29
THEORY-CHANGE IN SCIENCE

John Worrall

Introduction

According to an historical sketch enjoying wide circulation, once upon a time, in the "bad old days" of logical empiricist hegemony, philosophers of science believed that the progress of science is cumulative. When a new scientific theory replaces a previously accepted one, it simply generalizes the older one (or perhaps two or more older theories). The (alleged) paradigm case was Newton's "synthesis" of the laws of Kepler and of Galileo: Kepler's laws govern planetary motions; Galileo's govern terrestrial free fall and projectile motion; Newton's theory provides an account of *all* motion anywhere in the universe that, when applied to the planets and to terrestrial objects respectively, yields Kepler's and Galileo's laws as special cases.

Despite tenacious defense, this cosy picture – so the widespread story continues – could not indefinitely resist the impact of the two great revolutions of the twentieth century. For two centuries Newton had been supposed to have discovered the truth about the universe, but then his theory was rejected in favor of a relativistic rival that fundamentally contradicts it in several important ways: for example, replacing the Newtonian assertion that time is absolute, with the claim that two events may be simultaneous in one frame of reference but not in another. The "quantum revolution" involved breaks with entrenched ideas that seem, if anything, even more radical – for example, classical physics is deterministic, quantum theory seemingly inherently probabilistic.

And, so this story concludes, once these changes had been seen as "revolutionary," commentators (most notably Thomas Kuhn in his celebrated *Structure of Scientific Revolutions*) could emphasize that there had in fact been revolutionary change across the board in science. For example, the accepted view of the nature of light has changed from material particle, to wave in an elastic medium, to wave in a *sui generis* electromagnetic field, to photons – "particles" without rest mass obeying probabilistic laws.

Unsurprisingly, this sketch is at best a *highly* reconstructed rational reconstruction of history; but what *is* true is that many of the most important problems in philosophy of science since the 1960s have involved attempts to come to terms with (apparently *radical*) theory-change in science. Kuhnian theory-change seems to challenge the two most basic theses that single science out as epistemically privileged: the thesis of *scientific realism* and the still more basic thesis of *scientific rationality*.

Theory-change and scientific rationality

Kuhn claims that not only do successive theories separated by a "revolution" contradict one another, they are embedded within "paradigms" that involve different methodological standards. This certainly appears to entail a particularly striking version of relativism – if there are no "trans-paradigmatic" standards standing outside the scientific fray, then it seems impossible to deliver the verdict that the newer "revolutionary" theory is objectively superior to the older one: all one can do is record the empirical fact that (most of) those in the relevant scientific community came to believe that it was superior by dint of embracing the new paradigm.

Laudan (1984) agrees that if everything – theories *and* methods of appraisal (and also for him the aims of science) – were taken to change all at once in science then we would indeed be landed with "big-picture relativism." But Laudan holds that, while Kuhn may have been wrong that methods of appraisal of theories *always* change when fundamental theory does, he is certainly right that methods of appraisal are not fixed but are subject to at least occasional change. We learn *how* to do science better as we do better science! Delivering this (seemingly attractive) verdict requires some way of underwriting the claim that later scientific theories are in general better than earlier ones, while at the same time allowing that the methodological standards through which we make such judgments are themselves rationally modifiable. Laudan argues that this feat can in fact coherently be achieved via his *reticulated model* of theory-change.

The basic idea of this model is that a theory T_1 may be accepted as superior to some erstwhile entrenched rival T while some methodology M is in force, but then T_1 itself, once accepted, turns out to justify a change in methodology from M to M_1. Laudan sees this idea as a version of *normative naturalism* that somehow delivers norms which are both genuinely normative and empirically-governed.

There are however difficulties with Laudan's interesting attempt. He claims, for example, that the wave theory of light was accepted while Newton's inductivist methodology, which eschews genuine theories and theoretical notions, was applied in science; but this acceptance then forced the abandonment of the inductivist methodology in favor of a more liberal hypothetico-deductive approach. It is easy to see how, *once accepted*, Fresnel's theory, with its commitment to the undeniably theoretical "luminiferous aether," would fail to cohere with inductivism as Laudan construes it. But how could Fresnel's theory have been accepted in the first place if Newton's methodology really did rule against any genuinely theoretical entities and if that methodology really was accepted by scientists?

