theory and an upstart competitor . . . they provide *the* shared basis for theory choice' (Kuhn 1977: 322). But, these objective factors, Kuhn claims, supply 'no algorithm for theory choice'. At any rate when the choice between rival theories is a live issue in science, these objective factors never dictate a choice. This is for two main reasons. First, single factors often turn out to deliver no unambiguous preference when applied to the theories *as they stood at the time when the choice was being made*. For example, it is often assumed that the Copernican heliostatic theory was empirically more accurate than the Ptolemaic theory. This eventually became true but only as a result of the work of Copernicus, Kepler, Galileo, and others, who had clearly then 'chosen' the Copernican theory for other reasons (if for any *reasons* at all). Secondly, even where single 'objective factors' do point clearly in the direction of one of the rival theories, different factors may point in opposite directions: while simplicity (in a certain sense) favoured Copernican theory, consistency (with other, then accepted theories) undoubtedly favoured the Ptolemaic theory.

Hence, according to Kuhn, the objective factors must always be supplemented by 'subjective' factors in order to deliver a definite preference: 'every individual choice between competing theories depends on a mixture of objective and subjective factors, or of shared and individual criteria' (Kuhn 1977: 325).

Many interesting issues are raised by this later account. Here I give bare outlines of a few of them. (Readers interested in more details should follow up the references supplied in the Bibliography.)

1. Isn't Kuhn's laundry list of 'objective factors' lacking in necessary structure? He presents it as if each of the factors is independent of all the others and that the 'objectivist' could give no reasons for regarding some of the factors as more important than others. In fact many of those philosophers who aim to show that theory choice is an objective affair (Poincaré, Duhem, Lakatos, and many others) would see one of Kuhn's objective virtues—that of predictive (empirical) success—*both* as intimately connected with other virtues (for example, 'fruitfulness') *and* as dominant over others (such as detailed empirical accuracy).

After all, Duhem's point about theory-testing means that scientists always can, in principle, develop detailed theories within a given framework that will capture given known phenomena. What scientists cannot guarantee is that the general framework will be predictively successful. The Copernican who (1) acknowledges that Ptolemaic theory had developed much the more extensive fit with the phenomena (after all it had had several hundred years' start) but who (2) insists that hard work on detailed assumptions within his theory will eventually allow it to match Ptolemaic theory in this regard, and who (3) points to, say, planetary stations and retrogressions as predictive successes for his theory unmatched (and perhaps unmatchable) by the Ptolemaists—such a Copernican is surely winning the objective argument.

Or consider Kuhn's 'objective factor' of consistency with other theories. The suggestion is that because, for example, Copernican theory clashed with accepted Aristotelian cosmology (whereas Ptolemaic theory was consistent with Aristotle), it remained reasonable for those who ranked consistency over, say, predictive success to continue to advocate the Ptolemaic theory. But in fact inconsistency with other accepted theories may surely be reasonably regarded sometimes as a virtue rather than a vice, since it sets an agenda of problems for further research—with, however, the important proviso that there is some independent (empirical) reason for thinking that these problems *can* be satisfactorily solved. Again it seems to be predictive success that supplies that independent reason.

2. Is Kuhn correct that his account diverges comparatively little from that 'currently received' in the philosophy of science? The answer seems to be that because of the essential role it ascribes to 'subjective factors' in supplementing

the always indecisive objective ones, it does indeed sit well with the personalist Bayesian account of rational belief. But it does so on account of the features of personalist Bayesianism that many other philosophers find objectionable (basically, as we shall see later, (alleged) over-reliance on subjective prior probabilities). Although Kuhn advertises his more elaborate account as a decisive rebuttal of the claim that his view makes theory change in science a matter of 'mob psychology', it should be remembered that his account still has the explicit consequence that 'there are always some good reasons for each possible choice (i.e. both for sticking to the "old" theory and for adopting the "new" one)' (Kuhn 1977: 328). And it still has the explicit consequence that the eventual resolution of 'revolutions' is simply a matter of a new consensus happening to emerge around the new paradigm: 'valid' reasons for adherence to the older view run out when and only when the community ceases to take seriously those who subjectively see such reasons.

1.3. The Personalist Bayesian Account of Rational Belief

1.3.1. *Probability and Evidence*

No amount of evidence can deductively prove a scientific theory (this is 'Hume's problem'); and no amount of evidence can strictly disprove a scientific theory (this might be called 'Duhem's problem'). Some might be tempted to give up on the whole idea that evidence supplies the reason for accepting certain theories over others. But this seems much too quick. Another possible (and, given the undeniable and staggering empirical success of science, surely altogether more plausible) conclusion is that proofs and disproofs were always too much to ask: evidence may establish some theories as at any rate *more* rationally believable than others (indeed some of the less believable theories may be rendered close to incredible by the evidence). An obvious suggestion for trying to make this idea precise is to use the notion of probability. Theories may not be proved by the evidence, but they may be much more probable than any of their rivals in the light of the evidence. Theories may not be disproved by evidence, but they may be made so improbable by it that it hardly makes a difference. There is a long history of attempts to develop these suggestions. These attempts have the great advantage of being able to exploit the simple and mathematically precise theory of probability, which helps make discussion of the issues much sharper than is usual in philosophy.

1.3.2. *Carnap and Probabilistic Inductive Logic*

Readers who are not already acquainted with the axioms of probability should consult the brief treatment by David Papineau in the companion to this volume

or one of the references in the Bibliography. Although initially developed to apply to 'events' (the event of a head coming up on the toss of a coin, the event of a radioactive atom undergoing decay in a given time interval, etc.), the probability calculus can also be interpreted as applying to sentences. The philosopher of science Rudolf Carnap in fact developed the idea that the probability of a sentence S—the chance that S is true—is the proportion of the 'possible worlds' in which S holds compared to all 'possible worlds'. Thus a tautology has probability one since it is true in all possible worlds, and a contradiction has probability zero (false in all possible worlds). The probability of the sentence *either* S *or* S' is the measure of the union of those possible worlds in which S is true and those in which S' is true *minus* the possible worlds in which both are true (otherwise the possibilities in which both are true would be 'counted twice'). So, in accord with the probability axioms,

$$p(S \vee S') = p(S) + p(S') - p(S \,\&\, S').$$

An important notion in probability theory is that of the *conditional probability* $p(S \mid S')$, the probability of S conditional on S'. In the case of probabilities of events, we can, for example, ask for the probability that a draw of a single card from a well-shuffled pack has produced a heart, given that the card drawn was a red card. (We may have caught a glimpse of the card and seen that it was red, but not been able to discern whether it was a heart or a diamond.) Intuitively the answer is $\frac{1}{2}$ (whereas the unconditional probability of a heart is of course $\frac{1}{4}$). Similarly, the probability that the card drawn was a spade, given that it was red, is of course 0 (again the unconditional probability being $\frac{1}{4}$).

Provided that $p(S') \neq 0$, then the conditional probability $p(S \mid S')$ is equal to the ratio $p(S \,\&\, S')/p(S')$. So, on Carnap's interpretation, $p(S \mid S')$ measures the proportion of possible worlds in which S' holds that are also ones in which S holds. The idea was that the probability calculus, interpreted in this way, would provide a natural extension of deductive logic. The inference from premiss S' to conclusion S is, remember, deductively valid if and only if every interpretation of the language in which these sentences are expressed (every 'possible world') that makes S' true also makes S true. So when S' implies S, S must be true in all the possible worlds in which S' is; and hence the above ratio is 1; that is, $p(S \mid S') = 1$. If there is at least one interpretation in which S' is true and S false, then S' fails to entail S deductively. But Carnap's idea was to give formal sense to such intuitively appealing ideas as that S' may 'almost imply' S, while for other such pairs of sentences S' may 'almost imply' that S is false. In the former case, S would be true in nearly all the possible worlds in which S' is, and so, in the Carnap interpretation, $p(S \mid S') \approx 1$; in the latter case S would be false in nearly all the possible worlds in which S' is true and so $p(S \mid S') \approx 0$.

Carnap's idea was that this extension of the logic of deductive entailment to a logic of partial entailment would underwrite such claims as the following:

although the (total) evidence does not entail any of the available theories $T_1, \ldots, T_n$, it none the less implies one of them, say T_3, to a higher degree (perhaps to a much higher degree) than any of its rivals: $p(T_3 \mid e) >> p(T_i \mid e)$ for any $i \neq 3$; and although the evidence does not entail the falsity of some theory T, it 'almost' does so: $p(T \mid e) \approx 0$. These judgements would be just as objective as the straightforward deductive judgements of which they are natural generalizations.

A successful account of partial entailment between sentences would at the same time be, on this approach, an account of *rational degrees of belief*. Just as a rational person believes tautologies (such as 'either probability theory is hard or it isn't') totally and disbelieves totally contradictions (such as 'probability theory is hard and it isn't'), and just as she must rationally believe that Socrates is mortal, given that she believes that all men are mortal and that Socrates is a man, so she must rationally believe Newton's theory to degree r, if she accepts total evidence e and if $p(\text{Newton} \mid e) = r$. The approach, then, would—it seems—deliver an entirely objective answer to the rational preference problem, for example:

Q: Why is it rational (scientific) to prefer the Darwinian theory of evolution to the creationist account, given all evidence e that we have (about fossils, homologies, etc.)?

A: Because $p(\text{Darwin} \mid e) >> p(\text{creationism} \mid e)$; that is, because e entails Darwinian theory to a much higher degree than it entails creationism, and hence Darwinian theory is (objectively) much more likely to be true, given e, than is creationism.

This is a terrific idea. Unfortunately it doesn't work. It fails for the same reasons that the so-called classical interpretation of probability fails. The fatal flaw lies in the implicit idea that there is always only one way of describing or dividing up the possibilities, so that there is always one answer to the question 'What is the proportion of possible worlds in which S' holds to possible worlds in which both S' and S hold?' The flaw surfaces even in apparently straightforward cases.

Suppose there are three individuals that either possess or do not possess some single property P. We might, for example, be considering a population of three ravens each of which might be black (have the property P) or white (fail to have the property P). We want to know what the probability is that the sentence 'Two of the three ravens are black' is true. What are the possible worlds here? Well, we might say that the proportions of black to non-black ravens are what supply the equal possibilities—that is, one possible world is the one in which all three are black, another possible world is one in which two (*any* two) out of three are black, and so on. It is clear that relative to this account of the possibilities (the possibilities being given here by what Carnap called *structure descriptions*) the sentence 'Two out of three ravens are black' holds in one out of four

possible cases (0 black ravens, 1 black raven, … , 3 black ravens) and so on this account its probability is $\frac{1}{4}$.

On the other hand, we could of course individuate each of the ravens (as their mothers no doubt do) and count 'possible worlds' by looking at each possible *distribution* of the properties black and non-black over all the individual ravens. Given this construal, there is, as before, one 'possible world' in which all the ravens are black, but, unlike before, there are not one, but three, possible worlds in which two ravens out of the three are black (the one in which raven_1 is white and the others black, the one in which raven_2 is white and the others black, and the one in which raven_3 is white and the others black). On this account, using Carnap's terminology, each possible world is characterized by a particular *state description* of the form

$$\pm Pa_1 \,\&\, \pm Pa_2 \,\&\, \pm Pa_3;$$

that is, by any of the threefold conjunctions which either asserts (+) or denies (−) that P holds of each of the a_i. Relative to this reckoning, there are a total of 2^3, i.e. eight, possibilities rather than four, and our particular sentence ('Two of the three ravens are black') holds in three of them. So this way of marking out the possibilities would give that sentence the probability $\frac{3}{8}$ rather than the probability $\frac{1}{4}$.

It is unclear how it could be argued that one of these two ways of counting possibilities is the correct one and the other incorrect. (If you are inclined towards the state description account as somehow clearly the more natural, you might like to reflect that this way of counting possibilities does not permit 'inductive learning'. That is, suppose you have already observed two of the ravens in our three-bird example, and both have turned out to be black. Still the probability, as arrived at through state descriptions, that the third one will also be black given this evidence is just the same (namely, one-half) as it was before any observations were made. This is because there are two state descriptions in which the first two ravens are black and in one of them the third is also black, while in the other the third raven is white. This may not seen *too* counter-intuitive, but consider a universe in which there are 1,000 birds and suppose you have observed the first 999 and they were all black; still on the state description measure, the probability that the 1,000th raven is black would be the same as the probability that the 1,000th bird is white, i.e. $\frac{1}{2}$. Of all the $2^{1,000}$ state descriptions in this case, there are again just two in which the first 999 are black, and in one of those the final one is black, while in the other it is white.) Moreover, the problems for Carnap's approach become even more intense when—as will of course standardly be the case in anything like a realistic scientific example—the universe is infinite. (A clear account of these difficulties can be found in Howson and Urbach 1993, chapter 4.)

Although the Carnap programme is long dead, I have outlined it here

because its failure sets the agenda for recent concerns. I have no doubt that some readers, when introduced to the more recent probabilistic approach to confirmation of theories, will feel that this approach is 'not objective enough'. It is as well to remember, however, that the fully objective probabilistic account has apparently unambiguously failed.

1.3.3. *The Basics of Personalist Bayesianism*

Can probability theory still be used to characterize correct scientific reasoning despite the failure of the Carnap programme? The influential 'Bayesian' school argues that it can.

A Bayesian 'agent' is thought of as having 'degrees of belief' in all the propositions expressible in her language (hence including any truth-functional combination or any logical consequence of any such propositions). Such an agent is 'rational' (scientific) if and only if

(*a*) at any given time, her degrees of belief satisfy the probability calculus (i.e. are formally representable as probabilities); and
(*b*) she *modifies* her beliefs from one time (t_1) to the next (t_2) in the light of the evidence that has accumulated between t_1 and t_2 in a certain way—via the 'principle of conditionalization'.

Concerning (*b*): suppose that all that has happened of epistemic relevance to theory T between time t_1 and time t_2 is that some statement e which was not known to hold at t_1 has now (at time t_2) been accepted (by the agent) as evidence (this is admittedly a rather vague supposition because of the clause about the acquisition of e as evidence being all that happened of epistemic relevance between the two times); the principle of conditionalization then requires that the agent's degree of belief in T at t_2 should be the same as her degree of belief in T at time t_1 conditional on e, that is,

$$p_{t_2}(T) = p_{t_1}(T \mid e).$$

This is sometimes expressed as the requirement that a rational agent's 'posterior' degree of belief in T after evidence e has accrued should be her 'prior' degree of belief of T conditional on e—that is, prior to e's having become evidence.

The approach is called 'Bayesian' simply because probabilities of the form $p(T \mid e)$ are crucial quantities for the approach and these probabilities are evaluated using Bayes's theorem. This theorem (which is a theorem of pure probability theory and hence entirely uncontroversial) in its simplest form states that, provided $p(e) \neq 0$, then:

$$p(T \mid e) = p(e \mid T) \,.\, (p(T)/p(e)).$$

The approach is called 'personalist' because it explicitly eschews Carnap's idea that there is only one correct, 'objective' value of the *prior* probabilities $p(T)$ and $p(e)$ occurring in this expression.

Although this freedom over the priors entails that two equally Bayesian-rational agents may have very different degrees of belief in the same theory in the light of the same evidence, it by no means follows that the approach is entirely subjective, that the approach simply describes the way different people in fact think and reason. Each of the requirements (*a*) and (*b*) imposes objective constraints on degrees of belief that may well not be met by particular real reasoners (indeed some of the constraints provably cannot be met by real reasoners).

For example, since the probability calculus is based on classical logic, the Bayesian requires that an agent is rational only if she is a perfect deductive logician—believing every logical truth absolutely, i.e. to degree 1. (Of course, even the smartest of you would be hard pressed actually to recognize that some complicated sentence involving, say, twenty-seven atomic propositions is in fact a tautology; hence you might not know that your 'real' degree of belief in it, as a Bayesian-rational agent, was in fact one.) Again, Bayesianism dictates that an agent is rational only if her degree of belief in the proposition that 'either the sun will rise tomorrow or blow up overnight'—assuming that she discounts entirely the possibility that it will *both* blow up *and* rise tomorrow—must be the sum of her degrees of belief in the proposition that the sun will rise tomorrow and in the proposition that the sun will explode overnight.

1.3.4. *Bayesianism and Rationality: The Two Main Questions*

The two constraints on 'rational degrees of belief' underlying Bayesianism have, then, some objective bite. The question arises whether an agent needs to satisfy those principles in order to count as rational. That is, would an agent automatically be irrational if her degrees of belief did not satisfy those constraints? A positive answer to this question clearly requires justification of the two Bayesian constraints as necessary conditions for rationality.

