Chapter 2

Philosophy of Science: Classic Debates, Standard Problems, Future Prospects

John Worrall

The Background

Immanuel Kant's celebrated investigation of human knowledge started from the assumption that we have achieved rock-solid, indubitable knowledge – in geometry through Euclid and in physics through Newton – and from the question of how this was possible (especially in view of Hume's demonstration of the invalidity of inductive inference). Contemporary philosophy of science is a rich and multi-faceted enterprise and so any one way of viewing it will inevitably leave out much of importance and interest. Nonetheless, many of the classic debates and areas of current concern can be introduced by investigating how Kant's questions require modification in the light of the development of science since his time and by investigating the attempts made to answer those modified questions.

Two radical – apparently "revolutionary" – changes of fundamental theory occurred in the early twentieth century, those associated with the theory of relativity and with quantum theory. The former had the more direct effect on Kant's presuppositions and questions. If, at any rate, we think of geometry as a synthetic description of the fundamental structure of space, then Einstein's revolution involved the rejection of Euclidean geometry in favor of the Riemannian version of non-Euclidean geometry. Instead, for example, of two straight lines that are parallel being extendable indefinitely without intersecting, the new geometry states that any two straight lines (geodesics) eventually intersect. Far from being certainly true, Euclidean geometry (at least as a "physical geometry") is – it seems – not even true. Similarly, although Newton's theory (of mechanics plus universal gravitation) continues to be empirically adequate over a wide range of phenomena (basically motions involving velocities small compared to that of light), its fundamental claims about the structure of the universe – that space is infinite, that gravitation acts at-a-distance, that time is absolute so that two events simultaneous in one reference frame are simultaneous in all – are entirely rejected by relativity theory. Again, far from being certainly true, Newtonian physics is, it seems, not even true. Indeed, given that relativity theory denies action at a distance, suggests that space is finite (though unbounded), and entails that two events that are simultaneous in one frame of reference will *not* be simultaneous in another frame that moves relatively to the first, it is difficult for many to see intuitively how Newton's theory could count as even "close to the truth" (supposing for sake of argument that Einstein's theory were the truth).

These developments transform Kant's question into a dilemma. Is there some way of interpreting (or reinterpreting?) scientific theories so that the apparently radical nature of the revolutionary shift from classical to relativistic physics becomes just that – merely apparent? If so, then it might still be possible to argue that science *when properly understood*, delivers, if not outright certainty, then some close approximation to it. If not, if we simply have to accept that scientific development has involved revolutionary change at the most fundamental theoretical level, then we presumably cannot reasonably rule out the possibility of still further revolutions in the light of which our current theories will seem just as false as Newtonian theory now seems to us. And in that case, the question becomes what makes science special at all from the epistemic point of view?

Why is Science Special from the Epistemic Point of View?

Let's begin on the second horn of this dilemma – conceding for the sake of argument that the apparently revolutionary shifts are real. In that case, there is no prospect of continuing to hold that scientific theories are *proved* or established by unquestioned empirical data. What is it, then, that makes science and the methods of science special from an epistemic point of view? (There are of course some thinkers – mostly sociologists of science – who would reject this question, and insist that the conclusion we ought to draw from the existence of scientific revolutions is that science is just one human system of beliefs amongst others (such as the Azande system of magic) with no justified claim to any special epistemic status. But the staggering predictive success of our theories in "mature" science is so strongly at odds with this view that it is difficult to take seriously.)

Demarcation and falsifiability

The question of what makes science special is often called "the demarcation problem." One celebrated answer – directly motivated by the Einsteinian revolution – is Karl Popper's falsifiability criterion: science is special because, even though its theories are not provable from evidential statements, they are *refutable* by such statements. The Einsteinian revolution is – Popper (1959) suggested – a direct vindication of this view (and indeed that revolution was a major motivation for the view): the "revolution" was a great step forward because it involved the refutation of a highly falsifiable, but hitherto unfalsified theory (Newton's), and its replacement by a still more falsifiable – but not yet falsified – theory (Einstein's). In contrast, non-scientific claims – metaphysical claims, such as that God exists, or claims that Popper categorized as *pseudo*scientific, such as the claims of astrology or of Freudian psychoanalysis, are (allegedly) entirely unfalsifiable: no possible evidential statement could contradict any such claim and hence establish its falsity. Science is special because at least we can know when we are wrong.

It is now (almost) universally accepted that Popper's account fails. One issue – raised right at the beginning by Reichenbach, for example – was whether the problem of induction, and, in particular, the so-called "pragmatic problem," can ever be solved in a purely falsificationist way. It seems positively irrational not to base our technological interventions – in building say bridges or aeroplanes – on the best available scientific theories. But would this judgment be underwritten simply by the report that those best available theories are so far unrefuted – that is, unrefuted in tests already performed? We surely also need some sort of reason to think that the past test-record of those theories reflects their overall truthlikeness and therefore at least their likely performance in *future* tests. (It is, after all, perfectly possible given simply deductive considerations that theories that have performed relatively badly in the past will, in the future, perform better than ones that have performed relatively well so far.) It seems that we need, then, some sort of link between past performance in tests and overall truth (or at least overall empirical adequacy). But this is exactly the sort of inductive assumption that was anathema to Popper.