Laudan's claims tend to conflate *professed* and *real* methodology, and also, like many of Kuhn's, seem to result from an over-inflated understanding of the admittedly vague term "methodology." If *any* claim about what types of theory for a given area are likely to prove successful is counted as "methodological," then it is no news that there has been clear methodological change over time in science. Many such "rules" are unsurprisingly paradigm- or research programme-dependent. Once, for example, Fresnel had produced a successful account of diffraction, scientists applied the "rule" that other optical phenomena should be explained in terms of waves in an elastic medium. But there is surely a *reason* why classical wave theories were once thought likely to work, but then the idea was abandoned. A reason based on judgments (about empirical support and the avoidance of ad hoc assumptions) that remained fixed. While classical wave theories were initially far and away the best empirically supported accounts of light, eventually a theory came along – Maxwell's theory – that was still better empirically supported, on those same principles, and yet rejected the luminiferous ether. There is no indication either from the history of science or from anything that Laudan or Kuhn says that there has been any change in these core principles of "little methodology."

Even if this is true, two questions immediately arise: *first* what *are* those core principles? and *secondly* what is their status – how can they themselves be defended?

Suppose, concerning the second question, we have agreed on some basic, abstract principles of empirical support. How could those principles themselves be justified? This issue – essentially of how, if at all, the principles of rationality can themselves be rationally defended – is one that has often arisen in the history of philosophy. It would seem that deductive logic dictates that the basic principles of rationality *cannot* in fact themselves be rationally defended – there is nowhere deeper to go (and even if there was, the issue would arise again with respect to those "deeper" principles). And hence it seems that the adoption of rationality must itself be arational. The best we can do is to defend those basic principles as very general, abstract *givens* or "dogmas."

This is, however, an uncomfortable admission for a rationalist to make and in philosophy of science, as in more general epistemology, a good deal of effort has gone into attempting to avoid making it. These efforts have often involved claims that certain logical circles, far from being vicious, are somehow acceptable (see Van Cleve 1984, though the idea goes back at least to Braithwaite and Goodman); they have also often involved defenses of *externalist* epistemological views (e.g., Papineau 1993); and finally, and cutting across these various efforts, it has been claimed, as we saw in discussing Laudan, that methodological rules can be regarded as themselves subject to empirical assessment and hence as *naturalized* (without sacrificing the normative force of those rules).

All these approaches have their adherents and the issues remain open – though my own view is that they each face insuperable objections.

Aside from the issue of their own status, what could the basic core principles of scientific rationality be? Despite many difficulties, it still seems clear that they somehow have to do centrally with *empirical support*. Moreover they will have to be principles of empirical support that deal adequately with the *Duhem problem*.

Kuhn's most direct challenge to scientific rationality was his claim that scientists normally treat difficulties for their theories as *anomalies* rather than as Popper-style refutations: as problems for further research and not as reasons to give up the paradigm. As anomalies mount up, both those who declare a "crisis" and look for a new paradigm and those who continue to believe that the older paradigm will eventually resolve its anomalies, are *equally rational*. This clearly threatens the idea that theory-change is invariably justified in terms of the newer theory/paradigm proving better empirically supported than the older one.

Kuhn's notion of an *anomaly* is easily, and better, explained via a Duhemian analysis. Duhem (1906) pointed out that although we often speak of testing single scientific theories against empirical data – Newton's theory against planetary positions and so on – when the deductive structure of such tests is properly analysed, the situation is seen to be more complex. Auxiliary assumptions are always needed – any attempt to test Newton's theory of mechanics plus gravitation against the observed positions of some planet will, for example, implicitly rely on an assumption (clearly a theoretical assumption) about the amount of refraction that light reflected from the planet undergoes in passing into the earth's atmosphere. Moreover, at least for many theories, the central theory itself breaks down into a "core" component and a set of more specific assumptions. For example, there is really no such thing as *the* wave theory of light. Instead, and in line with Lakatos's idea (1970) of competing research programs, there is a core idea: that light consists of some sort of waves in some sort of medium, together with an evolving set of more specific claims about the type of medium, about the waves therein and so on. Thus the full structure of an empirical test is more like the following:

Central theory
Specific assumptions
Auxiliary theories
Initial conditions

Therefore, empirical result E

Assume that, when the observation is made, E turns out to be false. All that logic guarantees is that at least one of the premises is false – it does not dictate which one and in particular it does not dictate that it is the central theory. Those scientists whom Kuhn describes as treating recalcitrant data as "anomalies" are just taking it that, at least as a first move, the "blame" for getting the data wrong lies either with an auxiliary theory or with one of the specific assumptions rather than with any theory basic to the paradigm.

