'Revolution in Permanence': Popper on Theory-Change in Science

JOHN WORRALL

Introduction

Science, and in particular the process of *theory-change* in science, formed the major inspiration for Karl Popper's whole philosophy. Popper learned about the success of Einstein's revolutionary new theory in 1919 (the same year in which his discontent with Marxism and Freudianism reached crisis-point), and Einstein 'became a dominant influence on my thinking—in the long run perhaps the most important influence of all.' Popper explained why:

> In May, 1919, Einstein's eclipse predictions were successfully tested by two British expeditions. With these tests a new theory of gravitation and a new cosmology suddenly appeared, not just as a mere possibility, but as an improvement on Newton—a better approximation to the truth ... The general assumption of the truth of Newton's theory was of course the result of its incredible success, culminating in the discovery of the planet Neptune ... Yet in spite of all this, Einstein had managed to produce a real alternative and, it appeared, a better theory ... Like Newton himself, he predicted new effects within (and without) our solar system. And some of these predictions, when tested, had now proved successful. ('IA', p. 28)

Popper saw the development of science, through the process of change in accepted theory, as *the* exemplification of 'the critical approach'. Science is rational because all of its theories are open to empirical criticism, and because it stands ready to reject any such theory if criticism succeeds, no matter how impressively the theory had performed in the past. Having identified this approach at work in science, he went on to claim that it is the basis of human rationality both inside and outside science: what constitutes a 'mistake' may differ from field to field, but the rational method is always that of standing ready to make mistakes, and especially ready to learn from them. This simple idea then becomes a basic theme in almost every other part of Popper's philosophy.

Popper several times—especially in his later work—cited the 'simple schema'—

 https://doi.org/10.1017/S1358246100005440

$$P_1 \rightarrow TT \rightarrow EE \rightarrow P_2$$

which, he claimed, characterized all rational problem-solving right across the scale from the amoeba to Einstein. In particular it is, he claimed, 'the schema for the growth of knowledge through error-elimination by way of systematic *rational criticism*' (*OK*, p. 121). Here 'P_1' refers to the initial problem situation facing the organism or scientist, 'TT' to the tentative theory proposed as a solution to the problem, 'EE' to the process of 'error-elimination' applied to the tentative theory, and P_2 to the revised problem-situation that results from this trial-and-error process.

Popper often insisted on the power of these simple ideas and clearly believed in particular that his problem-solving scheme reveals a great and simple truth. The impact of Popper's ideas suggests that, at least in some areas, this home truth really needed driving home. It is, however, difficult to think of *serious* thinkers who would challenge the scheme as it stands. Who—outside perhaps of a few extreme social constructivists—would deny that criticism and problem-solving play important roles in the development of science? Who would deny that we should learn from our mistakes and that science has managed to do just that? Who would deny that that all our theories are—to an extent at least—tentative? Certainly, the issue between Popper and rival philosophical theories of the development of science—such as those of Kuhn, Lakatos, Laudan, Shapere, van Fraassen, the Bayesians and others—is joined only when various notions are more precisely specified. What *exactly* constitutes *scientifically valuable* criticism, for example? Does producing the most valuable criticism involve holding *all* theories equally open to correction? How *exactly* is 'error' established in science? What *exactly* do we learn from our mistakes ('truer' theories or only ones that have higher empirical adequacy)? Can some theories, although always strictly speaking tentative, nonetheless become *probable* to a reasonably high degree? Are successive 'trials' informed by the successes and failures of previous ones? And, if so, exactly how? Popper, especially in his later work, insisted on interpreting various criticisms of his account of the development of science as attacks on his simple scheme and hence on 'the critical approach' in general. But, as I shall argue, some at least of these criticisms are more charitably, and more revealingly, seen as rival attempts to put some real meat on what is in truth a pretty skeletal skeleton.

The two criticisms of Popper's own attempt to fill out the details of the general scheme that I shall discuss are these. The *first* is that he basically mischaracterized the process of 'error-elimination' in science. And the *second* is that he basically mischaracterized the process by which 'tentative theories' are proposed. Put baldly: Popper's view was that science is entirely based on the method of 'trial-and-error', 'conjecture and refutation', and yet—so these criticisms allege—he seriously misidentified the nature *both* of the process of identifying error in science *and* of the process of theory-production or 'conjecture'. Both criticisms, especially that concerning error-elimination, have been developed and quite extensively discussed before. I hope however that I shall add something new to them.

1. Refutations: Popper *versus* Kuhn, Lakatos *et al.*

It is easy to get the impression when reading Popper (or, rather, it is difficult to avoid the impression) that the basic picture being presented is a very straightforward version of the trial-and-error schema. Tentative theories are put forward in response to problems; these theories are scientific only if empirically testable—that is, only if they have deductive consequences whose truth value can be agreed on in the light of experiment or observation; those theories are tested; some may, if we are lucky, survive for a while the severest tests we can subject them to—these theories are *temporarily* 'accepted'; but the process of testing must always continue and if a hitherto accepted theory eventually fails a test, then it is rejected and a new tentative theory sought. The chief vehicle of scientific progress, on this straightforward view, is the direct empirical refutation of theories.