Difficulties with falsifiability – the Duhem problem

Moreover, fundamental issues also arise about the assumed falsifiability of scientific theories. The most direct problem here had already in fact been explained in impressive detail some thirty years before Popper's work by Pierre Duhem (1906). Scientists often talk about testing scientific theories, such as Newton's theory (of mechanics and gravitation) by comparing that theory's predictions – about, say, planetary positions – with the "data." But Duhem pointed out that if the deductive structure of any such test is analysed carefully then further premises – often called "auxiliary assumptions" – always turn out to be necessary if the deduction of the observation statements at issue is really to be valid. Nothing that we are likely to characterize as a "single theory" in science – Newton's theory or Maxwell's theory of electromagnetism or quantum theory or whatever – has any empirical consequence when considered "in isolation," further auxiliary assumptions are always needed. For example, no consequences about planetary positions at some given time t follow from Newton's theory (of mechanics plus universal gravitation) and nor do they follow from Newton's theory plus "initial conditions" about the positions of those planets at some earlier time t'. What is needed, in addition, is a whole set of other assumptions that are clearly themselves theoretical rather than in any sense "directly given" by observation – this set includes assumptions, for example, about the mass of the planet concerned and the number and masses of the other bodies in the solar system, not to mention assumptions about how light travels between the planet concerned and our telescope. (So, in particular, a – clearly theoretical – assumption is needed about the extent to which light is refracted in passing from "empty space" into the earth's atmosphere.)

This apparently minor logical point has major consequences. Suppose we have some observation sentence O and are happy to say that we can decide the truth value of O on the basis of observation or experiment. If contrary to Duhem, we could invariably take any "single" scientific theory T and deduce a range of such results O from it, then, just as Popper emphasized, if some such O were established as false on the basis of observation, then it would follow that T must be false as well. (The so-called "principle of retransmission of falsity" says that if some premise, in this case the theory T, entails deductively some conclusion, in this case the observation sentence O, then if that conclusion is false, so also must be the premise.) In fact, however, as Duhem's analysis showed, the deductive structure of any real test of any real scientific theory always involves auxiliary assumptions – often quite a large set of them. But if we can infer O only from a conjunction of sentences $T\&A_1\& \ldots \&A_n$, then should we decide, on the basis of observation or experiment, that O is false, all that we can infer is that *at least one of* the set of theoretical claims $T, A_1, \ldots, A_n$ is also false. (The principle of retransmission of falsity when applied to deductive inferences with more than one premise does not, of course, say that if the validly deduced conclusion is false, then so are all the premises, but only that *not all* the premises can be true – at least one must be false.) In particular, we cannot infer that it is T itself that is false.

Duhem's analysis does *not* show that observation results never supply good grounds for holding that some "central" theory T is false; but it does show that these are never conclusive and that something more than falsification must be involved. There might, for example, be independent grounds for thinking that the auxiliaries $A_1, \ldots, A_n$ are more likely to be true than is T. If so, then the fact that the falsity of O shows that not all of $T, A_1, \ldots, A_n$ can be true would supply good grounds for rejecting T. Or, and this is what generally in fact happens in cases of scientific theory-change, while a theoretical system built around theory T can be made to yield O only by adjusting some of the auxiliaries A_i exactly with the requirement in mind that O be entailed, an alternative system built around some alternative theory T' involving non *ad hoc* auxiliaries is independently empirically confirmed (that is turns out to predict some further empirical result O' which is then confirmed). Either suggestion, however, brings in ideas of confirmation that are foreign to Popper's scheme.

Confirmation – the attempt at an "objective" account

Why not then go straight for confirmation as the solution to the problem of what makes science special? The Einsteinian revolution was a constructive proof of the fact – which in any event ought in retrospect to have been obvious – that we can never conclusively *prove* general explanatory scientific theories on the basis of observation or experiment; Duhem's analysis showed that we can never conclusively *falsify* them either. But perhaps we can nonetheless *confirm* scientific theories on the basis of empirical results. Perhaps what distinguishes a better scientific theory from a good one is that the former is better confirmed by the evidence; perhaps what explains "revolutionary" shifts in scientific theory – for example, that from Newton to Einstein – is exactly that, given the evidence that had accumulated, Einstein's theory was the better confirmed theory; and finally perhaps what distinguishes scientific theories from non-scientific ones (whether metaphysical or pseudoscientific) is that the latter are not even capable of empirical confirmation. The claim that "God exists" fails to be scientific, not because it cannot be proved from evidence, not because it can never be falsified by evidence, but because it can never be confirmed (and therefore can never be disconfirmed either) by any possible – intersubjectively agreed – evidence.

As a general framework suggestion, this answer still seems to me viable (indeed perhaps when considered in a very general way, it is the only viable answer). The problem has been that of giving a more precise account of the notion of "confirmation" – a more precise account that delivers all the above judgments and that seems both coherent and philosophically defensible.

A number of "non-standard" approaches have been tried (perhaps most notably Clark Glymour's (1980, 1987) "bootstrapping" approach), which have run into their own difficulties. But most attempts to put flesh onto the skeleton of the confirmation approach have, unsurprisingly, involved the notion of probability. What confirmation delivers, it is suggested, is greater probability of being true: the change from Newton to Einstein was the change from one reasonably probable theory (of course probable in the light of the evidence) to another that is still more probable in the light of the evidence; theory-change in science can be explained as rational because in the light of accumulating evidence the relative probabilities of rival theories naturally change; and finally, non-scientific theories are those whose probability cannot be affected one way or the other by the evidence.

Although there are intimations of the approach much earlier in the history of thought, recent discussions of this idea really stem from Carnap's (1950) groundbreaking work. His initial idea was to produce an entirely objective version of the account by developing a probabilistic "inductive logic" as a generalization of deductive logic. The crucial notion in all accounts is the probability of some theory given (or conditional on) some evidence. Carnap's original idea was that such conditional probabilities measure degrees of partial entailment – to claim that the probability that Einstein's theory is true, given the evidence is, say, 0.8 means that the evidence entails Einstein's theory to degree 0.8. (Here full – deductive – entailment would of course be degree 1, that is, the probability of *A* given *B* is 1 whenever *B* deductively entails *A*.) This idea might then be used to supply the rationale for scientific revolutions if it could be shown that the newer theory – say Einstein's theory – has higher probability in the light of the evidence available at the time of the "revolution" than had the earlier theory – in this case Newton's. Intuitively, although the evidence of course entails neither theory, it comes closer to entailing Einstein's theory than to entailing Newton's.