There are many cases in the history of science showing that this type of move, far from being under suspicion of possible "irrationality," has produced some of the greatest scientific breakthroughs. Perhaps the most famous was the discovery of Neptune: by holding on to Newton's theory despite its apparent clash with the facts about Uranus's orbit, Adams and Leverrier were led successfully to predict the existence of a hitherto unknown planet.

Treating a negative result as an anomaly is, therefore, *sometimes* good science. But in other cases it seems to be the very essence of pseudoscience. Consider, for example, creation "scientists" defending their basic theory that god created the universe in 4004 BC against the evidence of the fossil record by assuming, as Gosse famously did, that god created the rocks with the fossils already in them.

And even within science, such defenses of an entrenched theory often seem to be clearly bad science. When, for example, the wave theory of light made impressive predictions about the results of various diffraction experiments, some corpuscularists, just as Kuhn would suggest, "held out" for their preferred theory and claimed that these results were merely "anomalies" for their theory: eventually, by making the right (and clearly quite complex) assumptions about the "diffracting forces" that affect the particles as they pass the edges of opaque objects, these results could be given a corpuscularist account. Duhem's analysis shows that such a move is always logically possible. However, although corpuscularists might produce tailor-made assumptions about diffracting forces to accommodate, say, the outcome of the two-slit experiment, the strong intuition remains that this is a telling result in favor of the wave theory.

If we are to show that theory-change in science has been rational in the precise sense that later theories are invariably better empirically supported than their predecessors, then we shall need an account of empirical support that underwrites this intuition.

An obvious distinguishing feature in these cases is that the newer theory standardly *predicts* the empirical results, while the defenders of the older theory *accommodate* those results after the fact. So Fresnel's theory *predicted* the white spot at the centre of the geometrical shadow of a small opaque disc; corpuscularists suggested *after the event* that this result might be accounted for within their approach by making suitable assumptions about "diffracting forces." Darwinian theory predicts (in a way) the fossil record; creationists only accommodate the facts after the event by supposing that god chose to draw pretty pictures in some rocks when creating them. If then there were a general defensible rule of empirical support that *predictions count more* then we would have the rationale we are seeking.

The issue of *prediction vs. accommodation* is a long-running one that continues to be hotly debated. There seem, however, to be two obvious problems with the suggestion that predictions carry more supportive weight than explanations of (otherwise equivalent but) already established facts. The *first* is that while the suggestion yields the intuitively correct judgments in *some* cases, it does not do so in all. The facts about the precession of Mercury's perihelion were, for example, well known before the general theory of relativity was articulated, and yet all serious commentators regard that theory's explanation of Mercury's orbit as constituting important empirical support for it – at least as strong support as it received from the prediction of any temporally novel fact. The *second* problem is more general: the suggestion seems to stand without any epistemic justification – why on earth *should* the time-order of theory and evidence have any epistemological import?

It seems then that for all its sharpness, the *predictions-count-more* view cannot be the correct solution to the Duhem problem. And in fact the main defect of the creationist account of the fossil record, for example, is surely not that the facts were already known when the specific theory that captures them was first formulated, but rather that they *had to be* known since they were used in the construction of that specific theory. The basic idea of creationism gives no indication whatsoever that there should be particular "pictures" found in particular rocks – the specific theory that has them as part of creation is based entirely on the observations themselves. Similarly, in the optics case, the basic idea that light consists of material particles subject to forces gives no indication whatsoever that the particular "diffracting forces" emanating from a small disc should be such as to draw the particles passing the edge so that they hit the center of the geometrical shadow: that fact had to be given and to form the starting point of the construction of some force function that would do the job. On the other hand, those cases in which some already-known result seems to supply strong empirical support to a theory are characterized by the fact that the result follows from the central theory concerned, using only *natural* auxiliaries – not special assumptions that are tailored to the fact concerned. For example, planetary stations and retrogressions fall out naturally from the Copernican theory as straightforward consequences of the fact that we are making observations of the other planets from a moving observatory: a given planet's stations occur when we overtake or are overtaken by it. The issue is not about prediction versus accommodation, unknown vs. known facts, but rather all about non-ad hoc vs. ad hoc accounts of phenomena whether already known or not (though of course a scientist cannot tailor an assumption to an empirical result she does not yet know about!).