Discussion of the first constraint—that an agent's degrees of belief at any given time must be probabilities—has centred around the so-called *Dutch book* argument. The idea is that if an agent's degrees of belief fail to satisfy the probability calculus and hence fail to be probabilities, then she is provably committed to regarding certain bets as fair on which she is bound to lose whatever the real world turns out be like. Extensions of the Dutch book argument have also been developed in an attempt to show that an agent must conditionalize in order to avoid irrationality; while another argument for the principle of conditionalization has been that it is essentially analytic—the fact that your degree of belief in T conditional on e is r simply *means* that your degree of belief in T

would be *r* if you were to come to know *e* (but everything else remained the same). There are, however, arguments in the literature against both suggestions and some philosophers regard conditionalization as both substantive and unjustified. References to the interesting literature on these issues can be found in the bibliography.

The second main question about these Bayesian constraints is philosophically more challenging, if less clear-cut, and concerns their *sufficiency*: is an agent whose degrees of belief satisfy the two Bayesian conditions automatically rational (in the intuitive scientific sense)? This seems to come down to the question whether those judgements about correct and incorrect reasoning in science supplied by educated intuition are captured by the Bayesian account. (Of course there is also the possibility that the Bayesian may argue that some particular clash between his principles and educated intuition shows that intuition needs to be *re*-educated.)

There are certain straightforward aspects of scientific reasoning that seem entirely uncontroversial and of which the Bayesian position gives pleasingly neat explications. For example, *theories can be confirmed by observational or experimental evidence and are generally better confirmed by passing 'severe tests'*—that is, by tests the observed outcome of which was implied by the theory but was unlikely in the light of 'background knowledge'. To take a famous example from the history of optics, Fresnel's wave theory of light was confirmed by the discovery that the centre of the shadow of a small opaque disc held in light diverging from a point-source is illuminated. The claim that the centre of the 'geometrical shadow' should be illuminated had been shown to be a consequence of Fresnel's theory (it was Poisson who discovered this logical fact), yet the claim seemed extremely unlikely to be true (so much so, according to the standard story, that Fresnel's contemporaries firmly expected the theory to suffer a major defeat here). In fact, however, when the experiment was performed—by Fresnel himself and his friend Arago—it turned out as predicted by the theory.

The Bayesian account of the evidential confirmation of a theory is simplicity itself: evidence confirms a theory if (and to the extent that) it increases the theory's probability; if, that is, the probability of the theory once that evidence has been established is higher than its probability beforehand. Since the Bayesian principle of conditionalization requires that an agent's 'posterior' degree of belief in theory *T*—that is, her degree of belief in *T* once some piece of evidence *e* has been established (but nothing else of epistemic relevance has happened)—should be her earlier degree of belief in *T* conditional on *e*, that is, $p(T|e)$, and since $p(T)$ measures her ('absolute') degree of belief in *T* before *e* became established, then for a Bayesian

e confirms *T* if and only if $p(T|e) > p(T)$.

(Or better, since it is as well to remind ourselves that these are all *personal* probabilities, an agent will see *e* as confirming *T* just in case, for her, $p(T|e)$ is higher than $p(T)$.) By Bayes's theorem, assuming $p(e) \neq 0$,

$$p(T|e) = p(e|T).p(T)/p(e).$$

Since in cases of the kind now under consideration, *e* is entailed by *T* in conjunction with accepted initial conditions and auxiliaries, it seems reasonable to take $p(e|T) = 1$ (indeed if these accepted initial conditions and auxiliaries are regarded as part of background knowledge, then the Bayesian agent is bound to set the so-called likelihood term at 1). Hence $p(T|e)$ simplifies to $p(T)/p(e)$. This means, in turn, that unless $p(e) = 1$ (that is, unless the agent was completely sure about the outcome of the experiment ahead of time), then *e* must confirm *T* on the Bayesian account and moreover the extent of the confirmation—measured by the difference between the posterior degree of belief in *T*, $p(T|e)$, and the prior, $p(T)$—is greater the smaller is $p(e)$. (Given that $p(e) < 1$, $1/p(e)$ of course becomes larger as $p(e) \to 0$.)

Does Bayesianism do as well when it comes to other—perhaps rather more subtle and taxing—aspects of reasoning in science? When discussing Kuhn's work earlier, we found that the two most pressing problems that that work poses for the defender of the idea that theory change in science is a 'rational' process were the Duhem problem and the problem of prediction versus accommodation. (Concerning this second problem, the issue arose, remember, of whether a piece of evidence that has been 'shoved' into the 'box' provided by a paradigm might weigh less heavily in that paradigm's favour than evidence that 'falls out' of the paradigm without needing to be shoved.) What can Bayesianism tell us about these two problems?

1.3.5. *Bayesianism and the Duhem Problem*

Duhem pointed out that no assertion that we would naturally think of as a 'single' scientific theory—Newton's theory (of mechanics plus gravitation), the classical wave theory of light, special relativity theory, or whatever—has deductive consequences *of its own* that are directly checkable at the empirical level. In order to deduce from Newton's theory, say, predictions about planetary positions (supposing we take such predictions as directly empirically checkable) we need further independent assumptions—about what other massive bodies exist in the solar system, about the amount of refraction that light undergoes in passing into the earth's atmosphere, and so on. Should the observational consequence be found false, then all that follows deductively is that at least one of the premisses—that is, at least one from the set of assumptions consisting of the 'central' theory under test together with all the background or auxiliary assumptions—must be false. (Let *T* be the central theory and *A* the conjunction

of the necessary auxiliaries, then if neither T nor A alone but only the conjunction $T\&A$ entails some observational or experimental consequence e, then all we can infer from the observation that e is false is $-(T\&A)$, which is of course equivalent to $-T\vee -A$.) Indeed if the 'central theory under test' itself divides naturally into a 'core' assumption and a set of more specific assumptions (as does 'the' wave theory of light, for example, which consists of the core assumption that light is some sort of wave in some sort of medium together with more specific assumptions about the sorts of wave and the sort of medium), then the situation is one step more complicated. Now the natural formalization of the experimental test involves C (the 'core' assumption), S (the set of still 'central', but more specific, assumptions), and the auxiliaries A. But then again if only the conjunction $C\&S\&A$ entails some evidential statement e, then $-e$ entails only $-(C\&S\&A)$, that is, $-C\vee -S\vee -A$. And, so far as deductive considerations alone go, the scientist now has three options: (1) retain the whole central theory ($C\&S$) and reject an auxiliary; (2) accept the auxiliaries but reject one of the specific assumptions rather than the core assumption (this option will naturally, if rather confusingly, be described as 'modifying' the central theory rather than rejecting it), and finally, (3) reject the core theory.

Scientists, however, do not invariably exploit the freedom left to them by these purely deductive considerations. Sometimes it seems clearly right to reject an auxiliary in the light of an experimental 'anomaly'. A classic and much discussed case concerns the discovery of the planet Neptune. Here there were 'unexplained irregularities' in the observed orbit of the planet Uranus—that is, Newton's theory of mechanics and gravitation, taken together with then accepted auxiliaries and initial conditions, led to predictions about Uranus' orbit that were observably incorrect. Rather than reject the 'central' Newtonian theory, Adams in England and, independently, Leverrier in France suggested that the 'mistake' lay in the auxiliary assumption about the number and masses of other bodies in the universe that had significant gravitational effects on Uranus. Working backwards from the assumption that Newton's theory is true, they were led to postulate the existence of a hitherto unsuspected planet beyond Uranus—a postulate that was subsequently confirmed by astronomical observation. Here the retention of the central theory in the light of an anomaly seems eminently scientifically justified (to the extent that this episode is always counted as one of the great successes of the Newtonian theory).

But defending a cherished theory against apparent counter-evidence by rejecting some less central, specific assumption sometimes seems to be not a great scientific success but the hallmark of *pseudo*-science. Velikovsky's work provides a clear example. The central theory concerned in this case was Velikovsky's claim that a comet had long ago broken away from Jupiter and orbited the earth on a series of occasions producing such notable events as the parting of the Red Sea and the fall of the walls of Jericho. The anomalous evi-

dence was the absence of any records of similar cataclysms to those recorded in the Old Testament in the archives of at least some other record-keeping cultures of the time. Velikovsky pointed out in effect that this is an instance of the Duhem problem: the assertion that records of appropriate cataclysms would have been kept in the cultures concerned can be deduced only by postulating not merely his fundamental cometary hypothesis but also various auxiliary assumptions—for example, the assumption that the culture's scribes would have recorded any cataclysm on the relevant scale if they had witnessed it. (You would indeed think that 'events' on the scale of the parting of the Red Sea would be worth a line or two in anyone's diary.) Just like the Newtonians working back from the assumption of the truth of their central theory, Velikovsky assumed that his comet really did exist and really did perform the complicated dance he postulated, and so he was led to assume that cataclysms had been witnessed in the other cultures and that in fact they had proved so traumatic that 'collective amnesia' had set in. (Of course, Velikovsky postulated that this dread condition had afflicted those and only those cultures who were otherwise keeping records but had no records of appropriate cataclysms.)

It is surely a requirement on any adequate account of scientific rationality that it explain the difference between these two episodes—that it explain Adams and Leverrier's move, but not Velikovsky's, as scientifically reasonable. (Recall that an account that satisfies this condition will avoid many of the problems posed by Kuhn.) Does personalist Bayesianism satisfy the condition?

Colin Howson and Peter Urbach, developing an earlier idea of Jon Dorling's, claim that a straightforward analysis based on personalist Bayesian principles solves the Duhem problem entirely. Consider a case in which some theoretical system can be represented as the conjunction $T\&A$ and in which neither T nor A alone entails any observation statement; and suppose that some experimental consequence of the conjunction has turned out to be false. The situation with respect to deductive logic is symmetric: neither T nor A alone is falsified. But, as Dorling points out, there is no reason why in such cases the effect on T and A should always be symmetric in Bayesian probabilistic terms. The 'posterior probabilities' of T and A are, assuming $p(e) \neq 0$,

$$p(T\,|\,e) = p(T).p(e\,|\,T)/p(e), \text{ and}$$
$$p(A\,|\,e) = p(A).p(e\,|\,A)/p(e).$$

Plainly, if an agent judges $p(e\,|\,A)$ to be very close to $p(e)$, while she judges $p(e\,|\,T)$ to be very much less than $p(e)$, then her posterior probability for A will be very close to its prior while T's posterior will be very much lower than its prior.

Dorling has analysed various historical cases and Howson and Urbach give an especially clear treatment of Prout's hypothesis (that the atomic weights of all elements are integer multiples of the atomic weight of chlorine) facing experimental evidence that seemed to show the element chlorine to have an

atomic weight around 35.8. But the idea behind the analysis can be readily illustrated in the case we have cited of Newton's theory and the difficulties with Uranus.

Adams and Leverrier 'held onto' Newton's theory despite the 'recalcitrant' observations of Uranus' orbit, and instead 'laid the blame' on the auxiliary about the number of other planets in the solar system. Let T then be Newton's theory, A the necessary auxiliaries, and e the evidence about Uranus' orbit. Suppose that Adams and Leverrier's belief-states can reasonably be represented (of course in an idealized way) as follows:

(1) $p(A \mid T) = p(T)$ (i.e. T and A are probabilistically independent)
(2) $p(T) = 0.9$
(3) $p(A) = 0.6$
(4) $p(e \mid -T \& A) = p(e \mid -T \& -A) = \frac{1}{2}p(e \mid T \& -A)$.

(The probability $p(e \mid T \& A)$ is of course 0 here since e refutes the conjunction $T \& A$.) That is, these scientists start out having a higher degree of belief in Newton's theory than in the auxiliaries (though these are accepted auxiliaries so their degree of belief in them is none the less substantial); and they think that the observed orbit of Uranus is twice as likely to occur if Newton's theory holds but the auxiliaries are wrong, than, for example, if Newton's theory is false and the auxiliaries are correct.

It follows just by the pure mathematics of the probability calculus that

$$p(T \mid e) = 0.878 \text{ while } p(A \mid e) = 0.073.$$

So a Bayesian agent beginning with the above priors and likelihoods, and conditionalizing on the evidence, would find the credibility of Newton's theory scarcely affected by the 'anomalous' evidence from Uranus, while she would now regard the auxiliaries A, which she had initially regarded as more likely to be true than not ($p(A) = 0.6$), as scarcely credible at all. And the claim is that this analysis reveals the rationale for the markedly asymmetric reaction to this 'anomalous' evidence by Leverrier and Adams: in finding their 'central' (Newtonian) theory scarcely threatened and in 'blaming' the auxiliaries, they were, in effect, operating as rational Bayesian agents.

It is of course by no means clear what would count as an acceptable rational reconstruction of the 'belief dynamics' of some set of scientists. Not even the most committed Bayesian would claim that such accounts are simply descriptive—since no one seriously holds that the historical agents had degrees of belief that are exactly expressed by the numbers used in such Bayesian analyses. No Bayesian to my knowledge has explicitly analysed the Adams–Leverrier case; the numbers used above are taken from Howson and Urbach's treatment of Prout. But similar difficulties seem to arise in all cases. In particular it seems rather difficult to take the likelihood assumptions ((4) above) seriously. Even

accepting that no one expects precise numbers, I suspect that one would have got little more than blank stares from Adams and Leverrier if one had asked them what their degree of belief in the evidence about Uranus would be on the assumption that both Newton's theory, and the then accepted auxiliaries, are false.

However, the chief difficulty with this analysis arises once we concede, for the sake of argument, that the above account in terms of personal probabilities does provide a rational reconstruction of the reasoning of Adams and Leverrier. What we set out to find as a solution of the Duhem problem was not simply a reconstruction of the (intuitively correct) reasoning of Adams and Leverrier, but a reconstruction that differentiated that reasoning from the superficially similar, but intuitively highly suspect, reasoning of Velikovsky and followers. The Bayesian analysis fails signally to underwrite this distinction. Nothing is easier than to model Velikovsky's reasoning in Bayesian terms. We just have to make formally the same assumptions about his prior beliefs as in the Adams and Leverrier case. Let T' be Velikovsky's central cometary hypothesis and A' the necessary auxiliaries (including assumptions about the reliability of scribes in various cultures). Then, if his beliefs about these various claims ahead of the evidence (e') of the lack of suitable records in certain cultures are appropriately expressed by

(1′) $p(A' \mid T') = p(T')$ (i.e. T' and A' are probabilistically independent)
(2′) $p(T') = 0.9$
(3′) $p(A') = 0.6$
(4′) $p(e' \mid -T' \& A') = p(e' \mid -T' \& -A') = \frac{1}{2}p(e' \mid T' \& -A')$,

it plainly follows just as before that

$$p(T' \mid e) = 0.878 \text{ while } p(A' \mid e) = 0.073.$$

Hence, as a good Bayesian, Velikovsky will have conditionalized, and so his 'posterior' degree of belief in his central hypothesis will be hardly different from his prior degree of belief, while his degree of belief in the auxiliaries will have been radically reduced. And this indeed captures the way Velikovsky seems to have reacted to this evidence.

Now the Bayesian does not of course condone simply plucking alleged degrees of belief out of the air in order to defend a view favoured for some other (non-epistemic) reason. It may well be that, although Velikovsky could have defended his position as rational on Bayesian principles if he had had the degrees of belief involved, as a matter of fact he did not have (or rather cannot sensibly be idealized as having) the priors and likelihoods cited in (1′–4′). It may well be that he instead made a series of conditionalization errors. But the real Velikovsky's beliefs seem irrelevant to the philosophical point. The Bayesian account seems to be put into deep trouble just by the fact that, according to it,

were someone's belief-state to be one that could be appropriately modelled via assumptions (1′) to (4′) then they would be declared perfectly rational in seeing the evidence of no cataclysmic records as barely affecting the credibility of the central cometary hypothesis.

The natural reaction here is of course that some of the priors and likelihoods in (1′–4′) themselves seem intuitively highly suspect. Is it really reasonable, for example, to have a degree of belief as high as 0.9 in Velikovsky's theory of the earth's close encounters with this strange 'comet' ahead of the evidence of no cataclysmic records in some culture *C*? Or is it really reasonable to hold that that evidence is twice as likely on the assumption that Velikovsky's theory is true but that scribes in *C* were inaccurate as it is on the assumption of no close encounters and accurate scribes? However, such questions are *entirely out of place* in the Bayesian scheme which treats such assignments of degrees of belief as 'givens' in the analysis. If you hold that there is an objective difference between the reasoning of, say, Adams and Leverrier and Velikovsky in that the first reasoned scientifically and the second not, then you cannot hold that personalist Bayesian analysis is an accurate and complete account of objective, scientific reasoning.

The criticism that personalist Bayesian analyses are *too* personalist, too dependent on facts about agents' priors, to give an adequate account of the principles underlying correct reasoning in science is one that has often been levelled. There have of course been Bayesian attempts to counter the charge. I shall briefly outline these attempts shortly—after first looking at the Bayesian treatment of our second important methodological issue—that of whether successful predictions are more confirmatory for the theories that make them than are successful accommodations of known facts.