This idea, for all its simplicity and appeal, fails. The basic problem is essentially the same as the one that afflicts the so-called classical account of probability – which defines the probability of some event A as the ratio of the number of "equally possible" cases in which A holds to the number of all the equally possible cases. (Intuitively the probability that a fair dice when rolled will finish with "6" up is $\frac{1}{6}$ because there are six equally possible cases and just one in which the event "6 up" occurs; the probability that an even numbered face will be uppermost in the same situation is $\frac{3}{6}$, i.e. $\frac{1}{2}$ since there are again six equally possible outcomes and in three of them the event "even number uppermost" is instantiated.) The difficulty concerns the notion of partitioning the set of all the possible events in some experiment into the "equally possible" ones. In general, there are different ways of doing this and it seems impossible to argue that only one such way is "correct." And yet with a different partition of the events into equally possible cases we arrive at different probabilities.

Although this approach and this difficulty for it were originally developed in the context of probabilities of various *events*, an entirely analogous approach, and an entirely analogous difficulty, can be developed when thinking, as Carnap did, of the probability that a particular *sentence* is true. Suppose, for example, we are interested in hypotheses about the contents of an urn known to contain, say, 50 balls, each of which is either black or white but in an unknown proportion; suppose further that we are (for some reason) unable to break open the urn and our evidence is restricted to drawing some number of balls from the urn, with replacement, and noting their colours. What constitute the equally likely cases here? All possible *proportions* of black to white balls – all 50 black, 49 black 1 white, 48 black 2 white, etc.? Or are the equally likely cases specified by assuming that each individual ball has the same chance of being white as of being black? It seems difficult indeed to argue that one of these notions is the "correct" one. But it is no surprise that the two yield quite different probabilities for various hypotheses. Suppose we are interested in the hypothesis that exactly half of the balls are white and our evidence is that we have drawn 10 balls, 6 of which are white. The inductive support given to that hypothesis by that evidence, the degree to which the evidence partially entails the hypothesis, will be quite different depending on which of these two ways we slice up the "equal possibilities"; and this makes it very difficult to claim that there is one objective probability for the hypothesis in the light of the evidence.

Confirmation – the Bayesian account

This and a range of other problems led those pursuing the idea that "confirmation is probability" – including eventually Carnap himself – to abandon this "objectivist" partial entailment approach. The currently most popular version of this general idea takes the probabilities at issue in confirmation theory in fact to measure simply a person's degree of belief in the proposition at issue. An agent is considered to have degrees of belief in every proposition available to her and in every logical combination of such propositions. Such an agent is "rational" if

(i) *at any given time*, those degrees of belief can be represented as probabilities (that is satisfy the probability calculus) and
(ii) changes in her degrees of belief *from one time to the next* satisfy something called the "principle of conditionalization."

Although a thoroughly subject- (or agent-)based approach, this account does have clear objective elements. For example, condition (i) requires that if an agent's degree of belief in the theory that the initial escape velocity of matter from the big bang was v_1 is d_1, while her degree of belief in the theory that the initial escape velocity of matter from the big bang was v_2 is d_2, then (assuming that she – properly – believes that it is not possible for the escape velocity to have both values!), she must believe that the theory that the escape velocity was *either* v_1 *or* v_2 to degree $d_1 + d_2$. Also, if an agent has degree of belief d in some proposition P then she must have a degree of belief d' *at least as high* as d in any proposition Q that is a logical consequence of P.

Defenders of this view have produced various arguments for why condition (i) should be considered an absolute requirement on rationality. The most often-cited argument proceeds by identifying an agent's degrees of belief with fair betting odds (the worst odds at which the agent would be ready to bet on the proposition's being true) and showing that if those degrees of belief were *not* probabilities, did not satisfy the probability calculus, then the agent would be committed to accepting as fair a system of bets such that she would be bound to make a net loss, whatever way the world turned out to be (that is, which ever sentences were eventually accepted as true). This is the so-called "Dutch Book Argument."

A crucial notion in this approach is the *conditional probability* $p(a|b)$ – the probability that a holds on the assumption that b does. These are, of course, interpreted as measuring what your degree of belief in *a would be* if you came to accept *b*. The most important such conditionals for a theory of confirmation will of course be of the form $p(T|e)$ where T is some theory and e some statement that can be checked on the basis of observation or experiment. Principle (ii) in this impressively austere approach then says something like the following. Suppose that all that happens of any epistemic relevance concerning some particular theory T between two successive stages in science t_1 and t_2 is that some empirical statement

e that is simply *potential* evidence at t_1 has been checked and actually found to hold (that is, has become real evidence, an accepted part of "background knowledge") by time t_2. How should the agent's degrees of belief in T at times t_1 and t_2 be related? Given the understanding of $p(T|e)$ as measuring the degree of belief in T that you would have if you were to come to know e, advocates of this approach have suggested that it is obvious that the agent's "new" degree of belief in T at t_2 should be her "old" degree of belief in T conditional on e. That is, introducing subscripts on the probabilities for clarity;

$$p_{t2}(T) = p_{t1}(T|e)$$

And this is the "principle of conditionalization."

Conditional probabilities like $p(T|e)$ are calculated using Bayes' theorem, which, in its simplest form, says

$$p(T|e) = \frac{p(T)p(e|T)}{p(e)}$$

Because of the frequent use of Bayes' theorem, the approach we have been discussing is called the Bayesian approach to theory-confirmation, or – for reasons made clearer shortly – the personalist Bayesian approach.