Criticism of Popper's emphasis on experimental falsification came to a head as a result of the impact of Thomas Kuhn's views of science. There is, however, nothing of real relevance to this particular issue in *The Structure of Scientific Revolutions* that was not raised already in Duhem's *The Aim and Structure of Physical Theory*.[1] Indeed many of the Kuhnian theses that have created such a stir in philosophy of science seem at root to be (often rather less clear) restatements of Duhemian positions. Consider, for example, Kuhn's famous claims about 'elderly hold-outs', claims that, as we shall see, take us straight to the heart of the falsifiability issue.

According to Kuhn, if we look back at any case of a change in

[1] T. S. Kuhn, *The Structure of Scientific Revolutions* (first edition, 1962; second edition 1970, Chicago: Chicago University Press). P. Duhem, *La Théorie physique. Son objet, sa structure* (Paris, 1906); English translation *The Aim and Structure of Physical Theory* (Princeton: Princeton University Press, 1956).

fundamental theory in science, we shall always find eminent scientists who resisted the switch to the new paradigm long after most of their colleagues had shifted. Famous examples of such hold-outs include David Brewster, who continued to believe in the corpuscular theory of light long after Fresnel's wave theory was developed, and Joseph Priestly, who persevered in defending the phlogiston theory against Lavoisier's oxygen theory. These hold-outs are often (though not invariably) elderly scientists who have made significant contributions to the older paradigm.

Kuhn is of course right that important hold-outs existed; he may or may not be right that there are *always* significant hold-outs in any scientific revolution. But the challenging Kuhnian claim was not descriptive, but instead the *normative* claim that these elderly hold-outs were no less justified than their revolutionary contemporaries: not only did they, as a matter of fact, stick to the older paradigm, they were moreover, if not exactly right, then at least *not wrong* to do so. On Kuhn's view, 'neither proof nor error is at issue' in these cases, there being '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 who switch to the new paradigm be so condemned.

Why does Kuhn think the resisters no less rational than their more mobile colleagues? His full account of theory change is complex, but he is quite clear about the basic reason why resistance of the new theory fails to be irrational:

> 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.[2]

The fundamental point here is the one already made by Duhem, namely, that the sorts of assertions that tend to be called 'single', 'isolated' theories have in fact no empirical consequences of their own. A scientist may speak of testing Newton's theory of gravitation, say, by observing planetary positions. But when that test is subjected to a full deductive analysis, it is readily seen that a range of other assumptions are in fact involved—amongst others, assumptions about the number and masses of the other bodies in the solar system, about the non-existence (or neglibility) of any forces other than gravitational forces, about how the telescope works and about the extent to which light is refracted in passing

[2] Kuhn, *The Structure of Scientific Revolutions*, pp. 151–152.

into the earth's atmosphere. All these assumptions are needed if a genuinely observational consequence is to be deduced. Thus the smallest unit with anything like directly testable empirical consequences is a *group* of theories (or a *theoretical system*) based on a central theory, but including more specific assumptions and auxiliary and instrumental assumptions.

Often—as, for example, in the case of the corpuscular theory of light (a case that Duhem himself considered in detail)—the situation is further complicated: the 'central theory' itself breaks down into a 'core' and a set of more specific assumptions. Thus 'the' corpuscular theory of light consists of the basic assumption that light consists of *some sort* of material particles; to which scientists then need to add more specific assumptions about the particles (for instance about what differentiates those producing blue light from those producing red light), and about what particular forces act on those particles in particular circumstances (for instance in passing from one optical medium into another).

Duhem's point about the real deductive structure of observational and experimental tests implies, of course, that if a test outcome is negative, if the experiment or observation contradicts the predicted consequence, then, even assuming that we know the (negative) test outcome for sure, the only theoretical unit that we can strictly infer is false is the *whole set* of assumptions needed to derive the experimental prediction. That is, all that we know (directly) from such a 'refutation' is that at least one assumption from within this set is false, we do not know just from the negative result which *specific* assumption is false. In particular, we cannot of course infer that it is the 'central' theory.

It is easy to point to historical cases in which scientists retained a central theory despite the experimental refutation of the theoretical system based on it by rejecting instead an auxiliary assumption. And, importantly, it is easy to point to such cases in which *the scientists concerned seem obviously justified in doing so*. One especially famous example concerns the discovery of the planet Neptune. The predictions about the orbit of the planet Uranus made on the basis of Newtonian theory turned out to be wrong. But instead of regarding this as refuting that theory, Adams and Leverrier independently conjectured that there is a further and hitherto undiscovered trans-Uranian planet, and that once its gravitational effect is taken into account the correct predictions about Uranus's orbit will follow from the theory. Adams and Leverrier were in effect pointing out that a prediction about a particular planet's position cannot be deduced just from Newton's theory; instead, further assumptions are required—in particular one about the other gravitating bodies that are affecting the planet concerned. And they went on to suggest that the best way to deal with the refutation of the initial theoretical system by the observations of Uranus was, *not* by rejecting the central Newtonian theory, but by rejecting the initial auxiliary about the number of other bodies affecting Uranus. Roughly speaking, they 'worked backwards'—assuming the truth of Newton's theory—to discover the simplest assumption that would give the right empirical results, and this turned out to be that there is a further planet beyond Uranus that astronomers had not yet noticed. This claim turned out to be dramatically confirmed.