1.3.6. *Bayesianism and 'Prediction versus Accommodation'*

Kuhn claimed, remember, that one could not fault the hold-outs to revolutions—their belief that some explanation could be given within their paradigm for the allegedly crucial new evidence is demonstrably 'neither illogical nor unscientific'. One straightforward way in which an account of empirical support for scientific theories might counter this claim would be by underwriting a distinction between evidence that was predicted by a theory and evidence that was already known and 'merely' explained by a theory (or accommodated within the theory). Suppose it could be argued that the correct account of empirical support gives, *ceteris paribus*, greater confirmatory weight to a newly predicted fact than to an explained old fact. Suppose a 'revolutionary' new paradigm predicts some hitherto unsuspected phenomenon. Kuhn's hold-outs' belief that the evidential scales could be balanced simply by giving an account of that phenomenon within their old paradigm would then be mistaken—they

would, contrary to Kuhn's own claim, be shown to have been unscientific in the sense that they did not weigh evidence according to the correct principles.

So, for example, phlogistonists insisted that the fact that the residue of mercury burned in air weighed more than the initial mercury could be accommodated within their paradigm, despite the fact that it was committed to the idea that something (to wit, phlogiston) was always emitted from substances that burned. They were clearly right: one possibility (that does not in fact seem ever to have been taken seriously) would be to attribute phlogiston 'negative weight' (whatever that might mean); a more plausible possibility would be to hypothesize that the mercury *both* loses phlogiston *and* gains something else, a complex process that happens to result in a net weight increase. But if this were merely an accommodation (that is, if no such suggestion made independently testable predictions) and if accommodations count less than predictions (putting it very roughly) then the hold-outs for phlogiston would indeed be reasoning unscientifically if they counted this as balancing the evidential scales (Lavoisier's oxygen theory having, remember, predicted the result of the mercury experiment).

But does the correct account of empirical support lend a confirmatory premium to new evidence? And, if so, why? This is a long-running issue in philosophy of science. Keynes and Mill are notable representatives of the anti-predictivist side. Keynes wrote (in his *A Treatise on Probability*):

> [the] peculiar value of prediction . . . is altogether imaginary. . . . The question of whether a particular hypothesis happens to be propounded before or after examination of [its empirical consequences] is quite irrelevant.

This echoes John Stuart Mill, who wrote (in his *System of Logic*):

> it seems to be thought that an hypothesis . . . is entitled to a more favourable reception, if, besides accounting for all the facts previously known, it has led to the anticipation and prediction of others which experiment afterwards verified . . . Such predictions and their fulfilment are, indeed, well calculated to impress the ignorant vulgar, whose faith in science rests solely on similar coincidences between its prophecies and what comes to pass. But it is strange that any considerable stress should be laid upon such a coincidence by persons of scientific attainments.

The 'person of scientific attainments' who held this strange view and whom Mill chiefly had in mind was William Whewell. Whewell asserted that successful prediction 'gives a theory a stamp of truth beyond the power of ingenuity to counterfeit'. Similar views were expressed, for example, by Duhem, who held that the 'highest test of [a theory] is to ask it to indicate in advance things which the future alone will reveal' and that if the theory passes such a test, it is especially highly confirmed. (In Duhem's own terms this means that it is likely to be part of a 'natural classification'.) And subsequently by many other philosophers of science.

What does Bayesianism have to say about this contentious issue? The answer is interestingly mixed.

1.3.6.1. *Old Evidence never Confirms (and that's Wrong)* Some analysts—for example, Clark Glymour and, following him, John Earman—have argued that

1. the Bayesian system entails that *only* new evidence ever confirms a theory; and
2. this is completely contrary to intuitively sensible judgements made by scientists.

Glymour holds, in other words, that the answer that would be entailed by the correct account of evidential support is that there is no premium on predictions; and, since he also believes he can demonstrate that Bayesianism entails that only successful predictions count in favour of a theory, it follows that Bayesianism is not the correct account of evidential support.

The issue here has become known as the 'problem of old evidence'. Here is how Glymour argues that Bayesianism entails that only new evidence confirms a theory. If e is evidence at time t, that is, e is already known to hold, then it is part of 'background knowledge' at t, and—remembering that all probabilities in the Bayesian scheme are relative to background knowledge—this means that

$$p_t(e) = 1.$$

It is easy to prove that, if so, then no such piece of known evidence e can confirm any theory. As we saw earlier, e confirms T only if $p(T|e) > p(T)$. But, by Bayes's theorem, $p(T|e) = (p(e|T).p(T))/p(e)$. And so if for any old evidence, e $p_t(e) = 1$ (which entails $p_t(e|T) = 1$), it follows that $p_t(T|e) = p_t(T)$. Thus, by conditionalization, once you take e into account, your degree of belief in T is just the same as it was before—e is entirely neutral with respect to T, neither confirming nor disconfirming it.

Glymour claims that, to the contrary, there is a whole host of cases from history of science where a theory was not only positively confirmed, but strongly confirmed by some evidence e that was already well known. The case that Glymour especially emphasizes is of general relativity theory and the observations of the precession of Mercury's perihelion. The evidence about Mercury was already known when the general theory was formulated, but it provided, in the view of the scientific community, especially compelling evidence for Einstein's theory—more compelling, according to most people's intuitions, than that from newly discovered phenomena such as the Mössbauer effect.

1.3.6.2. *The Garber, Niiniluoto, Jeffrey Attempt to Solve the Old Evidence Problem* Glymour himself suggested a way in which the Bayesian system might be extended to overcome this problem of old evidence. This suggestion was subsequently endorsed and developed by a number of philosophers including Garber, Niiniluoto, and Jeffrey. The suggestion is that the important fact that was unknown in cases like that of general relativity theory and the precession of Mercury's perihelion was not the empirical evidence about Mercury's motion but rather the logical fact that the theory entails the (known) evidence about Mercury's motion. The confirmation arises from the *logical discovery* that the theory entails the evidence.

In order to make this account work a new primitive connective $\Rightarrow$ is introduced. This represents (though one is not formally allowed to recognize it) deductive entailment. It can then be shown that if

(1) an agent has degrees of belief that are probabilities not just in the scientific theories and statements available to her but also in all statements of the form $T \Rightarrow e$, $T_1 \Rightarrow T_2$, and so on; and if
(2) the agent's probability distribution satisfies certain assumptions,

then for some T, e pairs,

$$p(T|T \Rightarrow e) > p(T).$$

Hence although Glymour is right that old evidence in itself cannot Bayesian-confirm any theory, the discovery of the logical fact that the theory entails that old evidence (that is, the discovery that $T \Rightarrow e$) may confirm the theory.

This ingenious suggestion faces certain problems, however. First of all, since a Bayesian agent is supposed to be a perfect (deductive) logician, it is only in virtue of forgetting that $\Rightarrow$ really means deductive entailment and taking it as an undefined primitive that formal contradiction is avoided. And secondly, the intuitive justification seems, on reflection, very doubtful—surely what really confirms the general theory of relativity is empirical evidence about Mercury and not a logical (and therefore analytic) fact about the relationship between the general theory and sentences describing Mercury's motion.

1.3.6.3. *Howson and Urbach's Attempt to Solve the 'Old Evidence Problem'* Some Bayesians—notably Colin Howson and Peter Urbach—have argued that Glymour has it topsy-turvy: it is indeed correct intuitively that old evidence can confirm scientific theories, but the Bayesian position entails exactly this intuitively correct result. The crucial claim behind this particular Bayesian analysis amounts to the following:

> Probabilities are all implicitly relative to background knowledge; but when the impact of evidence e on some theory T is being weighed, the right probability for e is the degree of belief you *would have had* in e, were you in the cognitive situation you are in fact in, except that you did not know e.

When e is (or was) unknown, this 'correct degree of belief' is just your actual degree of belief at the instant before e becomes evidence for you. However, where e is already known, the relevant degree of belief has a counterfactual character: you pretend that your relevant 'background knowledge' is not B (as it in fact is), but rather, so to speak, '$B - \{e\}$'—what remains when you 'subtract' e from background knowledge B. It follows that $p(e)$ for known e need not be 1, and indeed in general will not be. And so, of course, contrary to Glymour, old evidence may Bayesian-confirm.

Several problems make this suggestion highly problematic. One is the formal problem that the set of sentences '$B - \{e\}$' is not well-defined. Simply 'deleting' e from the (deductively closed) set of B's consequences has no real effect, since e will simply 'reappear' as a consequence of lots of other consequences of B. (Let f be any other consequence of B, then the sentence $e \& f$ will also be a consequence of B.) On the other hand, suppose B is axiomatized in some way and we characterize $B - \{e\}$ as the set of consequences of the axioms with e removed. Since evidential statements like e will not normally be axioms, this operation generally leaves B entirely unaffected. However, suppose $\{B_1 \ldots B_n\}$ are axioms for B, where none of the B_i is the evidential statement e; since e is a consequence of B it is easy to see that the following axiomatization is logically equivalent to the first:

$$\{e, e \rightarrow B_1, e \rightarrow B_2, \ldots, e \rightarrow B_n\}.$$

But now the effect of 'deleting' e from this axiom set will be devastating (the remaining axioms all being relatively weak statements equivalent to sentences of the form 'either not e or B_i').

Howson and Urbach argue that despite these formal difficulties we make good enough intuitive sense of the required idea in many cases. Suppose, for example, a coin has been tossed a single time, and has landed 'heads', we can none the less easily understand the claim that the coin had a probability of one-half of producing tails *on that toss*. But are we not, then, in effect saying 'suppose my background knowledge were as it is except that I didn't know that the coin had turned up heads on this toss, then I would give a probability of one-half to each of the possible outcomes'?

But, whatever may be the case with the coin, it is not at all clear—as John Earman has also suggested—that sense can be made, for example, of what Einstein's cognitive situation would have been in 1915 had he not known about the precession of Mercury's perihelion. Einstein's whole view was so multiply affected by the Mercury anomaly, that it seems hard even to begin to think about what he would have believed had he not known about it. If so, then it is impossible to make any real sense of his degrees of belief in other propositions against this counterfactual background. Similarly, having studied Fresnel's thought and work for many years, I just have no conception of what it might

mean to talk of Fresnel's 'background knowledge' *as it would have been* had he not known about, say, the phenomenon of straight-edge diffraction. If this is right, then at least for these particular pieces of known evidence e, the Howson and Urbach counterfactual construal of $p(e)$ fails.

Suppose, however, we go along with this counterfactual suggestion for a while, does it lead to a satisfactory resolution of the predictivism debate? The Bayesian analysis now delivers the (surely correct) implication that known evidence does sometimes confirm. An investigator in 1915, somehow applying the counterfactual interpretation, assigns a probability of (let's assume, considerably) less than one to the statement that Mercury's perihelion precesses in the way it does. Given that the general theory of relativity, R, entails that statement e, then $p(R \mid e) = p(R)/p(e) >> p(R)$. Hence that investigator, having conditionalized as a good Bayesian, will see R as (strongly) confirmed by e.

However, as is now well known, a classical, non-relativistic account can also be given of Mercury's movements by making entirely *ad hoc* assumptions about the density distribution of the sun (that is, those assumptions about the density are fixed exactly so as to yield the known facts about Mercury and for no other reason). This possibility became known as a result of Dicke's work in the 1960s. Call this classical account C. C also entails e and so again an agent applying the counterfactual construal—say in 1960—will set $p(e) << 1$ and hence will derive $p(C \mid e) >> p(C)$. That is, such an agent will see e as confirming C as well. This in itself need not be problematic: after all we are surely happier with C given that it gets Mercury's motion right than we would have been if (*per impossibile* given the way it was constructed) it had got Mercury's motion wrong. And indeed so long as an agent attributes a very low prior probability (that is, ahead of e) to C, much lower than the prior she assigns to R, then the Bayesian analysis delivers the intuitively correct judgement (or at any rate the one that is firmly entrenched in the scientific community)—that R remains very much more probable than C once the evidence e is taken into account. Indeed, if we measure the support that e lends to theory T by the 'difference measure' ($S(T,e) = p(T \mid e) - p(T)$), then we have, since R entails e and therefore $p(R \mid e) = p(R)/p(e)$,

$$S(R,e) = p(R \mid e) - p(R) = p(R)(1/p(e) - 1).$$

Moreover, since C also entails e,

$$S(C,e) = p(C \mid e) - p(C) = p(C)(1/p(e) - 1).$$

Assuming that $p(e)$ has the same value in both cases and that that value is not one, we obtain the result that

$S(R,e) > S(C,e)$ (that is, that e supports R more than it supports C) if and only if $p(R) > p(C)$.

Or, to put the result in its more correct personalist terms: an agent will see *e* as supporting R more strongly than *C* just in case her prior for R is higher than her prior for *C*.

A similar analysis will show that, for example, some piece of evidence (say, some aspect of the fossil record) entailed both by some specific Darwinian theory and by some specific special creationist theory (in the latter case simply by dint of writing the—now alleged—fossil into God's creation) differentially supports the two theories in a way that depends entirely on their prior probabilities ahead of that evidence.

Such a Bayesian agent will make evidential judgements in these cases entirely in accord with the judgements made less formally by the scientific community exactly if she assigns a much higher prior to R than to *C* and a much higher prior to the Darwinian than to the special creationist theory. But what if she is firmly convinced of *C* or of special creation and assigns them much higher priors than their rivals? Then as a good Bayesian she will see the evidence as providing in a sense further support for her views, for she will see her previously favoured theories as the better supported by the evidence at issue. Analogously to the case of the Duhem problem: if you hold that, say, the fossil record should, on the principles of correct scientific reasoning, be seen as favouring Darwinism over special creation whatever one's initial beliefs, then you cannot hold that personalist Bayesianism is a correct and complete account of those principles.

1.3.7. *Rejoinders to the Charge of Over-Subjectivism*

Bayesians often respond to the charge that their accounts of scientific reasoning are too reliant on merely subjective priors by citing a range of results about the 'swamping' or 'washing-out' of priors. After first advertising the virtue of having a certain amount of 'subjectivism' in one's account of proper scientific reasoning (even in science it would sometimes be unfortunate if all reasoners held the same beliefs), Bayesians then go on to claim that this subjective element does not matter too much because in many situations the subjective element is eventually overwhelmed. In certain situations and subject to certain weak constraints on the priors, all Bayesian agents will tend towards agreement.

For all their formal interest, it is not clear how much philosophical weight these swamping results carry. Priors are only ever fully 'washed out' in the limit, which of course we never really achieve. The fact is that any actual theoretical preference—even the intuitively most bizarre—may be Bayesian rational: provided she began with a sufficiently high prior on her own theory and a sufficiently low prior on the rival evolutionary account, a creationist can have conditionalized away on the accumulating evidence and still have arrived, as of

this moment, at an overwhelmingly higher posterior for her scientific creationist theory than for Darwinian theory. The fact that this probability, though overwhelmingly high, is somewhat lower than her prior seems of little consolation; and there seems equally little consolation in the thought that, given certain kinds of future evidence, the sequence of her successive posteriors and that of her erstwhile opponent are destined to converge. If the judgement that a satisfactory theory of proper scientific reasoning ought to deliver is that creationists, Velikovskians, and the rest of the sorry crew are irrational *now*, then personalist Bayesianism is no such theory.

The second type of Bayesian response is that the charge of over-subjectivism is entirely misplaced. The principles of personalist Bayesianism are essentially natural extensions of logical principles. The requirement of coherence (that is, the requirement that one's degrees of belief satisfy the probability calculus) is a natural extension of the requirement of deductive consistency. (Indeed Ramsey—essentially the founder of the position—used the term consistency to cover both the purely deductive and the probabilistic notions.) Moreover, some Bayesians (as we saw earlier) defend the principle of conditionalization as analytic. The imposition of any further requirement (for example constraints on 'sensible' priors) would, however, definitely transcend logic and hence involve substantive assumptions about the world and our ways of comprehending it. But this would raise the awkward question how such assumptions could themselves have any reasoned credentials. How could a synthetic claim that is allegedly constitutive of reason itself be given a reasoned justification without getting involved in an infinite regress? Personalists like Dorling and Howson conclude that any further requirement would be 'arbitrary' and that Ramsey was therefore right to insist that inductive *logic*—personalist Bayesianism—must treat an agent's prior distribution of degrees of belief as simply a given. If this means—as we have seen it does—that a rational Velikovskian, or a rational scientific creationist becomes a possibility, then this must simply be accepted. If Velikovsky started off with some prior for his cometary hypothesis that seems ludicrous to you, then you will of course disagree with him (this simply means that your prior will be different), but so long as he has conditionalized properly on the accumulating evidence, you have no justification for regarding him as non-scientific or irrational—even though he still, even in view of all the evidence, orders the credibilities of the available theories radically differently from the way you do.