Bayesianism has a number of pleasing features. First, as already mentioned, it is impressively austere, appearing at any rate to define "inductive rationality" via only two assumptions. Second, it gives a gratifyingly simple account of what it takes for a theory to be confirmed by evidence: e confirms T just in case e raises T's probability, i.e. just in case $p(T|e) > p(T)$. And third, it is easy to see that this simple account captures a number of firmly entrenched intuitive judgments about confirmation. It is, for example, part of scientific folklore that if a theory passes a "severe test" (in Popper's terminology) then this confirms the theory more highly than would a less severe test – where a test is severe to the extent that its outcome is highly improbable in the light of background knowledge. One frequently cited example here is the prediction by Fresnel's wave theory of light that if the "shadow" of a small opaque disk held in the light emerging from a point source is carefully examined then the centre of the "shadow" will be seen to be illuminated, and illuminated indeed to precisely the same extent as it would have been had no opaque object been interposed. The usual story is that the idea that there should be such a "white spot" was so improbable in the light of background knowledge, that, once Poisson had shown that Fresnel's theory implied its existence, the scientific establishment was fully confident that Fresnel's goose had been cooked. The account of confirmation under consideration, using Bayes's theorem, straightforwardly captures this intuition. According to the Bayesian formula, the extent to which e confirms T (i.e. the difference between $p(T)$ and $p(T|e)$) is greater the smaller is $p(e)$ – i.e. the less likely e is according to background knowledge. (Remember that any probability lies in the interval $(0,1)$.)

Virtues like these, combined with major difficulties in alternative approaches, have convinced many contemporary commentators that Bayesianism is essentially "the only game in town" when it comes to providing a clear-cut, formal theory of confirmation (as opposed to simply some unsystematic list of intuitive judgments about theory-evidence relations). If so, then the only game in Confirmation Town leaves philosophers of science with a lot of work to do in adding to its rules.

Problems with Bayesianism

Of the difficulties facing the personalist Bayesian approach, I outline here one relatively specific "internal" problem and one issue that seems to me a major, general difficulty for the whole approach. The more specific difficulty has come to be known as the "problem of old evidence." There has been much discussion in philosophy of science going back to debates between John Stuart Mill and William Whewell (and beyond) about the relative confirmational value of a theory's predicting hitherto unknown "new" evidence and of its simply explaining already known "old" evidence. Certainly, many of the great confirmational successes for theories that are much heralded in the scientific folklore were predictions: the wave theory of light and the "white spot" at the centre of the "shadow" of an opaque disk (already mentioned) is one such example, and the prediction by the theory of general relativity of star shift (that stars would seem to be different distances apart during the day because of the gravitational effect of the sun) confirmed by Eddington's Eclipse Expedition is another. However, although there may well be some sort of special psychological effect of predictive success, it is difficult to see any principled reason why the time-order of theory and evidence should count *in itself*. Moreover, there are definitely cases where "old evidence" strikingly confirmed a theory – indeed confirmed it, in the eyes of the scientific cognoscenti, just as strongly as any piece of predicted "new" evidence could. Funnily enough, two such cases match the predictive successes just mentioned: Fresnel's explanation of straightedge diffraction (a phenomenon known for around 150 years when Fresnel proposed his theory) seems to have played just as strong a role as the "white spot" evidence in the acceptance of his theory; and, certainly, general relativity's success in accounting for the long-known "anomalous" precession of Mercury's perihelion counted for at least as much as its success with the "star shift." It seems clear that, whatever the truth about the "prediction versus accommodation" issue, it cannot be a blanket "old evidence always counts less." Yet, the Bayesian account of confirmation seems to yield the even stronger result that old evidence can *never count at all*.

This can be seen very easily from the Bayes formula and the fact that all probabilities in this approach are always implicitly relative to background knowledge – that is, to what we already take ourselves to know, at whatever stage of science we are considering. But if some piece of evidence e is "old" – already known, in background knowledge at some time t – then its probability at t, relative to that background knowledge, must of course be one. It follows, however, from Bayes formula and assuming that T deductively entails e so that $p(e|T) = 1$, that if $p(e) = 1$, then $p(T|e) = p(T)$. And that precisely means on the Bayesian account that e fails to confirm T.

There have been suggestions from its defenders for how this "old evidence problem" might be solved within the Bayesian framework, though none has won widespread assent. The more general problem seems to me, however, to have no possible solution within the purely personalist framework, but to require – at least – a major extension of it. The problem is that the Bayesian approach seems clearly too weak, to allow too wide a role to subjective opinion, to have any chance of capturing fully what is special about science.

Consult again the crucial Bayesian formula. The Bayesian agent is taken to be a perfect deductive logician, so that if T deductively entails e (usually *modulo* background knowledge) then she must assign a value of 1 to the term $p(e|T)$ – and similarly if T is a well-defined probabilistic hypothesis then she must assign whatever probability T – objectively – assigns to e. The other terms in the formula are however taken to be agent-relative. In particular, the so-called *prior probability* of T, $p(T)$, measuring the degree of belief that an agent has in the theory T ahead of whatever evidence we are now proposing to take into account is subjective – there is no truth of the matter as to what this prior probability is, the Bayesian simply takes it as a fact about a particular agent that she has a certain degree of belief.