In cases like that of 'the' corpuscular theory there is a further choice available to the scientist in the event of a clash between his overall theoretical system and evidence. This is an option that might loosely be described as 'modifying' the central theory rather than rejecting it entirely. In these cases, the central theory (as just indicated) itself has a 'core claim' (in the case of the corpuscular theory of light, the claim that light consists of *some sort* of material particles) alongside more specific assumptions (for example, assigning masses and velocities to the particles producing different kinds of light, making particular assumptions about the forces that affect these particles, and so on). A proponent of the corpuscular theory of light might find that, when she makes particular assumptions about the forces that operate on the light particles entering a transparent medium like glass, and adds plausible auxiliaries about her instruments, her overall system is refuted by observation of the amount of refraction the light actually undergoes. Seeing no way to replace the auxiliary assumptions she is making, she may decide nonetheless that it was her specific assumption about the forces operating on the particles at the interface that was wrong rather than the general assumption that light consists of material particles of some sort subject to some sort of forces. She will then produce a new theoretical system with the same auxiliaries as before and with the same core idea of the corpuscular theory, but with different specific assumptions about the particles and the forces on them.

The upshot of Duhem's analysis, then, is that when the deductive structure of a test in science is fully analysed, a whole group of assumptions is needed to derive the observational result—none of the assumptions alone being strong enough to entail any such observation. It follows just by deductive logic that, for any single specified theoretical assumption T, there must be for any set of observation statements a group of assumptions including T that entails all the observation statements in the set. Thus there are no crucial experiments: no result or set of results ever *forces* a scientist logically to give up any single theory. Duhem pointed out for example that any of the famous alleged crucial experiments against the corpuscular theory of light could have been accommodated within the corpuscular theory 'had scientists attached any value to the task'.[3]

Kuhn's discussion of hold-outs is, in large part, just a corollary of this Duhemian analysis. Kuhn simply adds to Duhem the historical claim that there are always (or usually) *some* scientists who 'attached some value to the task' of accommodating the allegedly crucial counterevidence within the older framework. But does Kuhn's striking claim that these hold-outs were 'neither illogical nor unscientific' also follow from the Duhemian analysis?

Kuhn is surely right that the hold-outs cannot be faulted as 'illogical': *if* deductive logic is the only constraint, then the hold-outs' insistence that the evidence regarded as crucial by the revolutionaries *can* be 'shoved into' the box provided by their favoured older paradigm not only fails to be demonstrably false, it is demonstrably true. Given that the paradigm-constituting central theory T, has no directly checkable observation sentences as deductive consequences, it follows that for any set of such observation sentences there must always exist a consistent theoretical system that entails T and also all the observation sentences in the given set.

But what of the claim that these hold-outs also fail to be 'unscientific'? Kuhn seems implicitly to assume (and the many sociologists of science influenced by him quite explicitly assume) that it follows from the fact that accommodation of the allegedly crucial result within the older system is always logically possible, that reason is powerless to judge against the hold-outs. But this, in effect, identifies scientific rationality with deductive rationality. An alternative conclusion (and surely the correct one) is that there are *further* articulable principles of scientific rationality which differentiate those cases where it is reasonable to defend a theoretical framework in the way that Duhem indicates is always logically possible from those cases where it is not reasonable.

Adams and Leverrier showed how to 'hold onto' Newtonian physics by attributing the clash between its predictions about Uranus's orbit and observations of that orbit to the omission of the gravitational influence of a hitherto unrecognized planet. Gosse, as is equally well known, showed how to 'hold onto' the theory that God created the universe with all its present 'kinds' in 4004 B.C.

[3] Duhem, *The Aim and Structure of Physical Theory*, p. 187.

by attributing the apparent evidence of now extinct species to God's decision to include in His creation things that look remarkably like bones of organisms of earlier species or remarkably like the imprints in the rocks of the skeletons of such organisms. There is surely a crucial difference between these two cases—a difference that has much to do with independent testability (the Adams and Leverrier switch makes new testable predictions—for example about the orbit of the new planet, while the 'Gosse dodge' is designed precisely to accommodate initially threatening data while permitting no further tests), and perhaps something to do with judgments of the relative 'plausibilities' of different possible auxiliary assumptions in the light of background knowledge. The main task that Kuhn's analysis sets for the holder of the view that radical theory-change in science is a rational process is—so it seems to me—precisely that of the first articulating the principles that underwrite the distinction between these two types of case, and then showing that those principles are themselves rationally defensible. It is exactly this task that Lakatos, Laudan, and subsequently many others have undertaken with varying degrees of success.