This response can itself be attacked in two ways. First, are the principles underlying personalist Bayesianism really 'logical'? Secondly, even if they are logical, wouldn't it simply follow that logic (even in the extended Bayesian sense) is not strong enough on its own to ground an adequate theory of proper scientific reasoning?

1.3.8. *Prediction need not be* Prediction

It seems that, depending on the details of the Bayesian analysis you favour, a Bayesian can endorse either of the possible answers to the old and new evidence problem. But which answer is intuitively the correct one? Does new evidence provide, *ceteris paribus*, more support and, if so, why? The question is still very much an open one—discussions of it have, however, in my view, been obscured by a failure to make the right distinction.

Why on earth should it matter from the point of view of how strongly it supports a theory exactly when some piece of evidence was discovered to be indeed evidence? Why should the fact that the precession of Mercury's perihelion had been well investigated before Einstein articulated the general theory of relativity while the gravitational starshift was discovered only later in itself matter at all? Perhaps it does seem especially impressive, psychologically speaking, when some theory turns out to make a prediction about the outcome of some experiment that no one ever thought of before, and when that experiment is performed it turns out as predicted by the theory. But if this were an essential principle of the way that science has always evaluated empirical support, then we should, I think, simply have to record this as an amazing fact about scientific 'rationality' which itself has no reasonable justification.

However, there is no need to suppose—in order to explain the theoretical decisions they make—that scientists do give extra theory-confirming weight to some piece of evidence just because it was unknown at the time the theory was articulated. There is, I suggest, no epistemically important distinction here beyond that between evidence that 'falls out naturally' from a theory and evidence that has to be 'shoved' into the theory. The important distinction is not that between old and new facts but that between evidence that has, and evidence that has not, been 'written into' or 'accommodated by' the theory—through, for example, fixing the value of some initially free parameter on the basis of the evidence.

For example, the 'scientific creationist' theory merely accommodates the 'fossil' record; by supposing that the creator simply chose to include in the creation some things that happen to look a lot like bones of animals of extinct species and simply chose to scratch patterns in the rocks that happen to look like the imprints of skeletons of such animals. (Indeed the whole creationsit approach is basically an exercise in accommodation: the fundamental theory says that God created the universe much as it presently is; observations reveal how it presently is and those details are then fed into the fundamental theory to create specific creationist theories that unsurprisingly entail the evidence.) On the other hand, Copernicus, for example, needed to make no special assumption in order to explain planetary stations and retrogressions, which were instead an inevitable consequence of the implication of his theory that those

planets were being viewed from a moving observatory (attached to the moving earth). Hence even though stations and retrogressions were known long before Copernicus formulated his theory, these phenomena were not accommodated by his theory in the epistemically important sense of the term. Similarly, the general theory of relativity has no free parameters that might be adjusted on the basis of facts about Mercury's perihelion advance, and this is why the theory's success in entailing the right orbit for Mercury counted strongly in its favour.

We can indeed mark this distinction as that between a theory's predicting a result and merely accommodating it *post hoc*, but only if we remember that prediction must be understood in a non-temporal sense that allows the prediction of old facts. This may seem a strange usage—but it is in fact one often adopted both in science and studies of science. The Logical Positivist Moritz Schlick, for example, wrote:

> the confirmation of a prediction means nothing else but the corroboration of a formula for those data which were *not used in setting up the formula*. Whether these data had already been observed or whether they were subsequently ascertained makes no difference at all. (emphasis added)

And French's respected textbook *Newtonian Mechanics* remarks that Newton's theory

> like every other good theory in physics had predictive value; that is, *it could be applied to situations beside the ones from which it was deduced*. Investigating the predictions of a theory may involve looking for hitherto unsuspected phenomena, or it may involve recognizing that an already existing phenomenon must fit into the new framework. (emphasis added)

Despite its many virtues (it is widely and understandably regarded as the best account of confirmation we have at present), the Bayesian system faces many problems, some of which have been examined in this section. It may be the best formal treatment of confirmation we have, but, if so, then even the best, it seems, needs extension and improvement.

1.4. Scientific Revolutions and Scientific Realism

1.4.1. *What is Scientific Realism?*

The state of science at any given time is characterized in part by the theories 'accepted' at that time. Presently accepted theories include the general theory of relativity, the quantum theory, various theories about elementary particles, and, for example, the modern 'synthesis' of Darwin and Mendel, as well as 'lower level' but none the less still theoretical claims, such as that chemical

elements have some sort of atomic structure, that electrons are negatively charged, that DNA has a double-helical structure, and so on. These theories talk about, amongst other things, electrons and other elementary particles, a space-time structure with a certain interesting metric, genotypes, species of once living but now extinct animals, and so on.

What exactly is involved in accepting a theory? What should we believe about accepted theories and their associated theoretical terms? The seemingly most straightforward and attractive answers are that (rationally) accepting a theory means (rationally) believing it to be true and hence believing the ontology that the theory postulates is real. Taken at face value, accepted theories (or at any rate many of them) straightforwardly attempt to describe a world of entities 'hidden beneath' the phenomena—entities that function in accordance with certain general laws and as a result produce (in conjunction with the laws governing our own constitution) the phenomena that happen to be manifested to humans. Taken at face value, those theories assert the existence of electrons, spacetime curvatures, genotypes, and the rest. It seems natural to say that 'acceptance' of them implies that it is reasonable to believe that the theories are true and hence that the entities they involve are real (and indeed that no other belief is reasonable).

This amounts to a particularly strong version of 'scientific realism'. Too strong to be sensible: scientific realists do recommend taking accepted theories at face value, but no real realist has any such straightforward view of what is involved in theory acceptance.

First, realist claims are explicitly restricted to accepted theories in 'mature' sciences. Although there is no agreement on a precise characterization of maturity, there is a measure of agreement on individual cases: current physics is definitely mature, and it achieved maturity with Newton at the latest; physics at the time of Aristotle, on the other hand was 'immature', and the same goes for pre-Lavoisierian chemistry, for optics before Newton (or perhaps Fresnel), and for nineteenth-century phrenology, as well as current parapsychology and other cases that may shade off into outright pseudo-science. In all these areas, theories were (or are) accepted in the purely factual, sociological sense of being believed by a group of people, whose research was based on that belief. But the realist will only defend those beliefs as rational when they concern theories from 'mature' sciences. The realist will not, for example, feel any need to defend the phlogiston theory as even approximately true, nor to endorse the entities it postulated—especially phlogiston itself—as real elements of the universe.

Moreover, no relatively sophisticated realist will advocate a uniform attitude towards accepted theories, even within mature sciences. For one thing, not every accepted theory is equally firmly entrenched. A scientist would find it close to incredible that the theory that there are electrons will not be preserved in future science, but would not have the same attitude towards the full quantum theory (indeed the quantum theory is known to require modification, see below). And the attitude of most scientists towards quarks or superstrings, for example, is that they may well exist, that there is a good deal of evidence for them and a good deal that could not be explained without them, but that they none the less retain, for the moment at least, a conjectural quality not shared by electrons. There are, in other words, different grades of 'acceptance': some theories are, at a given time, totally entrenched in that no alternative is being sought or even seriously contemplated; others are accepted in the sense that they are regarded in their field as the only serious contenders so far articulated, as having some degree of support, but as retaining a definitely conjectural character. The realist might, on this account, restrict her realism to accepted theories of the first, very firmly entrenched kind; or, perhaps more plausibly, distinguish between degrees of reasonable belief: the rational belief to have is that all accepted theories have a good probability of being true (or rather, as we shall see, of being 'essentially' or 'approximately true'), the probability reaching, or at any rate approximating, one in the case of the deeply entrenched theories.

A third way in which the notion of acceptance requires refinement is through the recognition that some scientific theories have an *idealizing* character. For some theories (the ideal gas law is the usual example) this is obvious, but many theories reveal elements of idealization, once examined carefully. Newtonian particle mechanics, for example, was certainly a successful theory and was firmly accepted in the eighteenth and nineteenth centuries—yet it was of course recognized that there might be no such thing as a Newtonian particle and that certainly none of the entities to which this theory was (successfully) applied strictly fit that description. The scientific realist is (or should be) in the business of arguing that theoretical science should be taken 'at face value'; but, even at face value, not all accepted scientific theories are intended to be straightforward descriptions of reality; instead, some idealize. Even within a basically realist account, some theoretical notions are to be viewed as connected with reality in rather more complicated ways than through direct reference to real entities.

Finally, the sophisticated realist will realize that 'accepted theoretical science' is invariably internally problematic. For example, it is scarcely controversial that the general theory of relativity and the quantum theory figure in the list of theories currently accepted in science. Yet it is well known that the two theories cannot both be strictly true—not for any philosophical, but for purely scientific, reasons. Basically, quantum theory is not a covariant theory as required by relativity theory; while the fields postulated by general relativity theory are not quantized as required by the quantum theory. It is generally held that a 'synthesis' of the two theories is needed, one that cannot of course (in view of their logical incompatibility) leave both theories fully intact. Quantum field theory

is intended to be the required synthesis, but it is not yet known how to articulate this theory fully. This does not mean, however, that the present quantum and relativistic theories are regarded as having an authentically conjectural character. Instead the attitude seems to be that these theories are bound to survive 'in (slightly) modified form' as limiting cases in the future unifying theory; this is why a 'synthesis' is being consciously sought.

For all these reasons, a sophisticated scientific realist is likely to be quite circumspect concerning what it is reasonable to believe about presently accepted theories. Realists do not generally claim that the rational belief is that those theories are true. Instead, as well as restricting their claims to theories in the mature sciences, they perhaps suggest that the rational belief is only that presently accepted theories are approximately true (or perhaps that it is probable to a certain—quite high—degree that they are approximately true).

1.4.2. *Arguments for Scientific Realism*

1.4.2.1. *The 'Miracle Argument'* It would be frivolous to try to deny that science has been strikingly successful at the level of empirical prediction and technological application. Television sets work, and they work, at least in part, because the prediction made by Maxwell's theory of the existence of electromagnetic waves turned out to be correct. Like it or not, atomic bombs work and they work, at least in part, because the predictions of the deep theories about the structure of matter on which they are based panned out at the empirical level. The empirical success of presently accepted theories in mature sciences, like physics, is hugely impressive. Moreover, aside from isolated (and usually soon recaptured) cases of 'Kuhn loss', science has seen a cumulative development at the empirical level: theories accepted now capture all the phenomena that theories accepted earlier did, and then some. How else can we account for this success except by assuming that what our theories say is going on 'beneath' the phenomena is 'essentially' or very largely correct? If, so the argument goes, what the theories say about 'transempirical' reality is true or 'close to the truth', then it is no wonder that those of their consequences that can be empirically checked by direct means turn out to be correct on such an impressive scale. But were those purely theoretical claims not correct (or not even intended to be descriptions of what goes on 'behind' the phenomena), then their empirical success would seem entirely mysterious.

This often rehearsed point can never be more than a plausibility argument—it is, of course, logically possible for a theory to be false and yet have an impressive range of true empirical consequences. (Trivially, every false theory has infinitely many true consequences.) Attempts to formalize the argument as itself a scientific explanation of science's success—for example, as some form of

(meta-level) inference to the best explanation—have generally come to grief. Moreover, there is, as van Fraassen has often emphasized, no *need* to explain science's empirical success—the scientific enterprise cannot, on pain of infinite regress, demand that everything be explained; and the weakest claim that would entail (though hardly explain) the empirical success of present scientific theories is, of course, simply that those theories are highly empirically adequate.

The empirical success of present scientific theories none the less seems a strong plausibility consideration in favour of the claim that they have somehow or other 'latched on to' the way things are. The main source of this plausibility is the fact that a significant part of the empirical success of science has been theory-led. As discussed earlier, theoretical frameworks based on different core claims are invariably elastic enough to accommodate known phenomena after the event, but the impressive thing is that many of the empirical laws that are now known were first discovered only as a result of being predicted by theories. Thus, for example, the fact that the centre of the shadow of a small opaque disc is bright was discovered to hold only as a result of its being predicted by Fresnel's wave theory of light. How else could that theory get such a surprising and hitherto unknown result right unless what it says about what is going on behind the phenomena is at least partially correct?

1.4.2.2. *The Argument that Realism is* Heuristically *more Potent* It is has often been suggested—long ago by Feigl and by Popper, for example—that the realist view of theories has proved itself heuristically the more fruitful view: those scientists who have made theoretical breakthroughs are invariably ones who insisted on interpreting present theories as (successful) attempts to describe an underlying reality rather than as merely instrumental frameworks.

One sharp form of this argument goes as follows. Everyone now acknowledges that 'background information' plays a significant role in the development of theoretical science. But when this uncontroversial point is properly thought through it in fact supplies an important argument for realism. Background information includes realistically interpreted assertions about causal relations and theoretical entities that are essential to the way in which that information is used—*both* in underwriting the acceptance of certain theories *and* in developing further theories (sometimes even successor theories). Moreover, the theories arrived at in this way have generally achieved significant instrumental success. We cannot understand the instrumental success of this aspect of scientific procedure unless we assume that the background 'information' relied on is indeed information—that the realist claims about causal connections ineliminably involved in the background information are at least 'essentially' correct and the associated theoretical entities real.

To take one especially simple case, Richard Boyd has argued that background information is used to locate the likely weak spots of proposed theories and hence to guide the construction of severe tests of that theory. What counts as a proper test of a theory about, say, the precise mechanism through which some drug *D* works to destroy bacterium *B* will be indicated by background information that underwrites plausibility judgements about likely alternative hidden mechanisms by which *D* might operate on *B*. These plausibility judgements are, in turn, based on already accepted theories of hidden mechanisms in other pharmacological cases. This method of identifying the proper tests of a theory has, so this argument goes, strikingly often led to the acceptance of theories that have thereafter continued to be empirically reliable. This, then, is one successful aspect of scientific practice which is based on a realist view of theories and theoretical entities and whose success can only be accounted for on the assumption that those background claims have really latched on to the way things are.

1.4.2.3. *The Argument that erstwhile 'Theoretical Entities' may later become 'Observable'* Finally, realists have argued that the distinction between theoretical and observational claims is itself subject to change as science progresses. An entity might start out as explicitly 'theoretical', as explanatory of things or events that can be observed; but, with the advance of technology, it might become observable, at any rate according to the ordinary scientific usage of that term. The bodies in the solar system aside from the earth, for example (understanding the term 'body' to imply that they are constituted of ordinary physical matter), were theoretical entities for a long time (they might after all have been, and were sometimes believed to be, gods or divine fire or whatever), but with the development of space technology they could be observed in a much more direct sense. Photographs and rock samples could be taken, and eventually a few people 'observed' one such body in pretty well as direct a sense as the rest of us observe the earth.

Many similar examples can be drawn from biology: entities like the mitochondria (organelles within the cell nucleus where combustion takes place) were initially introduced as explanatory entities (there had to be something in the cell that produces its energy!), but can now be 'observed' with the help of an electron microscope. (And if that doesn't count as 'real' observation, where, and on what principle, is the line to be drawn? Do I observe through my spectacles, or only—more directly but *very* obscurely—when I take them off?) Examples like these, it is argued, show that there is no principled, fixed distinction to be drawn between 'observable' and 'theoretical' entities; and hence they constitute strong evidence against any philosophy that gives a radically different epistemological status to 'observational' as opposed to 'theoretical' assertions. Anti-realism in its various versions is precisely such a philosophy. It treats observational claims and generalizations as straightforwardly true or false, and their associated entities as unambiguously real, but treats theoretical assertions and their associated ontologies quite differently, usually as codification schemes and useful fictions, respectively.

Van Fraassen has, however, tried to turn this argument around in favour of *anti*-realism, introducing in the process an interesting and much discussed account of what counts as 'observable'. For details refer to van Fraassen (1980).

1.4.3. *Arguments against Scientific Realism*

Several arguments against realism have been based on the philosophical idea that the position is *unnecessarily inflationist*—that it involves metaphysical assumptions that can be, and should be, excluded from science. (This is the main thrust behind van Fraasen's anti-realism, for example.) There is also a series of arguments based on *underdetermination* of theory by data—a methodological phenomenon which is taken by some critics *both* to render realism implausible *and* to remove entirely the force of one of realism's chief supports. (On underdetermination see Papineau's treatment in the earlier companion to this volume.) However, the most telling arguments, in my view, are to do with (1) difficulties concerning the idea of approximate truth and, especially, (2) the problems posed for realism by theory change in science.

1.4.3.1. *The Problems of 'Approximate Truth'* Theories are often accepted despite being known to be in need of modification; and the history of science shows that, often enough, theories that were accepted and *not* initially known to be in need of modification are in fact subsequently modified or rejected outright. Realists claim that it is none the less reasonable to believe that currently accepted theories in the mature sciences are 'essentially' or 'approximately' true. But what exactly does this mean?