This is not to assert that ad hoc maneuvers are automatically scientifically illicit. Adams and Leverrier created a theory specifically so that it would entail the already known (and initially anomalous) details of Uranus's orbit. Often, indeed, scientists obtain specific theories by *deduction from the phenomena* – where this really means deduction from the phenomena plus a general theory (or set of such theories) that they already accept. As I argued in "New Evidence for Old" (2002), we need in fact to differentiate two types of empirical support. *Deductions from the phenomena* supply support for the deduced theory, but *only* against the already-given background of the general theory: they supply no further support for that general theory. Thus, the creationist theory with the fossils gets (conditional) support from the fossils – they provide a very good reason to hold that particular version of the creation story if you are going to hold any version of that story at all; but the fossils give no (unconditional) support whatsoever to the general story. Similarly, in the Adams and Leverrier case the data from Uranus give very good support to their version of the Newtonian account involving a change in the number of planets presupposed, but the data alone give no unconditional support, I would say, to the general Newtonian theory. The difference in the two cases is, of course, that there is independent evidence in the Newtonian case: the revised theory is *read off* the Uranian data but then *predicts* the existence of a new planet, a prediction that can, of course, be checked observationally and which turned out to be true. In the creation case there is patently no such independent testability – writing the fossils into creation simply avoids the initial problem presented by those data but yields no further prediction that can be checked.

One important issue is whether the currently most widely held formal account of empirical support – that of the personalist Bayesians – can adequately capture the intuitive judgments of confirmation. However the merits and demerits of Bayesianism are discussed elsewhere in this collection.

Theory-change and scientific realism

The issue of *scientific realism* is clearly related to the question of scientific rationality, but is logically independent of it. It is logically possible to hold that there are fixed, objective rules of theory appraisal in the light of evidence that have governed all instances of theory-change in mature science, while at the same time being entirely agnostic as to whether following those rules is likely to take science ever closer to some aim – whether that aim be total empirical adequacy or the whole truth (as in scientific realism). Logically possible, but distinctly odd! Games specify their own aims – your team wins at football if it scores more goals, and there is nothing more to be said. But science is surely more than a game. Suppose we have agreed on the rules that dictate what it means for one theory to have greater empirical support than any of its rivals. It seems counterintuitive to claim that *all* we can say about the currently winning (*best-supported*) theory in some field is that it is indeed winning according to those rules. We would expect to be able to say something about what that judgment implies in terms of the likely relationship between that theory and the universe.

What (epistemological) scientific realists want to say, of course, is that the very best theories in the light of the rules of evidence are approximately true – not only at the empirical level but also at the level of "deep structure." The main motivation behind realism is the sometimes stunning empirical success of some theories in science: quantum electrodynamics, for example, turns out to predict the value of the magnetic moment of the electron correctly to better than one part in a billion! Intuitively speaking, realists have argued that it would be a miracle if some theory made such an amazing prediction and yet were not at least approximately true in what it said was going on *behind* the phenomena.

The chief obstacles to this view are precisely those posed by the facts about theory-change in science. If we accept that earlier theories in the history of science were quite radically false and yet enjoyed striking predictive success, then it can scarcely be claimed that it would be a miracle if present theories enjoyed the success they do and yet were not even approximately true. The history of science would be a history of *miracles*!

How (if at all) can realism about current theories be reconciled with the facts about theory-change? One line is the heroic one: accept that theories in the past were radically false but yet insist that our current theories are true. One might even *try* to make the line look less heroic by pointing out that, assuming a positive solution of the rationality problem, our theories now are epistemically superior to their historical predecessors; so why should not current theories be approximately true even though their predecessors were not? But this line is surely unsustainable. Suppose we really must admit that Newton's theory now looks radically false in the light of Einstein's theory. Although, of course, the evidence that we have now for Einstein's theory is more extensive than that for Newton's theory in the nineteenth century, the difference is only one of degree. On what grounds, then, could the realist deny the possibility that Einstein's theory might itself eventually be replaced by a theory bearing the same relation to it as it does to Newton's (of course, this would be on the basis of still more extensive evidence) and therefore come itself to look radically false?

Realists well-grounded in the facts about theory-change have not taken the heroic line, but instead have argued against the thesis that those theory-changes have invariably been *radical*. One possibility is to accept that there have indeed been radical changes in *fundamental* theory, but to point out that such changes do not seem to affect theories *lower down* the theoretical hierarchy. Correspondingly, any claim of approximate truth in the case of *fundamental* theories (e.g., concerning the basic structure of space and time) would be abandoned, and realism restricted to theories lower down the hierarchy (maybe those concerning atoms). Such an approach might be called "partial scientific realism."