It is true, of course, that, in applying this apparatus to some particular theory as it and the evidence for it develop over time, the Bayesian will usually tell a story of how the current prior for T is the end result of a series of applications of the principle of conditionalization on earlier pieces of evidence. But, even then, this series will, of course, have started with some initial prior which will then, by definition, be "purely subjective." Bayesians cite various interesting theorems about the "washing out" of priors which show that, in certain circumstances, two agents with radically different priors on some theory T will nonetheless converge to the same probability for T as evidence of certain kinds comes in. The fact however that in such circumstances (which may not in any event match real cases) any two agents will, in the – of course never actually attained – limit, agree hardly seems sufficient to capture what we generally think of as scientific rationality.

It will surely be generally agreed that, given all the evidence that we currently have from the fossil record, homologies, and various experiments, not to mention the results of various dating techniques, that the Darwinian theory of evolution together with its view of the earth as extremely ancient is altogether more rationally believable *now* than the "scientific" creationist view that the earth was created essentially as it now is, stocked with essentially the "kinds" that it currently has, in 4004 BC. If ever there was a non-defeasible desideratum on an adequate account of the relationship between scientific theories and evidence this is surely it. Yet, it is trivial to show that given any relative degrees of belief in Darwinism (D) and

Creationism (C) – say p(D) = 0.000001 and p (C) = 0.999999 – it is entirely possible for an "agent" to have arrived at those degrees in full accordance with Bayesian principles. She could have conditionalized away on all the evidence and still have arrived at degrees of belief that any satisfactory account ought surely to brand as absurd. Of course, this will require the supposition that the agent started the process – ahead of the consideration of any evidence – with even more extreme priors. But the personalist Bayesian explicitly eschews any restrictions on these priors. Any proof that such a "scientific" Creationist is bound to agree with us Darwinians in the indefinite long run is no consolation – it seems clear that the creationist holds a view *now* that is counter to good scientific reasoning, and the Bayesian just cannot deliver that judgment.

The way forward?

Here, then, is a problem that, in my view, remains very much open to future research. Personalist Bayesianism seems at best to capture only a part of scientific rationality. It needs to develop and to defend further requirements – placing at least restrictions on acceptable priors. It is by no means clear how this is to be done, however, within a genuinely Bayesian context. The alternative of course would be to develop another "game in town" – another different systematic attempt to capture good scientific confirmational practice in a precise, and philosophically defensible, way.

One – altogether more radical – suggestion that has been taken up by many recent philosophers is that the sort of approach embodied in Bayesianism and similar enterprises involves an entirely mistaken set of aims and priorities. According to the currently (and increasingly) strong movement towards a "naturalized" philosophy of science, philosophers have for too long been obsessed with traditional issues bequeathed to us to by the likes of Descartes and Hume. We should not be looking for anything like a *logic* of science or of scientific confirmation. Any such system would, in any event, itself rest on assumptions (assumptions which moreover must certainly go beyond deductive logic); and, as centuries of philosophy ought to have taught us, we should be powerless against the sceptic who then asks for justification of those principles themselves. We cannot ask for, and so should not seek, any firmer ground than science itself on which to build our epistemological claims. Philosophy of science should be pursued in a naturalized, scientific way, simply recording the methods of science.

The naturalizing movement with its greater emphasis on philosophers knowing about the details of science has undoubtedly led to many significant improvements. (Though it has to be said that it is easy to get the – of course, absurd – impression from recent treatments that earlier philosophers (the likes of Reichenbach, Hempel and Popper, not to mention still earlier figures like Poincaré and Duhem) knew nothing of the details of science!) Following Kuhn (1962), Lakatos (1970) and others, we now have a much more nuanced view of scientific theoryconstruction; we have a much richer set of descriptive tools for analysing science and its development involving models, idealizations and the like and a better understanding of the intricacies of scientific "observation." But, as for the general idea that a fully naturalized view can somehow establish the specialness of science, without any rate *vicious* circularity, by itself adopting a scientific approach – this seems to me a very difficult line to argue. The problem again remains an open one for future investigation.

Accumulation in Science, Despite "Revolutions"?

I explained at the beginning how the Einsteinian revolution turned Kant's problem into a dilemma. So far, we have been investigating the prospects for a program that admits the revolutionary nature of scientific change and tries, none the less, to rescue the epistemic specialness of science. Attempts to escape the *other* horn of the dilemma involve conceding that the idea of scientific rationality would indeed be in deep trouble if scientific development were as "revolutionary" as it might at first appear to be and therefore accepting the challenge of arguing that once science, and in particular scientific theories, are *properly* understood, the revolutionary nature of scientific theory-change disappears (or perhaps "largely" disappears). We should now investigate this second possibility.

Revolution in permanence? The "pessimistic meta-induction"

First, let's be clear about the extent of the apparent difficulty. As many commentators would see it, the relativistic and quantum revolutions are simply the tip of the iceberg and their chief effect ought to have been to take off the blinkers so that philosophers could see that "revolutions" (of varying degrees of magnitude) are, in fact, ubiquitous in science. Long before the turn of the century, and even allowing for the sake of argument that science only really started with "the" Scientific Revolution, there had been plenty of less well-publicized but none the less definite cases of seemingly radical theory-change in science. Consider, for example, the history of optics – even when restricted to the modern era. In the eighteenth century, the theory that light consists of material corpuscles had been widely accepted only to be replaced in the early nineteenth century by the theory that light sources do not emit matter but rather energy – matter within the light source vibrates and causes the neighboring particles of the all-pervading "luminiferous ether" to vibrate and hence these vibrations spread through the ether until absorbed by some receptor or other (such as the human eye). This theory, in turn, was replaced by what might be called the mature version of the Maxwell electromagnetic theory of light that denies the existence of the mechanical ether and attributes light instead to the "vibrations" of electric and magnetic field vectors.