Realists generally adopt a Tarskian correspondence notion of truth; and would surely like an account of approximate truth which parallels Tarski's clear and definite analysis. Popper attempted to provide such an account for the weaker notion of one theory's being *closer to the truth* than another.

The basic idea is that one theory *A* (considered as a deductively closed set of statements) is closer to the truth (has 'higher verisimilitude') than another theory *B* if and only if *either A* has more true consequences than *B* without at the same time having more false consequences, *or A* has fewer false consequences than *B* without at the same time having fewer true consequences. Of course, every theory has (denumerably) infinitely many consequences and any false theory has infinitely many true and infinitely many false consequences. The 'more than' relation involved here cannot, then, be defined in terms of set cardinality, for in that sense every false theory has exactly as many false

consequences as it has true consequences and exactly as many of both as any other false theory. Popper suggested that—at least for a significant range of cases—the subset relation might be used to better effect to provide the required ordering: A being defined as having more true consequences than B if the set of true consequences of B is a proper subset of the set of true consequences of A.

Unfortunately, as Tichy (1974) and Miller (1974) proved independently, it follows from Popper's definition that again any two false statements have the same verisimilitude—no false statement has more Popperian verisimilitude than any other. The proof is straightforward.

Popper's account is that theory A has more verisimilitude than theory B if either of two conditions holds:

(1) the set of true consequences of A properly includes the set of true consequences of B while the set of false consequences of A is included in the set of false consequences of B [intuitively: A has (strictly) 'more' true consequences but no more false consequences than B]; or

(2) the set of false consequences of A is properly included in the set of false consequences of B while the set of true consequences of B is included in the set of true consequences of A [intuitively: A has (strictly) 'fewer' false consequences than B without paying for this by also having fewer true consequences].

Can condition (1) hold for any pair of false theories A and B? Well, suppose, in line with the first part of condition (1), that the set of A's true consequences properly includes B's: this means that there is at least one sentence, call it t, that (a) is true, (b) follows from A, but (c) does not follow from B. A is false—let f be any one of its false consequences. The conjunction $t\&f$ (a) is false, (b) follows from A, and (c) does not follow from B (if it did then t would too, but t is by assumption one of the extra true consequences of A not shared by B). Hence there is at least one false consequence of A (namely, $t\&f$) that is not a consequence of B; and so the second part of condition (2) cannot hold, if its first part does. Hence condition (1) never holds of any pair of false theories.

How about condition (2)? Suppose, in line with the first part of the condition, that the set of A's false consequences is properly included in B's: this means that there is at least one sentence, call it g, that (a) is false, (b) follows from B, and (c) does not follow from A. (So, intuitively, g is one of the sentences that is to show that A is 'less false' than B.) Let h be any false consequence of A. The conditional $h \rightarrow g$ (a) is true [the truth-table for the conditional gives the truth-value true when both the antecedent and the consequent are false], (b) follows from B [$h \rightarrow g$ is equivalent to $-h \vee g$ and B entails g], but (c) does not follow from A [if it did then, given that h does, so, by *modus ponens*, would g; but g is, by assumption, not a consequence of A]. So, if the first part of this second condition is satisfied (so that A lacks a false consequence that B has), then A must also lack a true consequence that B has (namely, $h \rightarrow g$) and so the second part of the condition is not satisfied. Hence, condition (2) also fails to hold for any pair of false theories.

If 'more' and 'fewer' are understood in this set-inclusion sense, then the Tichy–Miller result shows that Popper's idea cannot work because if A has more true consequences than B, then it cannot have fewer false consequences, while if A has fewer false consequences, it cannot have more true ones.

Other definitions of increased verisimilitude and of closeness to the truth have been mooted, but none has won general acceptance (largely because of what seems to many critics an undue dependence of the proposed measure of a theory's verisimilitude on the language in which the theory is expressed). This leaves the sophisticated realist in an unfortunate position; he claims that present theories in the mature sciences are rationally held to be approximately true, but he is unable to give any acceptable precise analysis of what approximate truth means.

1.4.3.2. *The 'Pessimistic Induction'* An eighteenth- or nineteenth-century realist would have claimed that it was reasonable to believe that Newton's theory is 'approximately' or 'essentially' true. Newton's theory involves action-at-a-distance forces of gravity acting across an absolute and infinite space, with a separate, absolute notion of time, according to which any two events simultaneous for one observer are simultaneous for all. A contemporary realist, however, advocates a realist attitude towards a quite different accepted theory—the general theory of relativity. According to the latter, Newton's theory is, of course, a good approximation in a whole range of applications; indeed its predictions, though they always strictly differ from the truth ('the truth' as of course seen by relativity theory), are empirically indistinguishable from it for bodies moving with a sufficiently small velocity compared to that of light. But, good empirical approximation though it undoubtedly is, Newton's basic theoretical claims are simply wrong: there is no action-at-a-distance; instead bodies move along geodesics in a curved spacetime, of which time is an integral, non-separable part, with the consequence that, quite contrary to the absolute classical conception, simultaneity is frame-dependent. It is difficult to see, even intuitively, how Newton's basic theoretical claims could be said even to 'approximate' what relativity theory sees as the truth.

Similarly an early nineteenth-century realist in optics would have held that Fresnel's elastic solid ether theory was at least 'essentially' correct, and hence that something like the elastic solid ether and waves in it exist. Yet, it too was later replaced by theories apparently radically different from it. And unless we stretch the notion of 'something like' to (and beyond) breaking-point, it seems difficult indeed to see Fresnel's fundamental theoretical assumptions, involving as they do an elastic solid medium filling the whole of space, as 'something like'

the truth (as of course current theories see it). It is hard to see any ontological resemblance between probability waves in a quantum field and mechanical waves in a material medium.

The 'pessimistic induction' encourages the conclusion that changes of an equally revolutionary kind will, at some time in the future, affect the theories that are presently accepted in science (or at least the conclusion that it is not reasonable to hold that there will definitely not be such changes). How, then, can it be reasonable to hold a realist view of presently accepted theories?

Notice that this argument not only attacks the realist thesis, it also threatens to take all the force from the realist argument from predictive empirical success. The realist asks whether we can really take seriously the possibility that our present theories are as predictively successful as they are if their transempirical claims are 'way off beam'. The argument from scientific revolutions has the effect of simply pointing such a realist towards the history of science, which, allegedly, provides a whole catalogue of theories that were predictively successful, but were none the less quite radically false (or so presently accepted science tells us).

1.4.4. *Realist Rejoinders to these Arguments*

It is difficult to see any explicit response in the writings of leading realists such as Richard Boyd to the difficulties in articulating a precise notion of approximate truth. The implied response seems to be that we have a firm enough intuitive grip on the notion of truth-likeness not to worry about the formal difficulties. Does the realist have a more systematic response to the apparently most telling argument—that from scientific revolutions?

There seem to be three moves that realists could make. First, they could argue that accounts of the 'revolutionary' discontinuities in science have been greatly exaggerated. This might seem a rather desperate measure, but it can be argued with some degree of plausibility at least in some cases. Returning to the example of the elastic solid 'luminiferous ether', one suggestion—advocated, for example, in Kitcher (1992)—is that, although the elastic solid ether was indeed rejected by later theories, it in fact was an essentially idle component of the classical wave theory of light. (No realist should advocate a 'realist attitude' towards all theoretical claims, even theoretical claims within successful theories. Some play no effective role and are, in Kitcher's terminology, 'presuppositional'. Newton's assumption that the centre of mass of the universe is at absolute rest is surely a case in point.) Hardin and Rosenberg (1982), analysing the same ether example, have made the interesting—though I think ultimately unacceptable—suggestion that the elastic solid ether was not in fact rejected in passing to electromagnetic theory; rather Fresnel and other classical wave theorists had (without of course being aware of it) been referring

to the electromagnetic field all along when talking about the elastic solid ether.

A second move the realist might make is to restrict his realism to those theories and those theoretical entities for which the history of science provides no basis for the pessimistic induction. This sort of restricted realist (Jon Dorling, for example, appears to favour this position) looks at the history of science and, wherever she believes she can tell a roughly 'continuous' story, advocates a realist attitude to the theory that presently stands at the culmination of the story. Where, however, the discontinuities seem undeniable, such a 'selective realist' admits there is no ground for a realist view. So she might be realist about atoms and electrons, but not about, say, curved spacetime. Aside from issues about what counts as 'continuity' between two different theories, this suggestion seems *ad hoc*: no principled reason is given why some 'revolution' might not in the future occur regarding some item of scientific ontology that has so far had an (allegedly) untroubled history.

A third ploy is to try to argue that even where there have been radical discontinuities at the most basic theoretical level in some science there is a 'level' (below that of the fundamental theories but above the purely empirical) at which there has been continuity or quasi-continuity despite 'scientific revolutions'. Consider again the case of the elastic solid luminiferous ether. Freeze the history of science at the point where Maxwell's electromagnetic theory with its primitive electromagnetic field had been accepted. From that vantage-point, there is an easy explanation of the success of Fresnel's elastic-ether theory of light: from the later point of view, Fresnel clearly misidentified the *nature* of light, but his theory none the less accurately described not just light's observable effects but also its *structure*. There is no elastic solid ether of the kind Fresnel's theory (problematically but none the less importantly) involved; but there is an electromagnetic field. The field is not underpinned by a mechanical ether and in no clear sense 'approximates' it. Similarly there are no 'light waves' in Fresnel's sense, since these were supposed to consist of the motions of material ether-particles. None the less disturbances in Maxwell's field do obey formally similar (in fact, and unusually, mathematically identical) laws to some of those obeyed by the 'materially' entirely different elastic disturbances in a mechanical medium.

Unless—surely very much in the spirit of anti-realism—we think of these theoretical notions as characterized by their observable effects, then we have to allow that Fresnel's most basic ontological claim that the vibrations making up light are vibrations of real material ether-particles subject to elastic restoring forces was entirely wrong. A change in the value of a *sui generis* electric force and a movement of a material particle from its equilibrium position are more 'like chalk and cheese' than are real chalk and cheese. But if Fresnel was as wrong as he could have been about what oscillates, he was right, not just about

many optical phenomena, but also that those phenomena depend on the oscillations of something or other at right angles to the light. His theory was more than empirically adequate, but less than true; instead it was *structurally correct*. There is an important 'carry-over' from Fresnel to Maxwell, one at a 'higher' level than the merely empirical, but it is a carry-over of structure rather than content. Both Fresnel's and Maxwell's theories make the passage of light consist of wave forms transmitted from place to place, forms obeying the same mathematics. Hence, although the periodic changes which the two theories postulate are ontologically of radically different sorts—in one material particles change position, in the other field strengths change—there is none the less a structural, mathematical continuity between the two theories.

This third realist response need not be unacceptably 'Whiggish' (Whig history' allows unbridled use of hindsight and sees everything in history as a cumulative progression towards our present state). This response does not assert—indeed it explicitly denies—that Fresnel's theory really amounted to a subtheory of Maxwell's all along, that Fresnel's theory was always about Maxwell's field. Fresnel's theory itself is not a subtheory of Maxwell's, but a structurally identical *facsimile* of Fresnel's theory is. And it is the fact that this facsimile is entailed by the later theory that explains why, from the later vantage-point, the empirical predictive success of Fresnel's theory appeared as no lucky accident.

This account, in terms of the material falsity, but structural correctness, of Fresnel's theory is essentially that given by Poincaré in his *Science and Hypothesis*. The generalization of it—that the structure of earlier theories rather than their content is what is retained through revolutions—might be called 'structural realism'. This may be the most promising line for the realist—though like the other realist responses it certainly faces difficulties and needs much further elaboration. (See the recommended readings for further details.)

2. NATURALIZED PHILOSOPHY OF SCIENCE

What status do the correct principles of scientific methodology (whatever they may turn out to be) have? As indicated above, philosophers have found it hard to agree on *what* these correct principles exactly are; could this be because they have had an incorrect view of the *sorts* of principle they are and hence of how those principles are themselves to be validated? Suppose, for example, we decide that it is a principle of correct scientific, evidential reasoning that *ad hoc* accounts of phenomena weigh less heavily in favour of the theory that provides them than do predictive accounts. What status does this principle have? How

might it be justified if it were challenged? Until recently, the almost universally held answer would have been that it needs to be justified as an a priori principle—part of the 'logic of science', not itself dependent on any information provided by science. It is, however, becoming increasingly common for philosophers of science to argue that methodological principles are themselves part of the scientific enterprise and receive *empirical* justification in the same way as scientific theories.

Kuhn, for example, has claimed that

> Rather than being elementary logical or methodological distinctions, which would thus be prior to the analysis of scientific knowledge, [rules of appraisal] now seem integral parts of a traditional set of substantive answers to the very questions upon which they have been deployed. That circularity does not at all invalidate them. But it does make them parts of a theory and, by doing so, subjects them to the same scrutiny regularly applied to theories in other fields. (1962: 9)

Larry Laudan is still more explicit:

> scientific methodology is itself an empirical discipline which cannot dispense with the very methods of inquiry whose validity it investigates. Armchair methodology is as ill-founded as armchair chemistry or physics. (1987: 24)

And Ron Giere explicitly suggests that while the 'endless dispute' about the right account of scientific rationality, the right inductive logic (broadly construed), 'is generally taken to indicate the difficulty of the problem, it . . . may [in fact show] . . . that there is something fundamentally mistaken about the whole enterprise' (Giere 1988: 37). Philosophers should be looking not for a priori principles of 'rationality', but simply for correct *descriptive* accounts of how scientists as a matter of fact operate.

Philosophers of science nowadays tend to be much better informed than their predecessors about the details of the sciences they are philosophizing about, and this, I have no doubt, has led to important improvements in the field. Philip Kitcher asks, 'How could our psychological and biological capacities and limitations fail to be relevant to the study of human knowledge?' The answer is surely that they could not. Similarly, purely descriptive facts about the history of science cannot fail to be relevant to philosophy of science. None the less, it's a big step from 'relevant to' to 'constitutive of'; a big step from 'you won't do good philosophy of science if you don't know about scientific theories and practices' to 'philosophical claims about science are on a par with scientific theories and are to be assessed in the same way'.

Here I shall outline some of the lures that have tempted some philosophers to make this big step; and raise the issue of whether a fully naturalized view of the philosophy of science can in fact be adopted without abandoning the idea that science is epistemically special.

2.1. 'Epistemology Naturalized'

Perhaps the most celebrated advocate of 'epistemology naturalized' in recent times is Quine. Without getting into questions of Quinean scholarship, the picture that most commentators see as emerging from Quine is as follows. We should think of all our knowledge—mathematical, methodological, and logical, as well as scientific—as facing experience as a corporate whole. Some parts of this 'web of belief' may be more firmly entrenched, and so harder to shift, than others, but this is a merely pragmatic issue. All parts—and this includes methodological and logical claims—are basically on a par. The stimuli that the external world provides us sometimes force changes in our webs of belief. There are in principle indefinitely many ways in which we might make such changes—not only do we have the Duhemian choice between modifying either a core theory or an auxiliary one, we also have the option of modifying a logical or methodological principle instead of any 'substantive' one, central or auxiliary. The changes that are in fact made are governed by the attempt to maximize *simplicity*. There is no reason in principle why, given some particular recalcitrant experience, it should not turn out to be simplest to modify our 'analytic' methodological or logical principles rather than our 'synthetic', substantive scientific principles (though there are, no doubt, as a matter of fact rather few cases of this kind compared to cases in which it turns out to be simplest to modify a non-'central', 'synthetic' assumption).

An obvious question arises about this view. Suppose we start off with 'web of belief$_1$'; this is bombarded with some initially recalcitrant experiences, and we switch to web of belief$_2$ as the simplest available modification. Is the criterion of simplicity itself a web-of-belief-dependent notion or does it stand outside the webs as an independent, unjudged judge?

If the former, then, having adopted web of belief$_2$ (complete with simplicity notion$_2$), we shall no doubt be able to justify that web as the simplest available modified system; but that justification is internal—dependent on elements of the web itself. If simplicity is simply simplicity-as-seen-from-within-a-web-of-belief, then of course there is the possibility that some other group of investigators adopting a quite different notion of simplicity might prefer a radically different web of belief$_3$ as replacement for web of belief$_1$. No independent reason to think the majority right and this minority wrong could then be given. This is relativism.