Duhem's points have often been emphasized in recent discussions and what I've said might seem like stressing the by now obvious. Duhem's simple and vitally important message is, however, still often misconstrued. Two respects in which my account differs from some others in the literature should be given special emphasis. *First*, what is true about theory-testing is that when the deductive structure is properly analysed the set of necessary premises is quite *large*—larger perhaps than one might at first expect. What is *not* true (at least not in any interesting sense) is that 'the whole of our knowledge' (whatever that may be!) is involved in the attempt to test any part of it. The slip from 'we need lots of assumptions to get consequences that are really directly checkable empirically' to 'there is no end to the assumptions we need' is, I think, just sloppy: all deductions are finite.[4]

[4] Quine seems to make this slip. This seems to be partly based on a failure consistently to distinguish between 'Indefinitely many assumptions are needed if any observational statement is to be derived' (false) and 'Although "only" finitely many assumptions are needed and so at least one of that finite set must be rejected if the observational consequence proves false, there is no *a priori* limit on the assumptions that might in turn be affected by that initial rejection' (true, but unsurprising). It also seems partly based on a flirtation with the idea that not even *deductive logic* can be taken as fixed here. But if not even a *core* of logic is taken as constituting the framework of the discussion, then it is not clear that any sense can be made of *any* assertion about testing.

The *second* point on which I disagree with some other Duhem-inspired analyses concerns the allegedly inevitable fallibility of basic statements. Suppose Newton's theory of gravitation, for example, is being tested by observations of the position of some planet, say Neptune, at some given time. One formalization of the deductive structure of the test will involve Newton's four laws as the only explicit general premises, plus 'initial conditions' about Neptune, along with positions and masses of the other bodies in the solar system at time t, and finally a 'closure' assumption—of the form 'only gravitational effects have any non-negligible effects and the only non-negligible gravitational effects are those produced by the other massive bodies in our solar system: the sun and the planets aside from Neptune itself.'[5] A conclusion will then be drawn (actually only with the help of mathematical approximations—another story that I shan't go into here) about the position of Neptune at time $t+\Delta t$. No mention in this formalization of optics, atmospheric refraction or the rest. But of course the acceptance of the 'initial conditions' concerning planetary positions at t, as well as that of the test prediction about the position of Neptune at $t+\Delta t$, depend implicitly on auxiliary theories about optics. Scientists can only be construed as 'observing' planetary positions by pointing telescopes at the sky if a range of background theories is taken for granted.

There are no rules about how to formalize informal deductions, and so long as the auxiliary assumptions concerned are regarded as uncontroversial it will be natural to leave them implicit. This formalization, however, involves initial conditions and a test prediction that are theory-laden, not just in the trivial sense that all statements about the objective world are bound to be (the assertion that Descartes's tedious demon does not exist is of course a theory), but in a sense that involves the serious possibility of later correction. If the test is formalized in this way, then there appear to be two options in the case of an inconsistency between 'observation' and theory: reject the theory (still in fact a theoretical *system*, though a comparatively slim one) or reject the test result—where the latter means asserting that *either* the initial conditions in fact failed to hold *or* the apparent result was 'wrong'.

Some episodes from history of science are naturally told as ones in which the second option was adopted: some observational claim was 'corrected' in the light of theory. A celebrated example con-

[5] An alternative formulation would simply involve an 'initial condition' about the *total* force acting on Neptune at t. But the more complex assumptions about the planets as well as the 'closure' assumption would, of course, simply be hidden in such an initial condition.

cerns Flamsteed and Newton. Again speaking very roughly: Newton wrote to Flamsteed, the first Astronomer Royal, to ask him to check some of his theory's predictions about planetary positions; Flamsteed wrote back telling Newton that the predictions were incorrect; Newton replied that his theory is correct, it was Flamsteed's observations that were wrong and if he cared to 'recalculate' them, using the formula for atmospheric refraction that he, Newton, supplied, then he, Flamsteed, would find that his observations really confirmed the theory's predictions. Since it was Newton's view that prevailed, this looks like a classic case of theory overriding observation and hence not just of the in principle theory-ladenness of observation but of the real in practice corrigibility of observation statements.

There is nothing wrong with this way of telling the story so long as it is realized—as it seldom is—that it is equivalent to the following, perhaps less exciting, but more revealing account. The clue to the second formulation of the test is Newton's suggestion that Flamsteed should *recalculate* his data. This implies, of course, that there was some 'crude data'—basically records of the angles of inclination of certain telescopes at certain times (that is, when certain clocks showed particular readings)—which were never questioned in this episode. In order, however, to 'calculate' planetary positions from these 'crude' data, Flamsteed had to make various assumptions of a low-level, but nonetheless clearly theoretical, kind. These assumptions included one about the amount of refraction that light undergoes in passing into the earth's atmosphere. If these assumptions are teased out and added to Newton's theory plus the original auxiliaries, then this creates a still larger theoretical system that, unlike the original, has deductive consequences at the crude data level. Using this more extensive articulation of the test, this episode is restored to one in which unchallenged data clashed with a theoretical system and in which the dispute was simply over which of the assumptions making up that system should be rejected. Flamsteed was suggesting that it should be Newton's theory and Newton that it should be Flamsteed's (theoretical) assumption about atmospheric refraction.