A different approach – at least *allegedly* – is the widely discussed view called "entity-realism" (see, e.g., Hacking 1983). This claims to be different since it claims to eschew realism about theories altogether in favor of realism about entities. But how do we know that some (alleged) thing is an entity rather than a nonentity, that is, how do we (take ourselves to) know that there is something in reality corresponding to some term involved in our theoretical framework? The answer given is that we know this if we can *manipulate* the "entity" in question. But *why do we believe* that we are *manipulating* an electron in certain circumstances? We certainly do not ever see the electrons, let alone the *manipulation* of them. The answer is, of course, that we believe this because we accept certain *theories* that tell us that this is what we are doing and in the light of which we interpret certain observable signs (tracks in a cloud-chamber or whatever) as produced by electrons. Theories are inevitably involved. Entity-realists seem simply to be telling us that we should be realists about certain types of theory (ones that are sufficiently low-level and well entrenched) and not about others (ones that are more fundamental).

Like other versions of partial realism, entity-realism is at best agnostic about realism concerning our fundamental theories. Yet it is fundamental theories like Newton's theory with its prediction of the hitherto-unsuspected existence of Neptune that provide the most striking predictive successes and, hence, the seemingly best reason for being a realist. No one, independently of any issue about theory-change, should be a fully gung-ho realist about our fundamental theories. Quantum mechanics and general relativity are, for example, to say the least, uneasy bed-fellows, so all informed commentators expect one or both to be *corrected* in some not-yet-fully articulated "synthesis." Hence no one should claim that our current fundamental theories are outright true, but surely one should not give up so easily on the view that they are *approximately* true?

There are two versions of scientific realism on the market that – unlike partial realism – do not give up. One is defended by Philip Kitcher (1993: Ch. 5) and Stathis Psillos (1999: Ch. 5). They suggest that we should be realist about fundamental theories all right, but only about *parts* of those fundamental theories. Kitcher proposed a distinction between the *working* and *presuppositional posits* of a theory. It is *only* the latter that are rejected in *scientific revolutions*, while the working posits are invariably preserved. It therefore seems reasonable to make the optimistic meta-induction that those working posits will continue to be preserved through all future theory-changes – the reason for that preservation being that, unlike the presuppositional posits, they are *true*. Kitcher claims, for example, that Fresnel's assumptions about light waves are working assumptions, his claims about the elastic ether that carries the waves being merely presuppositional. The working posit – in the form of the idea that light is a (transverse) wave – was thus carried over in the theory-change to Maxwell's electromagnetic theory; and only the presuppositional (or idle) assumptions were abandoned.

This sounds like an attractive position. But it may be overly optimistic about what claims are really preserved through change – if we think not of the differences between Fresnel's theory and the next theory of optics, namely Maxwell's, but between it and our current theory of light, then since this involves probabilistic waves associated – by an entirely new quantum mechanics – with particles without rest mass, it is just as difficult to see Fresnel's waves preserved within that theory as it is his elastic solid ether. (Waves, that is, in some full-blooded contentual sense; there are, of course, wave *functions* in quantum mechanics – but this points toward *structural realism*. Indeed it can be argued that Kitcher and Psillos's position, when fully articulated, merges with the latter.)

Structural realism (SR), pioneered by Poincaré, attempts to deliver the "best of both worlds" (see Worrall 1989). It respects the pro-realist intuitions by agreeing that their striking predictive success is a clear indication that theories in the mature sciences have latched on to reality (no doubt in some approximate way); and at the same time it insists that, after all, the development of theoretical science, including fundamental theory, *is* cumulative (or quasi-cumulative) – but at the level of *structure*. Essentially, metaphysical ideas about how the mathematical structures involved in our best theories are instantiated in reality may seem to change radically as science progresses, but those mathematical structures themselves are invariably retained (usually *modulo* the *correspondence principle*). Maxwell's theory may do away with the elastic solid ether on which Fresnel's theory was based, and so Fresnel was indeed as wrong as he could be about *what* waves to constitute the transmission of light, but his theory continues to look structurally correct from the vantage point of the later Maxwell theory, which agrees with it that optical effects fundamentally depend on *something or other* that waves at right angles to the direction of the transmission of light. Hence Fresnel's equations – though not his preferred interpretation of the terms within them – are retained in the later theory. According to this view, Fresnel was, from the vantage point of the successor theory, as wrong as he could be about the *nature* of light (there is no such thing as the elastic solid ether and *a fortiori* no such thing as waves transmitted through it), but he was correct about its *structure* (light really does depend on *something or other* that vibrates at right-angles to its direction of transmission).

The question of whether SR is defensible currently attracts lively debate. The general feeling underlying many criticisms appears to be that SR is not strong enough to count as *really* a version of realism. Whether this is correct is an open question.