If, on the other hand, we want to claim that there is an objective, independent right and wrong in these cases—that it is, say, objectively simpler to adopt relativity theory than to explain the null result of the Michelson–Morley experiment by sticking to classical physics and adopting the Lorentz–Fitzgerald contraction hypothesis—then we must regard the criterion of simplicity at least as outside the war. *Some part of methodology and logic* must be regarded as privileged (and therefore not on a par with our substantive beliefs) if relativism is to be avoided. And to regard some part of methodology and logic as privileged and so not on a par with our substantive beliefs is exactly to deny (fully fledged) naturalism. It seems that if this Quinean position is fully naturalized, then it entails relativism of standards; and conversely if the position endorses the idea that one scientific shift may be objectively preferrable to alternative shifts, then it is not fully naturalized.

2.2. Scientific 'Reductions' of Philosophy of Science

There are several arguments in the recent literature along the following lines. First, the idea of a special scientific rationality is deeply suspect, if only because so many philosophers have tried to articulate it and have produced so little consensus. Secondly, we should draw the obvious conclusion from this—namely that there is no such defensible notion. Thirdly, this might seem to lead inevitably to historical and community relativism—people, groups who hold beliefs at odds with ours, cannot legitimately be held to have irrational, mistaken beliefs, but merely a different (not in any sense worse) belief system. But, fourthly, this is not in fact so, because science itself supplies us with all sorts of reliable information about how accurately to gather information, how accurately to create knowledge.

One forceful (and admirably clear) representative of this line of argument is Ron Giere. According to Giere (1988), the 'endless dispute' about the right account of scientific rationality 'is generally taken to indicate the difficulty of the problem. But it may also be taken as a basis for suspecting that there is something fundamentally mistaken about the whole enterprise'. This is a conclusion drawn also by social constructivists such as David Bloor; and Giere is quite clear that the constructivists end up in an inescapable and entirely unacceptable relativism, according to which there is no 'objective' preference ordering of scientific theories given the evidence but only different orderings depending on one's social circumstances. However, such relativism is *not* an inevitable consequence for the naturalizer: 'There is [indeed] something important missing from the sociological account. [But t]hat something is not rationality, but causal interaction between scientists and the world. . . . There is a way to avoid relativism without appeal to 'standards' at all. This is to focus on *cognitive processes*, such as those involved in representation and judgment, which are shared by all scientists' (1988: 56). Thus philosophy of science should in fact be pursued as a branch of cognitive science—using our knowledge of the biological and psychological bases of our cognitive processes to explain how science has developed and certain theories have been accepted. On this approach relativism 'is largely [!] avoided, [since w]e know [*sic*] on biological grounds that

the processes of representation and judgment are similar for most scientists' (1988: 58).

Several problems for this approach seem to stand out. First, we will surely avoid describing different, conflicting sets of standards for 'good science' on this approach only by heavily (though no doubt implicitly) restricting in advance our choice of who is to count as a genuine scientist. No doubt we could get close to the rough consensus envisaged by Giere (and also arrive at sensible appraisal rules) by studying the 'processes of representation and judgment' of Newton, Maxwell, Einstein, and the like. But what if the list also included the likes of Immanuel Velikovsky, or Duane T. Gish (one of the leading proponents of 'scientific' creationism)? The response that the latter are not 'proper scientists' carries of course normative or evaluative baggage—'not proper' according to which criteria? This suggestion for naturalizing philosophy of science is analogous to the suggestion that we can readily naturalize ethics: to arrive at a correct account of ethical principles simply describe how people as a matter of fact make moral judgements and decisions (though be careful to restrict yourself to the judgements and decisions of moral saints!).

Secondly, there is a slightly less obvious and so perhaps more insidious form of relativism inherent in the approach. Cognitive science, biology, neurophysiology, or whatever naturalists of this stripe recommend we employ to describe the sorts of causal interaction between the scientist and the world that produce knowledge consists, of course, in a set of theories. We do not simply observe the relevant causal interactions: theories tell us about them. These theories are presumably well supported by the evidence (or at any rate better supported than rivals). *Why accept those theories rather than other possible rival ones which would give us different accounts of knowledge?* Again we face a dilemma: *either* the answer to this question is that these theories themselves constitute knowledge in some objective sense—in which case we avoid relativism but only by in effect admitting that underlying this alleged naturalism is a non-naturalized criterion of the traditional philosophical sort; *or* the theories of cognitive science (or whatever) that underpin this account of knowledge have no special status themselves: all we can say is that some people believe them, but if others believe differently and hence give a different cognitive science account of knowledge, then that is equally legitimate. This second horn of the dilemma is again pure relativism.

2.3. Ways into Naturalism via History of Science

A final way into naturalism is based, not on cognitive psychology or any other scientific theory, but instead on the history of science. The background is the Kuhnian claim that even if we restrict ourselves to what most of us think of as

good science, we do as a matter of fact find changes in methodological principles from era to era (or paradigm to paradigm). Accepted so-called methodological standards are, historically speaking, no less paradigm-dependent than other more obviously substantive assumptions about the world. This at least makes the claim that methodological principles are a priori unattractive.

If, however, we draw the obvious conclusion that this reflects an absence of such absolute standards, then this seems to open the door to relativism even more explicitly than in the other approaches. If even the standards for judging good science are paradigm-specific, changing with changing paradigm, then nothing stays fixed on the basis of which a judgement of cognitive improvement can be made 'from outside'. If scientists switch, then, of course, having already adopted the new paradigm, they see their new position as an improvement. But if on the contrary they stick with the old paradigm, then they judge—perfectly correctly from the point of view of the standards which they still hold—that the new theory is inferior to the one they already have.

There is a good deal of plausibility in some of Kuhn's examples of 'methodological change', but they all involve taking a very broad view of what counts as a methodological principle. Of course 'methodology' is a very loose term. If we regard any principle that constrains the sort of theory scientists prefer now or have preferred at a given stage in science, then there are undoubtedly changes in methodology. One fairly obvious example is that in the eighteenth century, say, it would have been taken as an implicit requirement on any theory of some range of physical phenomena that it be deterministic, but nowadays, with the success of quantum mechanics, this is no longer so. But such changes in 'big methodology' do not imply that there have been changes in even the basic formal principles of theory preference (such as that inconsistent theories are unacceptable, theories should not be accepted before being tested against plausible rivals, and so on). Indeed the obvious way to explain why science no longer demands deterministic theories is by showing why it is that indeterministic quantum mechanics is such a hugely successful theory—*according to these basic formal principles of theory preference.*

There are, none the less, many interesting defences of 'naturalized' views in the recent literature. Larry Laudan in particular has argued explicitly that the thesis that methodology is part and parcel of science can be defended without either falling into relativism or sacrificing the normative, evaluative role of methodology. And various philosophers, Philip Kitcher as well as some reliabilists, have argued that the circle apparent in the naturalized view—that it takes for granted parts of science as constituting genuine information, while trying to articulate the standards for genuine information—is not a real problem. Readers wishing to know more about this fascinating and

still very much open field will find a guide to the relevant literature in the Bibliography.

3. PHILOSOPHICAL PROBLEMS OF CURRENT SCIENCE

Whether or not philosophy of science can be fully 'naturalized', whether or not it can be considered itself as part of science, it is certainly true that some of its most interesting problems involve close consideration of the details of scientific theories. In particular many of the most fascinating problem-areas in current philosophy of science concern foundational issues in current scientific theories. In this final section I shall try to give a flavour of this sort of problem-area. In Section 3.1, I consider in outline the measurement problem in quantum mechanics; in Section 3.2 I consider two particular issues that arise in connection with the (neo-)Darwinian theory of evolution.

3.1. The Measurement Problem in Quantum Mechanics

Is quantum mechanics a coherent theory? Exactly how much of a revolution in our classical ways of thinking is necessitated by acceptance of the theory? Although mastery of a range of demanding technicalities is of course necessary for a full appreciation of the issues here, the most central question can be elucidated using rather little technical machinery.

Stated at its most abstract, quantum theory ascribes to each physical system at any instant a quantum state; and it supplies two rules or laws that govern how that state will evolve over time. One of these—the Schrödinger equation—is entirely deterministic: it dictates that, left to itself and given the forces acting on it and the constraints governing it, the system's state at some later time $t + \Delta t$ is a specific function of its state at time t. The other rule—usually called the projection postulate—is indeterministic: it dictates *probabilities* for the outcome of various possible measurements that might be made on a system that is, at the instant the measurement is made, in a particular quantum state.

If a system happens to be in a special state with respect to some observable, say position, at the time that observable is measured (one of the 'eigenstates' of the operator corresponding to that observable), then the theory entails that the probability is unity that a specific value of the observable (a value characteristic of that eigenstate) will in fact be observed, all other possible values of that variable therefore having zero probability. For any state that is not an eigenstate of the particular operator corresponding to the observable being measured, a range of different possible values of the observable will have non-zero probabilities.

If no measurement is in fact made on the system at time t, then the theory entails that the system's state will continue to evolve in accordance with the Schrödinger equation: the probabilities for various possible outcomes of the observation at t then simply being the probabilities that those outcomes *would* have had if (counterfactually) an observation had been made at t. Suppose, however, a measurement *is* made at t on a system not then in an eigenstate of the corresponding operator. This measurement will produce some particular result—if the system consists of a single electron and the observable is position, then the electron will be detected at some particular position. According to quantum theory, the system will now be in the position-eigenstate corresponding to whatever value of position was in fact measured. This means that if that same position measurement were to be repeated immediately, then it would, with probability one, produce exactly the same result. The system is 'thrown' discontinuously into an eigenstate of the corresponding operator by the initial measurement (this is the so-called 'collapse of the wave packet'). Hence an immediately repeated observation of the same observable would (with probability one) produce the same result.

Can the indeterministic, probabilistic nature of quantum mechanics involved in the projection postulate be regarded as simply a reflection of our ignorance of the true state of the system? Suppose we are experimenting on some single particle, say an electron. Its initial state (dictated by its method of preparation) evolves over time and at some particular time—a time when its state is not an eigenstate of position—we make a position measurement on it. Suppose that the electron is in fact observed at a particular position x. The state of the electron at the time the observation was made specifies a probability (neither zero nor one) that the observation would produce the value x that it has in fact produced. Can we assume that the electron did in reality already have the position x which the measurement simply revealed, the probabilistic element of the quantum state description thus simply coding our ignorance of its detailed position ahead of the measurement?

Unfortunately (at any rate for conceptual conservatives) the answer is definitely 'no'. If quantum mechanics is correct—and masses of striking experimental results support it—then the world is really radically different from the classical picture. We have to get used to the idea that electrons, for example, may—in a perfectly definite sense—be at a given time everywhere and nowhere. One (deservedly) classic illustration of this feature of quantum mechanics is the two-slit experiment.

Suppose electrons are fired in random directions from some point-source at a screen in which there are two small slits symmetrically placed with respect to the source (see Figure 4.1). Quantum mechanics (correctly) implies that if the frequencies with which electrons arrive at the various points on the observation screen are recorded, then the result will be the 'interference pattern' illustrated

in the figure, curve 1. This is strange: when the electrons are observed at the observation screen, they are observed as discrete scintillations—as particles with definite positions; if they had definite positions throughout (whether or not *we* knew what those positions were), then each of them must presumably have *either* come through slit 1 *or* come through slit 2 in arriving at the screen from the source; but this seems to entail that if we could first just take the electrons that came through slit 1 alone and record where they arrived at the screen, and *then* could take the electrons that came through slit 2 alone and record where they arrived at the screen, and then simply added the two individual single-slit frequencies, we would get exactly the same total effect as in the original two-slit case. But this process can indeed be realized—by first running the experiment with slit 2 closed and recording the scintillations at the observation screen, and then running the experiment for the same length of time with slit 1 closed (see Fig. 4.1, curves 2). The observed frequencies with which electrons arrive at the various points of the observation screen in this second, two-stage experiment *is* of course the direct sum of the two individual slit frequencies (Fig. 4.1, curve 3); but this is an entirely different result from that obtained when both slits are open at the same time.

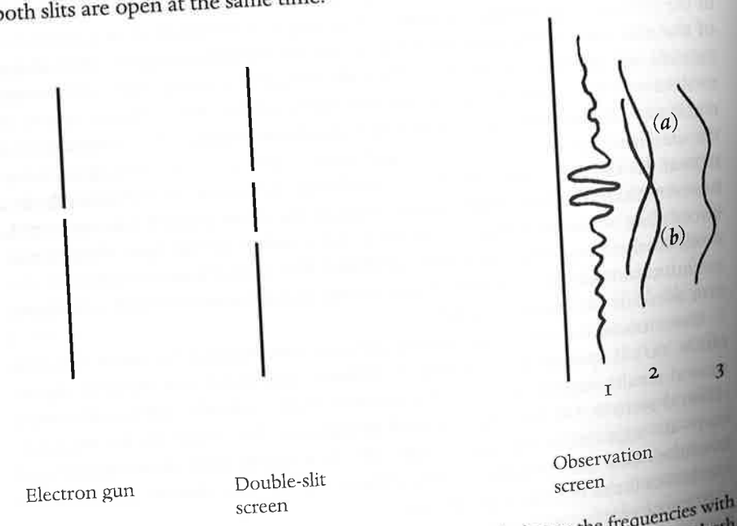

FIG. 4.1. The two-slit experiment with electrons. Curve 1 indicates the frequencies with which electrons are detected at the different points of the observation screen when both slits are open. The curves in 2 indicate those frequencies when (*a*) only the upper slit or (*b*) only the lower slit is open. Curve 3 indicates the frequencies obtained when only slit 1 is open for half the time, then only slit 2 is open for the rest of the time

The electrons in the experiment in which both slits are open cannot, it seems, be said to have arrived at the observation screen by *either* going through slit 1 *or* going through slit 2. Instead the electrons that pass through the double slit are, according to quantum mechanics, in a *superposition* of passing through slit 1 and passing through slit 2; and this superposed state cannot be simply an expression of our ignorance of which slit they 'really' passed through, because the fact that the quantum state is this superposition has real physical effects. (As you would expect from the above, quantum theory implies that if, for example, you set up some recording device for electrons behind each slit, then there would indeed be one electron recorded at one or other of these devices for every electron emitted from the source and eventually recorded at the observation screen, but the introduction of these devices would result in the final outcome being, as in the two-stage single-slit experiment, the direct sum of the two one-slit frequencies. The recording devices 'force' the electrons initially in a superposed state into eigenstates corresponding either to passing through slit 1 or passing through slit 2, and hence destroy the 'interference effects'.)

So, assuming of course that quantum mechanics is correct, we have to get used to the extremely strange idea of superposition. And systems are always in superpositions for some observables. Any quantum system in an eigenstate for one observable—so that the system in that sense has a 'definite value' of that observable (though remember that only means that a measurement of that observable will produce a that value with probability one)—will not at the same time be in an eigenstate for another complementary observable. For example, if an electron's position has been measured, so that it is, momentarily, in the eigenstate corresponding to its measured position, then it is inevitably in a superposition of momentum states. And, again, this superposition cannot be interpreted as meaning that, although the electron has a definite momentum, we do not know what momentum (and therefore what velocity) that is.

Is there any reason, though, why the 'strangeness' of the idea of superposition should mean that quantum theory is incoherent—that it cannot be regarded as a theory that is to be interpreted realistically? After all, science has been in this situation often enough before. The Newtonian idea of gravity as an action-at-a-distance force, for example, was often charged with incoherence when it was introduced (or, at best, as a throwback to pre-scientific, magical ways of thinking). Even Newton himself was unhappy with the notion and hoped eventually to 'reduce' gravity to some sort of mechanistic medium effect. The continued failure to develop an (independently testable) mechanistic explanation of gravity in the end led to the quiet acceptance that this notion could indeed perfectly well be regarded as primitive—a basic feature of the universe not in need of any explanation in terms of an underlying mechanism. And in fact, once scientists had stopped worrying about gravity as action-at-a-distance (not because it got any less 'strange' objectively speaking, but simply

because they got used to having the idea around), then further developments in science—notably Coulomb's electrostatic force—were happily based on action-at-a-distance ideas. Similarly, when Maxwell's idea of the electromagnetic field was first introduced, it was assumed—as almost a foregone conclusion (even by Maxwell himself)—that it eventually would have to be explained in terms of the antecedently 'understood' notion of a mechanical ether. That is, it was assumed that the varying electric and magnetic forces at each point in space could not be regarded as *sui generis* (after all, who had thought of such strange ideas before?), but would instead have to be accounted for as the results of some complicated contortions of a mechanical medium that fills space. Again the eventual outcome was that various attempts to 'reduce' the electromagnetic field to the ether failed, and scientists quietly came to regard this as a non-problem: that is, they came to see no obstacle to regarding the field as a primitive, basic notion that requires no further explanation. Is there any reason why familiarity should not eventually engineer a similar fate for quantum superposition?