Then, of course, as part and parcel of the quantum revolution came the photon theory with its probability waves. From particles to vibrations in an elastic solid, to changing strengths of a *sui generis* electromagnetic field, to photons governed by probability waves – these seem radical shifts indeed. And, of course, according specially to Kuhn (1962), similar revolutions took place in all other branches of science too.

Instances of revolutionary change supply the premises for the "pessimistic meta-induction" that has received a good deal of attention in philosophy of science in the past few decades. This argument is simply an elaboration of the problem from which I began. It is surely a characteristic of revolutionary theory-change that the new theory contradicts the old so that, if we assumed for the sake of argument that the new theory were true, we would be forced to the conclusion that the older theory was false. But what possible grounds could we have for thinking that scientific revolutions are now at an end – that we now have the final theories in all scientific fields? Newtonians in the eighteenth and nineteenth centuries believed – on the basis of very strong evidence – that the fundamental truth about the universe had been discovered; and they turned out to be wrong. No physicist in the nineteenth century, again on good evidential grounds, dreamed that the fundamental processes in nature might be inherently probabilistic and yet that, according to most views, is precisely what presently accepted theories are telling us is true. As we saw, theories of the basic constitution of light have undergone radical shifts. From the standpoint of the current photon theory, the theory that light consists of vibrations transmitted through an all-pervading elastic solid ether looks about as false as any theory could be – after all, for one crucial thing, the newer theory denies entirely the existence of such an all-pervading mechanical ether.

How, then, can we have any faith that that currently accepted photon theory will not, in its turn, eventually be replaced by a theory in whose light it will appear just as false as it itself makes the classical wave theory appear? And, if the findings of science are, at this fundamental level, as transient as this account makes them seem, how can we have any confidence in the process? Even if we could produce persuasive arguments for the methods of science as characterizing a rational process – that is, even if we could solve the problems sketched earlier with, say, some de luxe theory of confirmation – then even so, if that "rational" process produces conclusions that are subject to periodic radical, chalk-and-cheese change, it seems difficult to see why we should regard science as so special.

Notice that no one is asserting here that the "pessimistic meta-induction" is by any means a compelling argument – it is after all inductive and not deductive. It is perfectly *possible* that our scientific predecessors were unlucky (or misguided) and that we have now hit on the truth. And indeed the intuition underwriting the programs discussed earlier is exactly that science, and scientific theories, have improved over time. But it is difficult to see that improvement as in any sense qualitative – nineteenth century physicists had a good deal of evidence for their theories. We now have a good deal more evidence in the light of which very different theories seem true. But, then, since science will presumably continue to "improve" and evidence continue to accumulate, what grounds could be given for holding that our current theories will resist radical change in the light of that accumulating evidence? The pessimistic meta-induction does not need to establish that we have good grounds for thinking that our current theories will eventually be "radically" replaced; the weaker conclusion that we have no good grounds for thinking that they will not be so replaced is sufficient to pose the problem.

Resisting "pessimism" by restoring an essentially cumulative view

Instrumentalism So how, then, could philosophers of science before 1962 have been blind to what ought to have stared them in the face? The answer is that many of them at least were not at all blind to this phenomenon. Although we tend to think of the "pessimistic meta-induction" as a new philosophical argument, starting with Hilary Putnam or Larry Laudan (1981), in fact it can be found fully formed in Poincaré's (1905) *Science and Hypothesis*:

> The ephemeral nature of scientific theories takes by surprise the man of the world. Their brief period of prosperity ended, he sees them abandoned one after the other; he sees ruins piled upon ruins; he predicts that the theories in fashion today will in a short time succumb in their turn, and he concludes that they are absolutely in vain. This is what he calls the *bankruptcy of science*.

As the way he introduces it suggests, Poincaré was not only aware of the problem he was confident that he had an answer to it:

> [The "man of the world's"] scepticism is superficial; he does not take account of the object of scientific theories and the part they play, or he would understand that the ruins may still be good for something. No theory seemed established on firmer ground than Fresnel's which attributed light to the movements of the ether. Then if Maxwell's theory is preferred today, does it mean that Fresnel's work was in vain? No, for Fresnel's object was not to know whether there really is an ether, if it is or is not formed of atoms, if these atoms really move this way or that; his object was to predict optical phenomena.

Underneath the apparently radical theory-changes (producing the seeming "ruins") there is, Poincaré suggests, a steady accumulation of "real" knowledge in science.

There are two importantly different versions of this claim – versions which Poincaré himself did not *always* clearly differentiate (though I think there is, in the end, no doubting his preferred position). As it stands, the last part of this quotation suggests an "instrumentalist" view of science. Scientific theories, like Fresnel's theory of light, may *seem* to make true-or-false assertions about the underlying structure of reality, about material ethereal atoms held in place by elastic forces and about the vibrations of those atoms, which we cannot of course directly observe, but which allegedly constitute light and hence explain the optical phenomena that we *can* observe. However, the real role of scientific theories is not even to attempt to describe a reality "underlying" the phenomena, but instead merely to codify those phenomena in a coherent, efficient and "simple" way, and hence to enable their prediction. And at the level of "phenomena" – the results of experiments, such as various interference, diffraction and polarization experiments – Maxwell's theory, while attributing those phenomena to a radically different process, none the less agrees (exactly) with Fresnel's theory. Maxwell's theory, of course, goes on to make further predictions – about, for example, radio waves; but where the two theories both make empirical predictions they always exactly agree.