It is sometimes supposed that there are two separate reasons why empirical refutations of theories can never be conclusive: the Duhem problem *and* the inevitable fallibility of basic statements. Less confusion results, I claim, if we recognize only one problem—a 'big Duhem problem'. In any interesting historical case, there is always a level of 'data' low enough so that all sides in a dispute agree on the data. It's just that a very large number of assumptions need to be articulated and included in the theoretical system under test, if sentences at that level are really to be deduced from that system; and this means of course that, if the theoretical system turns out to be inconsistent with the (crude, unchallenged) data, there are a large number of options for replacing some part of the system to restore consistency.

In sum, the lesson that ought to have been derived from Duhem's analysis is then the following. It is always large theoretical systems that clash with empirical results; but if the system is made large enough, *every* such clash can be represented as one in which the empirical result at issue was, if not entirely unquestion*able*, then certainly never question*ed*. The chief methodological problem bequeathed by the analysis is simply that of formulating rules for ranking different modifications of such systems in the light of initially refuting evidence. Why did Adam's and Leverrier's 'modification' of the Newtonian system strengthen the basic Newtonian gravitational theory, while the 'Gosse dodge' does nothing to strengthen the empirical support for the basic idea of creationism?

The consequence of Duhem's analysis for the idea that science is characterized by the falsifiability of its claims seems quite straightforward and yet has often been obscured. Of course we can use terms as we wish, but it seems prudent in view of the above to say that there are indeed falsifications in science but that what gets *falsified* or *refuted* are large theoretical systems and never 'single' scientific theories. (Lakatos in particular was responsible for large amounts of terminological confusion here: talking, in one breath, of theories ('single', 'isolated' theories) as 'irrefutable' and, in the next, of scientists 'saving' such theories from refutation.) Duhem's analysis does *not* imply that there are no arguments for the falsity of 'single' theories provided by science. It implies only that such arguments do not consist of the deduction of a false observational sentence from that single theory, nor even of a deduction of such a false observation sentence from the theory *plus other relatively uncontentious assertions*. Instead the argument for the falsity of theories such as the claim that light beams are made up of material particles of some sort consists of two parts. First, a demonstration (usually based on a long sequence of reactions to refutations of theoretical systems built around the same core) that what other assumptions are needed in order to render that theory consistent with data is (a) implausible and (b) (crucially) entirely lacking in *independent* empirical support; and *secondly* of the production of a rival core theory, inconsistent with the first, and a demonstration that this second core theory can be incorporated into a system which enjoys *independent* empirical support and whose auxiliaries

 https://doi.org/10.1017/S1358246100005440

are at any rate much *more* plausible. So, for example, there is surely no doubt that the idea that light consists of material particles is false. The argument for its falsity, however, depends on no empirical refutation but (a) on the fact that various phenomena (for example, those of diffraction) could only be incorporated into particulate theoretical systems in an entirely *ad hoc*, non-independently testable fashion and (b) on the fact that those phenomena 'fell out naturally' from theoretical systems built around the rival wave theory.[6]

Kuhn's elderly hold-outs, then, did not perversely fail to see an experimental refutation of their favoured central theory. They were not wrong in that precise sense, but that does not mean that there was no objective argument based on experiment which told against their position. The argument that their favoured central theory eventually came to be regarded as false *for good* (*empirical*) *reasons* does not, however, consist of an empirical refutation (not even 'in hindsight'), but of the demonstration that systems built around that central theory 'degenerated' while systems built around a rival central theory scored impressive independent empirical success.

2. Falsifiability and 'conventionalist stratagems': was Popper a Duhemian about 'refutations' all along?

The Duhemian point that underlies this line of criticism of the idea of direct falsifiability of 'isolated' scientific theories seems both simple and undeniable. And indeed one aspect of Popper's reaction to Kuhn and to Lakatos was that, far from being a criticism of his position, he had emphasized it all himself long ago. Popper pointed to his explicit assertion made, not in response to Kuhn, but already in the original *Logik der Forschung*, that 'no conclusive disproof of a theory can ever be produced' (*LSD*, p. 50), and to his explicit acknowledgment that one reason for this was the ever present possibility of 'evading' an attempted refutation by using 'conventionalist stratagems' (where one type of conventionalist stratagem involves 'introduc[ing] *ad hoc* hypotheses').[7]

[6] For further details see my 'Falsification, Rationality and the Duhem Problem: Grünbaum vs Bayes', in J. Earman, A. I. Janis, G. J. Massey and N. Rescher (eds), *Philosophical Problems of the Internal and External Worlds* (Pittsburgh and Konstanz: University of Pittsburgh Press, 1993).