No reason at all, so I would argue, *if* strangeness were the only problem. The fact that we find certain notions strange just means that they conflict with our previous ideas (ideas that themselves have lost their strangeness only through longevity). But there is an extra problem with quantum theory. *The* problem, the so-called measurement problem, is that, supposing the theory to be generally applicable, its two basic principles seem to be in outright conflict—they have contradictory implications about certain observable situations.

Here's how that comes about. If quantum theory applies generally, then it applies to 'macroscopic' objects as well as to 'microscopic' ones (or, rather, no such distinction is made by the theory). It might seem counter-intuitive to think of, say, cricket balls, no less than electrons, as in superposed states. This would entail, for example, that a cricket ball with a definite position has an 'uncertain' momentum (where this very unfortunate term must be taken to mean, remember, not that the theory cannot predict exactly what the momentum is, but rather that the ball has no definite momentum). But velocity is just momentum divided by mass and surely cricket balls in definite positions do have definite velocities? Well, if quantum theory is generally applicable, then it does unambiguously entail that there is an 'uncertainty' in the momentum, and hence in the velocity, of the cricket ball with a definite position. But the fact that the ball's mass is enormous compared to that of an electron means that, unlike that of the electron, the ball's velocity uncertainty is too small to be observable, and hence means that there is no clash with everyday experience. (Like many things in quantum theory, this is not universally accepted—but it does seem to be the general view.)

But if quantum theory applies to 'macroscopic' systems, if it assigns quantum states to 'macroscopic' objects, then measuring instruments, and, significantly, systems consisting of measuring instruments plus, say, single electrons, have quantum states that evolve in accordance with the laws of the theory. Of course, working out for some particular complex system of measurement device plus electron what quantum state it actually is in, at some given time, would be an enormously complicated, in fact entirely impractical, task. But the argument here is independent of the details of the state and depends only on certain very general features of it and its evolution (general features that the theory guarantees the state and its evolution must have whatever the details).

Suppose we have a pair of 'complementary' observables (this means that any system that is in an eigenstate of the operator corresponding to one observable is necessarily not in an eigenstate of the operator corresponding to the other). To make things as simple as possible, let's avoid position and momentum (each of which can take any of infinitely many values) and choose instead a pair of observables, each of which can take only two possible values. There are such operators according to quantum theory ('spin' for example), but again the details are unimportant. Just in order to have names for them, let's call the observables *wealth* and *age*, the first having the two possible values *rich* and *poor*, the second the two possible values *young* and *old*.

The fact that these are complementary observables entails, remember, that an electron that is, say, *rich* (or, more precisely, is in a state that will produce with probability one a reading of *rich* on a wealth detector) has no age—it is instead in a superposition of *young* and *old*, each of which has (in the simplest case that we shall suppose holds here) a half chance of being produced by an age measurement. Suppose then that we do indeed have a perfect wealth-detector—once again the details are of no significance, all that we need to know is that the detector has, say, two lights, one green, one red, and that if an electron that is in fact *rich* is fed into it, the green light invariably flashes (really 'flashes with probability one', but that wrinkle too won't matter here), while if a *poor* electron is fed into the detector, the red light invariably flashes.

Suppose a rich electron is just about to enter an activated wealth-detector. Whatever the details of the state of the overall system, it must at that instant be an eigenstate of both the operators (machine-)*ready* and (electron-)*rich*; and since quantum theory is empirically adequate and gets results like this right, the dynamics must imply that, once the electron is in the detector, that state has evolved into one that is an eigenstate *both* for *green light* and for *rich*. On the other hand if a poor electron is about to enter the detector, then the evolution is from a state that is an eigenstate for *ready* and *poor* to one that is an eigenstate for *red light* and *poor*.

But now suppose that an electron that we know (because of its preparation) is *old* is fed into that same activated wealth-detecting machine. The initial state Π of the combined system at the instant the electron is about to enter is an eigenstate for *ready* and *old*. What does quantum mechanics say will happen to

the combined system's state when the old electron enters the machine? Relative to the wealth observable, the electron is in a superposed state which gives a half chance of it measuring *rich* and a half chance of it measuring *poor*. The initial state of the whole system therefore is a superposition of *ready–rich* and *ready–poor*. In fact

$$\Pi = (1/\sqrt{2})\mu_1 + (1/\sqrt{2})\mu_2$$

where μ_1 is the eigenstate for *ready–rich* and μ_2 that for *ready–poor*.

The first principle of quantum theory—the Schrödinger equation—tells us what will happen to Π as the electron enters the wealth-detector. Again we don't know exactly what state Π is, and therefore there is no question of a full solution of the equation. But we only need one general feature of that equation, namely that it is *linear*. This means that, since $\Pi = (1/\sqrt{2})\mu_1 + (1/\sqrt{2})\mu_2$, and since we already know how μ_1 and μ_2 evolve in the circumstances at issue (namely into eigenstates of *green light–rich* and *red light–poor* respectively), the time evolution of Π *must*, whatever the details, be of the form

$$\Sigma = (1/\sqrt{2})\textit{green light–rich} + (1/\sqrt{2})\textit{red light–poor}.$$

(I am taking lots of liberties with notation here, but I trust the message is clear.) So the Schrödinger equation predicts that Π will, in these circumstances, change into Σ. I shall return to what this means in a moment. The problem, as we shall see next, is that the second principle of quantum theory—the projection postulate—predicts something quite different.

The procedure we have specified, of passing an electron into a wealth-detector, is of course a measurement of the wealth of an electron in state Π. Hence the projection postulate applies and it says that since Π, the initial state here, is no eigenstate of wealth, the measurement effects a discontinuous (and probabilistic) change of the state of the electron and hence of the electron plus measuring device into either the eigenstate *green light–rich* or the eigenstate *red light–poor* (each with probability one-half). This result predicted by the projection postulate is the actual result: that is, on any particular occasion when an old electron is fed into the wealth-detector then in fact either the green light flashes or the red one does and in either case, if the electron's wealth is immediately remeasured, then the same light flashes again; and, moreover, if the experiment is run many times, lots and lots of old electrons being fed into the wealth-detector, then, on average, half the time the green light flashes and half the time the red light flashes.

The state Σ that the overall system must be in according to the Schrödinger equation is, on the contrary, a very strange superposition of green light flashing and red light flashing: for a system in state Σ there is no fact of the matter about whether it is the green or the red light that flashes.

This conflict between the two basic principles of quantum mechanics—the conflict brought about at root by the fact that the projection postulate gives a special role to 'measurements', while the Schrödinger equation applies in principle to all objects including so-called measuring devices—is 'the measurement problem'. It seems to show that quantum mechanics, as it stands, and for all its empirical success, cannot be generally correct.

The problem, as many readers will have recognized, is related to the famous Schrödinger cat problem. But I leave readers to satisfy their curiosity about the cat—and indeed about whether it was curiosity that killed it—via the recommended readings for this section (e.g. Albert 1992; Healey 1989). These also contain discussions of a range of further fascinating logical and foundational issues raised by quantum mechanics.

3.2. Fallacies about Fitness

3.2.1. *Is Darwinian Theory Based on a Tautology?*

The second area of current science that raises interesting methodological and philosophical issues, some of which I want to outline, is the Darwinian theory of evolution. (This theory is still, of course, very much alive in current biology in the form of the 'neo-Darwinian synthesis'—basically Darwin, plus Mendel, plus lots of more recent modifications and extensions.)

Charles Darwin himself never used the phrase 'survival of the fittest'. This slogan was instead coined by Herbert Spencer. None the less, it is—for good or (mostly) ill—often thought of as capturing Darwin's leading idea. At various stages throughout the career of evolutionary theory, discussion of its credentials has been bedevilled by the charge that the idea is in fact nothing more than a tautology: Darwin claims that it is the fittest organisms that survive; but which organisms *are* the fittest?—those that survive.

An organism's fitness is not about survival as an end in itself but survival as a means of reproduction. But this threatens merely to redefine the difficulty rather than eliminate it: if survival of an organism involves representation in succeeding generations (of the organism's genes rather than of the organism itself, of course) and if the organism's fitness were measured by the number of offspring it leaves, then 'the fittest survive' threatens to reduce to the claim that the organisms that leave the most offspring are the ones that are most heavily represented in succeeding generations.

There are in fact a number of independent reasons why fitness cannot simply be a matter of the actual number of an organism's offspring (or even of the actual number of offspring that themselves reach reproductive age). Consider (to appreciate the most obvious such reason) two monozygotic twin chimpanzees, say, just about to reach sexual maturity, chewing on two opposite ends

of the same (long) shoot in the same jungle when lightning strikes one of the chimps dead, while the other survives (a little singed but essentially unscathed) to produce numerous offspring. Same genes, same environment—mere accident that one has many offspring, the other none: it would clearly be absurd to suggest that the surviving chimp had greater fitness than the other. As Mills and Beatty (1979) make clear, Darwinian fitness (if indeed it can properly be applied to individual organisms at all) is a dispositional or probabilistic notion: fitness is to do not with actual, but with *expected*, number of offspring—each of the chimps in this story has the same expected number of offspring, but an external chance event intervenes to make the actual numbers of offspring different.

This consideration, on its own, shows that the claim that the fittest survive cannot be merely analytic. To see this, suppose we have a range of coins with various biases giving them dispositions to produce heads varying from 0.5 (no bias) to 0.9 (heavy bias towards heads). Suppose that all the coins are tossed together some large number of times; the statement 'the coin which produced the largest number of heads was the most strongly biased one' may well be factually false, and so can scarcely be analytically true.

This consideration also shows that evolutionary theory cannot really be committed to the idea that all evolutionary changes are brought about by natural selection. It must (and of course does) allow for the possibility of various chance changes. Nor are chance and natural selection the only possibilities: a trait exhibited by an apparently successful species may be neither an adaptation (that is, a trait that itself yields its bearer greater expected fitness than if it were absent) nor merely present by chance. It might, for example, itself be evolutionarily neutral, but tied (by physico-chemical processes within the organism) to some trait that does have selective value: trait T may be selectively neutral, but the genes that code for trait T' (which does have a selective advantage) may also produce T, as a 'side-effect'.

Darwinian theory is altogether more complex and sophisticated than some of its critics like to acknowledge. But this complexity itself suggests the more challenging and interesting criticism that perhaps underlies the tautology objection. Accepting that fitness is expected reproductive success, how exactly can it be identified independently of observed facts about reproduction? Isn't it too easy to conjecture possible ways in which a trait might prove of selective value? Given that there is an enormous range of possible ways in which a trait might prove of selective value to an organism, and given that Darwinians are by no means committed to regarding every observed trait in an apparently successful species as an adaptation—as itself contributing to greater fitness—aren't they in a position of 'heads I win, tails you lose'? If some particular attempt to explain the presence of some feature of some organism runs into difficulties, there are always lots of other possibilities to hand. In other words, the serious question

that underlies the not-so-serious tautology objection is the question of whether Darwinian theory is really *testable*.

3.2.2. *Is Darwinian Theory Empirically Testable?*

This question was raised early in the career of Darwin's theory by his wonderfully named (and quite astute) contemporary Fleeming Jenkin. Jenkin pointed out that the Darwinian, faced with some aspect of natural history for which he has no immediate explanation, will scarcely be nonplussed:

> He can invent trains of ancestors of whose existence there is no evidence; he can marshal hosts of equally imaginary foes; he can call up continents, floods, and peculiar atmospheres, he can dry up oceans, split islands, and parcel out eternity at will; surely with these advantages he must be a dull fellow if he cannot scheme some series of animals and circumstances explaining our assumed difficulty quite naturally. (quoted from Kitcher 1982)

On the particular question of geographical distributions of organisms (why specific kinds of finch are found only on the Galapagos Islands, why marsupials are found only in Australasia, and so on), Jenkin claimed:

> The peculiarities of geographical distribution seem very difficult of explanation on any theory. Darwin calls in alternately, winds, tides, birds, beasts, all animated nature, as the diffusers of species, and then a good many of the same agencies as impenetrable barriers. There are some impenetrable barriers between the Galapagos Islands, but not between New Zealand and South America. Continents are created to join Australia and the Cape of Good Hope, while a sea as broad as the Bristol Channel is elsewhere a valid line of demarcation. With these facilities of hypothesis there seems no particular reason why many theories should not be true. However an animal may have been produced, it must have been produced somewhere, and it must either have spread very widely, or not have spread, and Darwin can give good reason for both results. (quoted from Kitcher 1982)

Not surprisingly, this charge has been seized on by modern critics of Darwinism. The leading 'scientific' creationist Duane T. Gish, for example, remarks:

> the architects of the modern synthetic theory of evolution have so skillfully constructed their theory that it is not capable of falsification. The theory is so plastic that it is capable of explaining anything. (quoted from Kitcher 1982)

Darwin himself, misguidedly, tried to meet this sort of charge by specifying a Popper-style 'potential falsifier' for his theory:

> If it could be demonstrated that any complex organ existed, which could not possibly have been formed by numerous, successive slight modifications, my theory would absolutely break down.

But how could anything so highly theoretical possibly be 'demonstrated'? Every theory is 'falsifiable' if this just means that there are statements that we might come somehow to accept and that are inconsistent with it (the theory's negation would always suffice). The real question concerns how theories come into contact with intersubjectively agreed observation statements: obviously the statement that some complex organ could not possibly have been produced by numerous slight modifications is no such observation statement.

The way properly to meet this 'unfalsifiability' objection is surely as follows. It will remind you of the earlier discussions of Kuhn and of prediction versus accommodation. Again the essential point is the one made long ago by Duhem. Single, 'isolated' scientific theories never have observationally testable consequences of their own but only when incorporated into (usually quite large) theoretical systems that involve a range of further theoretical assumptions. Major scientific achievements are characterized only in part by the central claims they make about the world; they also involve a specification of a set of problems and of patterns of reasoning for addressing those problems.

This is especially true in the case of Darwin's theory—many commentators indeed see the core of the theory as relatively insubstantial. Its power lies in the fact that it provides a framework within which particular explanations of a variety of phenomena have been constructed, explanations many of which are indeed *independently* testable. Contrary to Jenkin's claim, the Darwinian is not free to make any assumption he likes without fear of being proved wrong; assumptions about the movements of land masses, for example, may conflict with independent findings in geology. It is true that, should some particular assumption of this kind not work, there will always be others he can try—but then the crucial question becomes whether Darwinians have been able in some signficant number of cases to score successes by developing particular theories that have independent support. And the answer is overwhelmingly positive. If, for example, Darwinians conjecture that early mammals reached Australasia via Antarctica, then one would expect that appropriate mammalian fossils might be found in Antarctica—and there have indeed been important recent discoveries of such fossils (see the Kitcher references in the Bibliography).

Hundreds of similar examples can be cited. As usual, the classic examples are the most telling. The British biologist H. B. D. Kettlewell studied the peppered moth, *Bison betularia*. The speckled form of this moth was common in Britain before the Industrial Revolution and continues to predominate in rural areas. Melanic variants (with black colouring) became increasingly common in urbanized areas. What explains the change in frequency of the two forms near cities? One possibility is as follows. Industrial pollutants kill the lichens that normally grow on tree trunks; and the peppered moth often rests on tree trunks. (These are of course not simply assumptions plucked out of the air, but are rather themselves independently supported—indeed, parts of 'background knowledge'.) Once the lichens are gone, the tree trunks are uniformly dark. The speckled moths are less visible than the melanic ones against a lichen-covered bark, but the melanic variants are better camouflaged against the uniformly dark trunks. In general, better camouflage is a selective advantage since it lessens the risk of predation. The theory then is that the increased frequency of the melanic form of the moth in industrial areas is explained by its decreased visibility against tree trunks.

Some parts of this account come with independent support, but must the account as a whole be left as a mere possible explanation? Kettlewell performed experiments, capturing moths of both speckled and melanic forms, marking them, and then releasing them in areas with different degrees of industrialization. He later used a night-light trap to capture moths and noted the relative frequencies of the two variants amongst the marked moths that were recaptured. In industrialized areas, a greater proportion of the speckled moths were 'missing, presumed dead'. Rather than leave predation as a mere possibility, Kettlewell watched individual moths through binoculars after release and observed that some of them were indeed eaten, while resting on tree trunks, principally by robins and hedge-sparrows. Moreover, he performed other experiments which gave more support to this specific explanation by undermining alternatives. One alternative possibility, for example, would be that the melanic moths for some reason have greater fertility in industrialized areas. Kettlewell did experiments on the relative fertilities of the two kinds of moth reared in different environments and ruled out this alternative.

The case of Kettlewell's moths is deservedly famous because the evidence is so direct and compelling that it is often presented as a demonstration of natural selection at work. It shows that the way to counter the criticism that Darwin's theory is untestable is not by trying to argue that the basic postulates have experimentally decidable consequences, but instead to show that the basic Darwinian ideas can be, and have been, supplemented by various specific assumptions in such a way that the overall theoretical system thus created has significant independent empirical support.