There has been some discussion in the literature of so-called "Kuhn loss" of empirical content – (alleged) cases where some observational or experimental result correctly accounted for by the deposed theory in some "revolution" is not correctly accounted for by the newer theory. Kuhn's own examples of this alleged methodological phenomenon are entirely unconvincing. There are undoubtedly cases in the history of science where a new theory is accepted despite the fact that it cannot at that stage account for some already known phenomenon and where the older theory (which has, of course, at that stage the advantage of longevity) gave at least some sort of account of that same phenomenon. A good example from optics is prismatic dispersion – according to the simplest models of the elastic solid (or indeed elastic fluid) ether, all waves, no matter what their frequency, would travel through it at the same velocity and yet the phenomenon of prismatic dispersion (exhaustively studied, of course, long before Fresnel's wave theory by Newton and others) establishes that the different monochromatic components of solar light travel through the material of the prism (usually glass) at different velocities. The corpuscular theory of light, deposed in the early nineteenth century "wave revolution," gives hints of an explanation – for example a "fixed force of refraction" with the different monochromatic rays corresponding to particles with different masses. But this explanation was known to run into enormous difficulties. If there are any genuine cases of "Kuhn loss" in which some phenomenon was *satisfactorily* explained by the pre-revolutionary but not by the post-revolutionary theory, then they are few and far between. Moreover, it is of the nature of science that any "losses" would be high on the agenda for work aimed at making them good. This is true even where the older "explanation" is highly flawed – in the example just discussed, for instance, a central thrust of the wave optics research program after Fresnel was precisely to develop a detailed mechanical account of the ether that yielded dispersion.

It seems difficult to deny, I suggest, that the development of science has been, at least to a very good approximation, cumulative *at the observational or experimental level*. This need not mean that the "post-revolutionary" theory has exactly the same empirical consequences as the pre-revolutionary one (though in a restricted domain). That happens to be true in the Fresnel–Maxwell case cited by

Poincaré, but the more usual pattern is the one exemplified in the shift from Newtonian classical to Einsteinian relativistic physics. Every *precise* observational consequence of special relativity theory is strictly inconsistent with the corresponding observational consequence of classical theory. Those conflicting observational consequences, none the less, explain the same data across a wide range, because they are, within that range, *observationally indistinguishable*. It follows then that the apparently radical theory changes brought about by "scientific revolutions" pose no problem for the instrumentalist – as concerns what that account sees as the real purposes of science, there is essential continuity across scientific change. Science is special because it delivers more and more of the epistemic "goods" – it is just that those goods do not consist of ever deeper, ever "truer" explanatory theories but rather of ever wider codifications of ever more phenomena.

An interesting more recent slant on this old position is provided by Bas van Fraassen's (1980) "constructive empiricism." Van Fraassen countenances no positivist reduction of the theoretical claims of science – if a theory asserts that electrons exist, it asserts they exist: the claim cannot be regarded as merely shorthand for some complicated set of observational sentences or as some sort of non-assertive "inference licence"; and such a theory is either true or false (in the regular Tarski correspondence sense) depending on how the world really is. However, to explain the rationality of what goes on in science, there is no need to involve considerations of whether such a theoretical claim *is* true (indeed as we have been seeing such involvement poses major problems for ideas about rationality). Scientists should be seen as "accepting" theories, not as true, but only as *empirically adequate*. Although van Fraassen does not directly address the issue of (apparently) radical theory-change, his position provides the basis for a response identical to the one just considered – the progress of science through theory-change can be seen as the development of ever more empirically adequate theories, each new theory revealing that its predecessor was indeed highly empirically adequate but over a restricted range.

Although I shall not discuss them here, there are, of course, many problems with this instrumentalist view – all of them associated in one way or another with the fact that the view does not seem to give proper weight to the role of theory, especially in the *development* of science.

Resisting "pessimism" by restoring an essentially cumulative view

Positivism and structural realism Instrumentalism, at least in the way I am interpreting it here, allows that successive theories in science contradict one another, and hence allows that theory-change leaves "ruins" (to use Poincaré's term) in its wake. The instrumentalist insists, however, that there is none the less accumulation at the level that science is really all about – the codification of phenomena. There are ruins but they are insignificant.

A different view – a version of which Poincaré himself in fact adopted – is that, when properly viewed, *there are no ruins.* Once the cognitive content of scientific theories is correctly analyzed, we see that the apparent ruins are just that – (merely) *apparent.* It may seem as though Fresnel's theory, for example, makes ontological claims about a medium with the constitution of an elastic solid pervading the whole of space and about the particles of that medium vibrating in certain ways. In fact, however, when we understand properly what the theory says we see that this is not really the case.

One extreme version of this general line is, of course, an outright empiricism or positivism. This sees the real cognitive content of a "theoretical" claim as somehow "reducing to" some (infinite) set of observation sentences. In the case of Fresnel's theory, for example, all the apparent theoretical talk about ether particles, in fact, "reduces to" assertions about interference and diffraction patterns and the like. The logical empiricists did not, in fact, pay much direct attention to theory-change, and developed their account of theories to solve different problems. But if their account could have been made to work, then clearly the phenomenon of theory-change would present it with no problem, assuming that, as I have claimed, the development of science is essentially cumulative at the empirical level.

It has for a long time now been very widely accepted that any such empiricist account is untenable. Certainly, various particular attempted reductive analyses did not work; and the general view, as in the case of instrumentalism, is that no such account can do real justice to the role of theory, particular its heuristic role in the development of science.

The account that Poincaré himself endorsed is different (at least pre-analytically) from both instrumentalism and empiricism or positivism. Having said the Fresnel's theory was not in vain despite its displacement by Maxwell's, because it still allows us to predict optical phenomena as before, he elaborates as follows:

> The differential equations [in Fresnel's theory] are always true [that is, they are carried over into Maxwell's theory], they may always be integrated by the same methods and the results of this integration still preserve their value.
>
> It cannot be said that this is reducing physical theories to practical recipes; these equations express relations, and if the equations remain true, it is because the relations preserve their reality. They teach us now, as they did then, that there is such and such a relation between this thing and that; only the something which we then called *motion*, we now call *electric [displacement] current.* But these are merely the names of the images we substituted for the real objects which Nature will hide for ever from our eyes. The true relations between these real objects are the only reality we can attain . . . (Poincaré, 1905).