[7] The 'other' reason given by Popper for the inevitable inconclusiveness of empirical refutations is the alleged inevitable fallibility of basic statements. In fact, as indicated *above*, this is best treated as another, rather confusing way, of putting the same Duhemian point.

Popper *does* seem to have made the mistake—both in 1934 and later—of thinking that auxiliary assumptions are only ever *introduced* in order to 'save' a theory. In fact Duhem's point was of course that, whether we are aware of it or not, such auxiliary assumptions are *always* involved in empirical tests. For example, in discussing the Lorentz–Fitzgerald contraction hypothesis, Popper talks as if an assumption about the length of the interferometer arms was *introduced* as a result of the null-outcome of the Michelson–Morley experiment—the assumption, that is, that the length varies depending on the velocity of the arm through the ether. But clearly classical physics cannot predict any outcome of the experiment without making *some* assumption about the length of the arms: before the null-result, however, this was the 'natural' assumption that the arms were always the same length (since this is what 'solid rod' congruence measurements revealed). This slip is not as minor as might at first appear: it led Popper to make the further, related mistake of supposing that good scientific practice demands that 'conventionalist stratagems' are always to be avoided. He wrote (*LSD*, p. 82):

> We must decide that if our system is threatened we will never [sic] save it by any kind of *conventionalist stratagem*. Thus we will guard against exploiting the ever open possibility ... of ... attaining for any chosen ... system what is called its 'correspondence with reality'.

But this itself might only betoken a minor confusion, for he immediately added the important remark (ibid.):

> As regards *auxiliary hypotheses* we propose to lay down the rule that only those are acceptable whose introduction does not diminish the degree of falsifiability or testability of the system in question, but, on the contrary, increases it.

Popper characterized one form of conventionalist stratagem as involving the 'introduction' (in fact, modification) of auxiliaries *ad hoc*—that is, in a way simply designed to solve the problem posed by some refutation of the earlier theoretical system. It seems to follow then from this rule about auxiliary assumptions—accentuating the positive—that *if* a 'conventionalist stratagem' does not diminish, but in fact *increases* the degree of testability of the system, then it counts as a scientifically acceptable move. This, however, clearly contradicts his claim, made only a few sentences earlier, that in order to be scientific we must *always* eschew conventionalist stratagems. If this contradiction is resolved by taking the remark about independent testability as definitive, then it would seem to point us in the (right) direction—towards the acceptance that (relatively common) experimental falsifications falsify only large theoretical systems, and towards the acceptance that the really major experimental results are those that *confirm* the excess predictions of one such system compared to available rivals.

Popper's remarks from 1934 supply material, then, for something like the following reply to his 1960s critics: 'Kuhn and Lakatos are pointing to a slight inconsistency in the presentation of my original position, but once this is resolved in a fairly obvious way then my position in fact anticipates the point they make in alleged criticism of it. We are all agreeing on a Duhemian analysis of testing. *Of course* I agree that "single" theories are never testable in isolation. But theories like Newton's have been properly tested as parts of theoretical systems which involve independently confirmed auxiliary assumptions, while theories like Freud's have never been incorporated within genuinely testable systems, but only—at best—within systems in which already known experimental results have been accommodated after the event. When I said that Newton's theory is testable and Freud's isn't, this was simply shorthand for this rather more complicated claim.'

Although it runs counter to a lot of the straightforward falsificationist rhetoric in Popper's various writings, this would, I believe, indeed be Popper's response in a rational reconstruction of history. How far does the reconstruction caricature real history? I have been unable to find a satisfactory answer. Many of Popper's remarks suggest that he genuinely felt that Lakatos in particular was deliberately setting up a strawpopper only to use the real Popper's ideas to knock him down; and yet Popper's reactions in the Schilpp volume to both Lakatos and Putnam (who makes much the same point about falsifiability) seem monumentally to miss the point.

Popper's 'Replies' ('R.C.') contain a number of suggestions about the issue of falsifiability that do not obviously cohere with one another. There is some consideration of the old favourite 'all swans are white'. (Such examples are, of course, beside the point since Duhem's analysis applies to proper scientific theories, not simple observational generalizations, which clearly *do* have observational consequences 'in isolation'.) There is also some suggestion that Popper was going to show that Newton's theory unlike Freud's *really is* refutable, that Newton's theory can be brought into direct conflict with observational potential falsifiers without needing 'auxiliary assumptions' (which Popper insisted he had taken account of all along under the name of 'initial conditions'). In substantiation of this Popper remarked 'If the force of gravity were to become a repulsive force Putnam would soon notice the difference...' ('R.C.' p. 998). But of course the issue is not about a possible *change* in the laws of nature (even if that notion can be made sense of); and 'the force of gravity is repulsive' is clearly another theory not an observation statement. (Any theory, Freud's for example, is of course falsifiable if we allow any claim, no matter how theoretical, to count as a potential falsifier—for example (?), 'no one was ever affected psychologically by any real or imagined sexual trauma in childhood.') Popper elsewhere made more sense of this kind of response by claiming that, for example, the observational claim 'Mars (say) moved in a square orbit' is a potential falsifier of Newton's theory. But of course—Duhem's old point—that claim *is* consistent with Newton's theory taken in isolation. This is not to deny that, had such an observation been made, Newton's theory of gravitation would have been rejected. But this would *not* be on account of a direct refutation unmediated by auxiliary assumptions. Instead it would be because it seems clear in advance that any auxiliary assumptions that would produce a Newtonian theoretical system that implied the square orbit would have been massively implausible on other grounds. (Remember, no one is denying that there are circumstances under which the only reasonable thing to do is to give up some single scientific theory—the dispute is only over the *reason* for this.)