Other cases are, unsurprisingly, less clear-cut: often there is evidence but it is somewhat less direct; in some cases there are plausible Darwinian accounts but no independent support. Although critics like to latch onto the latter kind of case, there is again nothing special about Darwinian theory in this regard. All theories—in physics as well as biology—face anomalies that they can, at any rate temporarily, account for only in an *ad hoc* way, permitting no independent test. The power of such explanations is simply derivative on the power of other explanations within the same programme or paradigm that do have independent support. So, for example, there is no doubt that the best explanation in the early seventeenth century of the failure to observe stellar parallax (a difference in the relative positions of stars at different times of the year) was the

Copernican one that there must indeed be parallax—the earth revolves around the sun and therefore the angular separation of a given pair of fixed stars will be different depending on where the earth is relative to them—but the distance of the stars from the solar system is so great that, although the parallax is real, it is too small to be detected with available instruments. This explanation is, taken by itself, entirely *ad hoc*—there was no independent evidence about the distance of the stars from the sun; the assumption was made exactly so as to deliver the known fact of no observable parallax within the Copernican system. None the less it was the best available explanation, because the Copernican system was, unlike its Ptolemaic rival, independently testable and independently confirmed by *other* phenomena (such as planetary stations and retrogressions). (And by the seventeenth century the Tychonic third way had dropped out as a serious contender for other reasons.) Theories, having been significantly independently confirmed in various areas, assume the right (temporarily of course) to provide the best available scientific explanation even in other areas where they cannot be independently supported. Darwinian theory, similarly, has been independently confirmed in a striking number of cases, and hence wins the right to supply the best explanations that can presently be provided even in areas where it is not directly and independently testable.

In sum, doubts about the testability of Darwinian theory turn out to be insubstantial.

3.2.3. *Adaptation, Teleology, and Explanation*

The second logical issue that I want to raise about evolutionary theory concerns precisely such 'adaptationist explanations' as the one provided by Kettlewell. Why have melanic moths become more prevalent in industrial areas? Because their colouring provides better camouflage in areas where industrial effluent has killed the lichens on tree barks and hence darkened the trees; and this means that the moth is less susceptible to predation than the original speckled form. It is standard to think of explanations as answers to 'why' questions. The above seems to be a good answer to a 'why' question. It can, it appears, be straightforwardly turned into an explanation of the colouration of melanic moths. Why are these moths black? In order to help avoid predation.

This second explanation accounts for the presence of a trait by stating *what the trait is for*, what value the trait has for its possessor. It therefore seems to be a sharply different form of explanation from anything found in the physical sciences. We would, for example, hardly look for an explanation of light's property of being (partially) reflected when passing from one medium into another in terms of the value this property has for light. Instead, the property is explained by showing that more fundamental laws plus initial conditions entail that light

must exhibit the property (the explanation is in terms, if you like, of the property's causal antecedents).

Do, then, adaptationist accounts form an equally legitimate pattern of explanation different from anything found in the 'harder' sciences? Or are such accounts inevitably second-best: temporary place-holders for the real explanations in terms of causes rather than effects that science will eventually uncover?

Suppose that we had a complete theory of the biochemical pathways through which the melanic moths' genetic make-up produces the dark colouring. Suppose that we had a complete historical account of how the mutation that first produced the melanism occurred and of the subsequent mating patterns through to some present population that we are interested in. We should then have, it would seem, a complete account of the causes of the dark colouring of this population of melanic moths. ('These moths have the dark colouring because they have genetic constitution *G* and because *G* plus biochemistry plus the given environment entails the dark colour; moreover, they have genetic constitution *G* because they are related in the following ways to a moth whose genome underwent mutation *M* at time *t* (who in turn was related to a moth whose . . .).') What we would not, however, have—or so some have argued—is an answer to our original 'why' question. We should have a full explanation of the causes of the colouring of the present population of moths in whatever industrialized area we are interested in. But we should not have an explanation of why we are confronted with *that* population of moths, in that particular geographical area: we should have no explanation of why it was melanic moths rather than speckled ones who predominantly survived to reproductive age so as to produce those in the population under study.

Many commentators see this argument as posing a challenge to the idea of a single, unified model of scientific explanation. The challenge is to produce a selectionist explanation for the spread of melanism in moths compared to the earlier speckled form that does not involve the assumption that the gene that produces melanism spread because of the value of the dark-coloured phenotypic trait to the moths in their environment. If the above argument is correct, then the problem cannot be one of incomplete causal information; the problem, it is argued, arises once we have supposed that such complete causal information has been garnered.

What is it that makes some commentators reluctant to allow adaptationist explanation as a legitimate form of explanation, on a par with the more familiar causal form? The outline answer is that such explanations—in terms of the value of a trait to the organism or of the *function* of a trait—are felt to smack of teleology, and it is often thought that the maturity of a scientific discipline can be measured by the extent to which it has eliminated teleological modes of thought.

As with so many terms in philosophy, it is not completely clear what counts as teleology. Plato was, so far as we know, the first philosopher to defend teleological explanations, and he assumed that teleology always involved intelligent design. It was appropriate to explain a trait in terms of the value that trait has for its bearer if and only if the bearer's designer regarded it as good for it to have that trait. Aristotle, on the contrary, defended teleological explanations as independent of considerations of conscious design. For Aristotle trait *T* in organism *O* was explained by identifying what *T* did for *O*, what activity it permitted *O* to perform that contributed significantly to its life.

Adaptationist explanations certainly appear teleological in this Aristotelian sense. And indeed some biologists explicitly endorse teleology as an essential aspect of evolutionary theory. The distinguished biologist Ayala, for example, holds that 'the presence of organs, processes and patterns of behavior can be explained teleologically by exhibiting their contribution to the reproductive fitness of the organisms in which they occur'. (quoted from Rosenberg 1985) And he explicitly characterizes the use of teleological explanation as 'not only acceptable but indeed indispensable' in biology. On the other hand, Darwin is often credited with being the one who finally eliminated teleology from biology: for example, by destroying the argument from design.

The resolution of this difficulty is the recognition that there two senses of teleology that need to be differentiated: an *anthropomorphic* sense, involving human-style intentions and purposes, and another, non-anthropomorphic sense. Darwin eliminated anthropomorphic teleology from biology.

Thanks to Darwin, there is no need in biology to talk of genes, or cells, or whatever exhibiting anything akin to purposeful deliberation. Admittedly, biologists often describe their subject-matter in this anthropomorphic way. Here is a typical short passage from a standard text in biochemistry (cited in this connection by Alex Rosenberg (1985); emphases added): 'These enzymes are highly selective in their *recognition* of the amino acids . . . A much more demanding *task* . . . is to *discriminate* between similar amino acids. . . . How does the synthetase *avoid* hydrolizing isoleucine-AMP, the *desired* intermediate . . . ?' But this and its (many) kin represent simply a manner of speech that can be eliminated without cognitive loss. There is no need for, and every reason to avoid, purposes in biology.

However, the above considerations about adaptationist explanations seem to show that Darwin did introduce explanations into science that are teleological, not in the anthropomorphic sense, but in Aristotle's sense of involving considerations of what is good for the organisms at issue. One suggestion, endorsed by, for example, the eminent Darwinian C. G. Williams, is to introduce the term *teleonomic* for the sort of teleology approved by Darwin, leaving *teleology* to carry the 'bad' connotations.

Other interesting issues about 'functional explanation' in biology can be pursued via the readings listed in the relevant section of the Bibliography.

BIBLIOGRAPHY

1. RATIONALITY, REVOLUTION, AND REALISM

1.1. Radical Theory Change in Science, and 1.2. The Impact of Kuhn's *The Structure of Scientific Revolutions*

Core Reading

Duhem, P. (1906), *The Aim and Structure of Physical Theory* (Princeton; Eng. trans. of 2nd edn. 1954).

Kuhn, T. S. (1962), *The Structure of Scientific Revolutions* (Chicago; 2nd edn. 1970).

——(1977), *The Essential Tension* (Chicago) See especially chapter 13.

Lakatos, I., and Musgrave, A. E. (eds.) (1970), *Criticism and the Growth of Knowledge* (Cambridge). See especially the article by Kuhn; and that by Lakatos developing his influential methodology of scientific research programmes.

Popper, K. R. (1963), *Conjectures and Refutations* (London). The first chapter of this book provides an introduction to Popper's falsificationist view of science, especially helpful for those who have not already come across it.

Further Reading

Hoyningen-Huene, P. (1993), *Reconstructing Scientific Revolutions: Thomas S. Kuhn's Philosophy of Science* (Chicago). The only comprehensive study of Kuhn's views and their development.

Kitcher, P. (1993), *The Advancement of Science* (Oxford). In my view the most interesting and sophisticated of recent attempts to account for the development and rationality of science.

Salmon, W. (1990), 'Tom Kuhn meets Tom Bayes', in C. Wade Savage (ed.), *Scientific Theories* (Minneapolis).

Worrall, J. (1990), 'Scientific Revolutions and Scientific Rationality: The Case of the "Elderly Hold-Out" ', in C. Wade Savage (ed.), *Scientific Theories* (Minneapolis).

1.3. The Personalist Bayesian Account of Rational Belief

Core Reading

Earman, J. (1992), *Bayes or Bust? A Critical Examination of Bayesian Confirmation Theory* (Cambridge, Mass.). The best recent uncommitted critical analysis of the merits and problems of some of the Bayesian approaches (Earman's view of the Bayesian system of confirmation is essentially the same as Churchill's view of democracy as a political system—that it is the worst system apart from all the rest).

Howson, C., and Urbach, P. M. (1993), *Scientific Reasoning: The Bayesian Approach*, 2nd

edn. (La Salle, Ill.). The clearest introduction to the personalist Bayesian view and its application to standard problems in philosophy of science. See esp. ch. 2 for an introduction to the theory of probability; and ch. 7 for Bayesian analyses of the Duhem, and other standard methodological, problems.

Further Reading

For Carnap's programme, see:

Carnap, R. (1947), 'Applications of Inductive Logic', *Philosophy and Phenomenological Research*, 8: 133–48.

——(1950), *Logical Foundations of Probability* (Berkeley).

The Bayesian approach owes a lot to Ramsey's path-breaking work; see:

Ramsey, F. P. (1931), 'Truth and Probability', in *The Foundations of Mathematics and Other Logical Essays* (London).

On the Bayesian approach to the Duhem problem:

Dorling, J. (1979), 'Bayesian Personalism, the Methodology of Research Programmes, and Duhem's Problem', *Studies in the History and Philosophy of Science*, 10: 177–87.

Redhead, M. L. G. (1980), 'A Bayesian Reconstruction of the Methodology of Scientific Research Programmes', *Studies in the History and Philosophy of Science*, 11: 341–7.

On the old evidence problem and prediction and accommodation:

Achinstein, P. (1990), *Particles and Waves: Historical Essays in the Philosophy of Science* (Oxford). An important rival view on the issue of prediction.

Brush, S. J. (1989), 'Prediction and Theory-Evaluation', *Science*, 1124–9.

Glymour, C. (1980), *Theory and Evidence* (Princeton). Ch. 3: the source of the old evidence problem.

Howson, C. (1984), 'Bayesianism and Support by Novel Facts', *British Journal for the Philosophy of Science*, 35: 245–51.

——and Urbach, P. M. (1993), *Scientific Reasoning: The Bayesian Approach* (La Salle, Ill.), ch. 11. Also important for the response to the charge of over-subjectivism.

1.4. Scientific Revolutions and Scientific Realism

Core Reading

Churchland, P. M., and Hooker, C. A. (eds.) (1985), *Images of Science* (Chicago). Contains a series of critical responses to van Fraassen's position together with very interesting replies from van Fraassen.

Hardin, C. L., and Rosenberg, A. (1982), 'In Defence of Convergent Realism', *Philosophy of Science*, 49: 604–15.

Leplin, J. (ed.) (1984), *Scientific Realism* (Berkeley). This collection of papers represents a wide range of views on the issue and has a very useful editorial introduction. See esp. the papers by Boyd (defending realism), Laudan (his celebrated 'Confutation of Convergent Realism'—perhaps the most detailed development of the 'pessimistic induction'), Putnam, and van Fraassen.

van Fraassen, B. (1980), *The Scientific Image* (Oxford). Generally regarded as the definitive statement of latter-day 'anti-realism'.

Further Reading

On the notion of verisimilitude and its problems:

Miller, D. (1974), 'Popper's Qualitative Theory of Verisimilitude', *British Journal for the Philosophy of Science*, 25: 166–77.

Niiniluoto, I. (1987), *Truthlikeness* (Dordrecht).

Oddie, G. (1986), *Likeness to Truth* (Dordrecht). Both Oddie's and Niiniluoto's books attempt to develop alternative accounts of verisimilitude that avoid the problems of Popper's.

Tichy, P. (1974), 'On Popper's Definitions of Verisimilitude', *British Journal for the Philosophy of Science*, 155–60.

Works developing alternative or intermediate views between realism and anti-realism:

Cartwright, N. (1983), *How the Laws of Physics Lie* (Oxford).

Fine, A. (1984), 'The Natural Ontological Attitude', in J. Leplin (ed.), *Scientific Realism* (Berkeley).

Hacking, I. (1983), *Representing and Intervening: Introductory Topics in the Philosophy of Natural Science* (Cambridge).

Poincaré, H. (1905), *Science and Hypothesis* (New York).

Worrall, J. (1989), 'Structural Realism: The Best of Both Worlds?', *Dialectica*, 43: 99–124. This paper explains and examines Poincaré's structural realist view.

2. NATURALIZED PHILOSOPHY OF SCIENCE

Core Reading

Giere, R. N. (1988), *Explaining Science: A Cognitive Approach* (Chicago).

Kitcher, P. (1992), 'The Naturalists Return', *Philosophical Review*, 101: 53–114.

Laudan, L. (1990), 'Normative Naturalism', *Philosophy of Science*, 57: 44–59.

Quine, W. V. (1969), 'Epistemology Naturalised', in *Ontological Relativity and Other Essays* (New York).

Further Reading

Goldman, A. (1986), *Epistemology and Cognition* (Cambridge, Mass.).

Kornblith, H. (ed.) (1985), *Naturalistic Epistemology* (Cambridge, Mass.).

Laudan, L. (1984), *Science and Values* (Berkeley).

——(1987), 'Progress or Rationality? The Prospects for Normative Naturalism', *American Philosophical Quarterly*, 24/1: 19–31.

Rosenberg, A. (1990), 'Normative Naturalism and the Role of Philosophy', *Philosophy of Science*, 57: 34–43.

3. PHILOSOPHICAL PROBLEMS OF CURRENT SCIENCE

3.1. The Measurement Problem in Quantum Mechanics

Core Reading

Albert, D. Z. (1992), *Quantum Theory and Experience* (Cambridge, Mass.). This contains the most accessible and lively account of the measurement problem that I know of,

and my treatment in the text is indebted to it. (I wish I could be as enthusiastic about Albert's preferred 'many worlds' *solution* to the problem.)

Further references to the extensive literature can be found in Albert's book; for one of many variant treatments, see:

Cartwright N. (1983), *How the Laws of Physics Lie* (Oxford), essay 9: 'How the Measurement Problem is an Artefact of the Mathematics'.

Further Reading

There are many other fascinating logical and methodological problems in quantum mechanics concerned especially with hidden variables, locality, and the Bell inequalities. The best introduction to these issues (and to the literature on them) is:

Redhead, M. L. G. (1987), *Incompleteness, Nonlocality and Realism* (Oxford).

A non-technical account of the problematic nature of the Bell inequalities can be found in:

Redhead, M. L. G. (1989), 'The Nature of Reality', *British Journal for the Philosophy of Science*, 40: 429–41.

See also:

Healey, R. (1989), *The Philosophy of Quantum Mechanics: An Interactive Interpretation* (Cambridge).

Hughes, R. I. G. (1992), *The Structure and Interpretation of Quantum Mechanics* (Cambridge, Mass.).

3.2. Fallacies about Fitness

Core Reading

Kitcher, P. (1985), *Vaulting Ambition: Sociobiology and the Quest for Human Nature* (Cambridge, Mass.), chs. 2 and 3.

Mills, S., and Beatty, J. (1979), 'The Propensity Interpretation of Fitness', *Philosophy of Science*, 46: 263–86.

Further Reading

Hull D. L. (1974), *The Philosophy of Biological Science* (Englewood Cliffs, NJ). A good introduction to the field.

Kitcher, P. (1982), *Abusing Science* (Cambridge, Mass.).

Rosenberg, A. (1985), *The Structure of Biological Science* (Cambridge).

Sober, E. (1984), *The Nature of Selection: Evolutionary Theory in Philosophical Focus* (Cambridge, Mass.).

5

THE PHILOSOPHY OF RELIGION

M. W. F. Stone