Hence, Poincaré claims a continuity across theory-change in science that extends not merely to the observational, but also to the structural level – as is evinced, at any rate in the case he discusses, by the retention of the mathematical equations (and hence of the observational consequences). All that is "lost" are preferred "names of images." The real cognitive content is preserved entirely in tact.

Another problem that seems to me still very much an open one for current philosophy of science is whether some version of Poincaré's *structural realism* can be elaborated, extended to all cases of theory-change and be shown to avoid collapse into outright empiricism. If not, is there any serious hope for any form of scientific realism? The idea that one can retain the view that Newton's theory may be "approximately true" despite the Einsteinian revolution seems to me implicity to presuppose some such (apparently) reduced form of realism. Otherwise, at the "ontological" level, we do seem to have not approximation but outright rejection (of absolute space, absolute simultaneity, action-at-a-distance and so on).

Other Issues

I have tried to build my introductory account of some central issues in philosophy of science around a theme. But, just as I said it would from the outset, any such thematic treatment is bound to leave out much of value. I have not touched on some central issues – such as scientific explanation, the notion of causality and others. Many of these will be dealt with in what follows. I have also not been able to discuss those very important areas of philosophy of science which overlap with theoretical work in the sciences themselves. Analyses of conceptual issues in the theory of general relativity, quantum mechanics and statistical mechanics have all been at the forefront – and have, in turn, raised in especially sharp ways general philosophical issues about determinism, locality and the like. More recent work has seen an extension into the foundations of biology – particularly the structure of Darwinian theory and of genetics; and, especially via interest in causal models, into the foundations of the social sciences.

References

Carnap, R. (1950): *Logical Foundations of Probability*. Chicago: University of Chicago Press.

Duhem, P. (1906): *The Aim and Structure of Physical Theory*, English translation, Princeton: Princeton University Press, 1954.

Glymour, C. (1980): *Theory and Evidence*. Princeton: Princeton University Press.

Glymour, C., Scheines, R., Sprites, P. and Kelly, K. (1987): *Discovering Causal Structure: Artificial Intelligence, Philosophy of Science and Statistical Modeling*. New York: Academic Press.

Kuhn, T. S. (1962): *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Lakatos, I. (1970): "Falsification and the Methodology of Scientific Research Programmes," in I. Lakatos and A. Musgrave (eds.), *Criticism and the Growth of Knowledge*, Cambridge: Cambridge University Press, 91–196.

Laudan, L. (1981): "A Confutation of Convergent Realism," *Philosophy of Science*, 48, 19–49.

Poincaré, H. (1905): *Science and Hypothesis*, English translation, New York: Dover Books.

Popper, K. R. (1959): *The Logic of Scientific Discovery.* London: Hutchison.

Van Fraassen, B. (1980): *The Scientific Image.* Oxford: Oxford University Press.

Further reading

Boyd, R. (1973): "Realism, Underdetermination and the Causal Theory of Evidence," *Nous*, 7, 1–12.

Cartwright, N. (1989): *Nature's Capacities and their Measurement.* New York: Oxford University Press.

Earman, J. (1989): *World Enough and Space-Time.* Cambridge: MIT Press.

Earman, J. (1992): *Bayes or Bust? A Critical Examination of Bayesian Confirmation Theory.* Cambridge: MIT Press.

Earman, J. and Glymour, C. (1980): "Relativity and Eclipses: The British Eclipse Expeditions of 1919 and their Predecessors," *Historical Studies in the Physical Sciences*, 11, 49–85.

Howson, C. and Urbach, P. (1993): *Scientific Reasoning: The Bayesian Approach, second edition.* Chicago: Open Court.

Kitcher, P. (1985): *Vaulting Ambition: Sociobiology and the Quest for Human Nature.* Cambridge: MIT Press.

Mayo, D. G. (1996): *Error and the Growth of Experimental Knowledge.* Chicago: University of Chicago Press.

Popper, K. R. (1963): "Science: Conjectures and Refutations," in *Conjectures and Refutations*, London: Routledge and Kegan Paul, 33–57.

Redhead, M. L. (1987): *Incompleteness, Nonlocality and Realism: A Prolegomenon to the Philosophy of Quantum Mechanics.* Cambridge: Cambridge University Press.

Sklar, L. (1993): *Physics and Chance: Philosophical Issues in the Foundations of Statistical Mechanics.* Cambridge: Cambridge University Press.

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

Worrall, J. (1989): "Fresnel, Poisson and the 'White Spot': The Role of Successful Prediction in Theory-acceptance" in D. Gooding, T. Pinch and S. Schaffer (eds.) *The Uses of Experiment*, Cambridge: Cambridge University Press, 135–57.

Worrall, J. (1999): "Two Cheers for Naturalised Philosophy of Science," *Science and Education*, 8(4), July, Special Edition.

The Blackwell Guide to the Philosophy of Science

Edited by

Peter Machamer and Michael Silberstein

BLACKWELL
Publishers

First published 2002

2 4 6 8 10 9 7 5 3 1

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

Blackwell Publishers Ltd
108 Cowley Road
Oxford OX4 1JF
UK

Library of Congress Cataloging-in-Publication Data has been applied for.

ISBN 0-631-22107-7 (hardback); 0-631-22108-5 (paperback)

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

Typeset in 10 on 13 pt Galliard
by Best-set Typesetter Ltd., Hong Kong
Printed in Great Britain by T.J. International, Padstow, Cornwall

This book is printed on acid-free paper.