In the end it is quite unclear what position Popper held in the Schilpp volume 'Replies' ('R.C.'). On p. 998 he asserted—apparently unambiguously—that, while auxiliary assumptions might be needed for certain sophisticated tests, Newton's theory permits also certain 'crude' tests where such further assumptions are unnecessary (and so 'Newton's theory can be refuted without the help of initial conditions (i.e. auxiliary assumptions)'). But then in replying to Lakatos on p. 1004, although Popper continued to insist on the falsifiability of Newton's theory in contrast to Freud's, *this was only* '[d]isregarding the possibility of immunizing strategies'. Since this possibility and the consequent need to distinguish progressive and degenerating shifts in theoretical systems was the whole issue, I find it hard to know what sense to make of a discussion that explicitly 'disregards' them. (This point has also been made by John Watkins.[8]) Duhem's analysis shows how to avoid any loose talk about 'falsifying hypotheses', 'fallible falsifications' and the like. A genuinely falsifiable theoretical system cannot be 'immunized' against a falsification; and as for unfalsiable

[8] John Watkins, *Science and Scepticism* (Princeton: Princeton University Press, 1984).

single theories, there is no need to immunize them against non-existent falsifications.

I gratefully leave further exegetical investigation to others and settle for the following qualified conclusion: so far as the falsifiability aspect of the 'Popper–Kuhn' or 'Popper–Lakatos debate' goes, *either* there was never really anything at issue or Popper lost. There is, of course, 'error-elimination' in science but only, *directly*, of large theoretical systems. The way in which components of such systems come to be regarded as errors (in particular the way in which the 'central' components come to be regarded as false) is an altogether more complex process than simple empirical refutation. If some remarks in Popper suggest that he had insight into this process, others seems to suggest he had little, and in any event both Kuhn and Lakatos were altogether more successful in describing the details.

3. Conjectures: Popper versus Kuhn, Lakatos *et al.*

As in the case of theory-refutation, there is a simple and straightforward account of theory-*production* that it is difficult to avoid taking away from Popper's writings. This account sees the trials in the trial and error method of science as 'random', or better 'blind'—uninformed by, not deliberately structured to meet, the epistemic environment into which they are to be launched. Just as mutations in Darwinian theory are not environment-directed, nor pre-designed to solve some range of ecological problems or to fit some existing 'ecological niche', so scientific theories, on the view Popper seems to defend, are not predesigned to meet existing epistemic problems: they are *generated* in a way that is blind to existing epistemic needs, although, once generated, they are subjected to rigorous selective pressure. As Popper put it (*OK*, p. 144):

> The growth of knowledge—or the learning process—is not ... a cumulative process but one of error-elimination. It is Darwinian selection, rather than Lamarckian instruction.

As in the case of refutations, this simple view is quite seriously and straightforwardly incorrect. Again the first and most important thing is to see clearly just how wrong the view is. The purely exegetical issue of how far, or how consistently, Popper really held the view is of secondary importance and is accordingly held over to the next section.

Kuhn's criticism of the idea that scientific theories are refutable is of course a relatively minor part of the account of science developed in *The Structure of Scientific Revolutions*. His fundamental idea is that of 'normal science' practised under the aegis of a 'paradigm', where a paradigm involves not just substantive assertions about the world, but also heuristics or 'puzzle-solving techniques' (and some other, perhaps less reputable, things besides). Puzzle-solving techniques (and the idea of emulating 'exemplars') guide the construction of specific theories and in particular guide the reaction of normal scientists to experimental 'anomalies'. Although there is much in Kuhn's elaboration of this idea with which to disagree, the fundamental idea itself is surely correct and important. At any stage of the development of science, there are ideas about how to construct theories and how to modify theories should experimental difficulties arise, ideas that can, with effort, be sharply articulated and are as much a part of 'objective knowledge' as theories themselves. Alternative elaborations of this basic idea were being suggested around the same time by Hanson, Hesse, Post and others; and Lakatos's idea of a research programme, characterized in part by its 'positive heuristic' was an attempt—not an entirely successful one—to characterize this aspect of science more precisely and in more detail. There was also a largely independent development of the idea of rational heuristics within the Artificial Intelligence literature, beginning with Simon and Newell in 1958. More recently several philosophers of science have developed more detailed views along the same lines; while the idea of producing AI programs that will generate scientific theories, though still relatively in its infancy, has already begun to produce interesting results.[9]

The fact is that scientists don't simply guess their theories; they don't make 'bold' Popperian conjectures. Instead they arrive at their theories in a way that, while it no doubt involves intuition and creativity, can nonetheless be reconstructed as a systematic, and logical argument based on previous successes in science and parts of 'background knowledge' taken as premises. Kuhn's own attempt to flesh out this suggestion was not, I believe, especially successful; and although others, notably Elie Zahar, have in the meanwhile done better, a full analysis is still some distance away. However, even in the absence of a general analysis of theory-construction, it is not difficult to show, by examining the *details of particular historical episodes* that such an analysis must exist.

[9] A quite comprehensive review of both the philosophical and AI literatures on rational heuristic can be found in chapter 2 of Ken Schaffner's recent *Discovery and Explanation in Biology and Medicine* (Chicago: Chicago University Press, 1993). See also Elie Zahar's *Einstein's Revolution: A Study in Heuristic* (La Salle: Open Court, 1987).

Two sorts of situation recur time and again in the history of science. In one sort of case a general theory becomes accepted—in part because some specific theory based on it scores impressive empirical success; and it then proves fruitful to take (of course for the time being) that general theory as a premise and try to develop further specific theories based on it to account for further, related ranges of phenomena. It seems to be a fact about the history of physics at least that this happens often *and works*. The second related sort of case is where some general theory has been accepted, some specific theories based on it are successful, but the latest such specific theory developed out of that general theory, though initially empirically successful, then runs into empirical anomalies; scientists have then continued to assume the general theory as a premise and tried to use the anomalies to argue to a different specific theory on the same general lines. Again this has often proved successful.

These two possibilities are readily illustrated in the relatively straightforward case of the classical wave theory of light. The general theory that light consists of vibrations transmitted through an all-pervading mechanical medium was shown by Fresnel's treatment of diffraction in 1818 to be highly empirically successful. This treatment amounts of course to a specific theory based on the general wave idea. As is well known, Fresnel's theory of diffraction predicted, for example, that the centre of the 'shadow' of a small opaque disc held in light diverging from a single slit would be bright—indeed that the very centre would be just as intensely illuminated as if no disc were held in the light. And this prediction had been verified by Arago.[10] When Fresnel and later others came to deal, then, with further optical phenomena, such as the transmission of light through birefringent crystals, it was natural to use this general assumption of light as waves in a medium as a basis. The heuristic guidance given to scientists in this way is much stronger than might be supposed and, by dint of detailed analysis of particular historical cases, can be much more clearly articulated than was managed by either Kuhn or Lakatos or others who discerned this important point.

So, for example, when Fresnel came to develop his account of the transmission of light through crystals, he took it that he was looking for an account of the form of the wave surface within such crystals. He took it that the ether within the crystal is a mechanical

[10] For the real story of this historical episode see my 'Fresnel, Poisson and the White Spot: the role of successful predictions in the acceptance of scientific theories', in D. Gooding, T. Pinch and S. Schaffer (eds), *The Uses of Experiment* (Cambridge University Press, 1989).

medium—that is, that when one of the ether's parts is disturbed from its equilibrium position it is subjected to an elastic restoring force. Observation had shown that there are three types of transparent media: 'ordinary' unirefringent ones (ones in which only one refracted beam is created), and two different classes of birefringent media (so-called uniaxial and biaxial crystals). Previous studies of elastic media in general had established—it was part of 'background knowledge'—that the elastic restoring force acting on a part of the medium drawn away from equilibrium generally depends on the direction of the disturbance. Such direction-dependence could be expressed mathematically in terms of the coefficients of elasticity along three arbitrarily chosen mutually orthogonal axes through the medium. Fresnel was hence led to the theory that the three types of transparent medium are ones in which (a) all three coefficients of elasticity are different, (b) two coefficients are the same and the third different, and (c) all three coefficients are the same. Case (c) is the isotropic case of unirefringent media; case (b) is uniaxial birefringents crystals; case (a) biaxial birefringent crystals.

In the second type of case that I mentioned, the latest specific theory, developed out of some general idea, though initially successful, then runs into experimental anomalies, but a new specific theory is then looked for *based on the same general theory* together with *the initially anomalous data*. This can again be precisely illustrated in the case of Fresnel. Background knowledge in the form of accepted theories in the mechanics of 'continuous media' entailed that two types of waves can be transmitted through such media: pressure waves produced by the medium's resistance to compression and waves produced by the medium's resistance to shear (if any). The former are longitudinal—that is, the vibrations of the parts of the medium constituting the wave are performed in the same direction as that of the overall transmission of the disturbance through the medium. The latter are transverse—the vibrations occur at right angles to the transmission of the whole waveform. Fluids transmit longitudinal pressure waves; only solids exhibit resistance to shear and hence only solids can transmit transverse waves (along, in general, with longitudinal ones). Since the light-carrying ether has to allow the planets to move freely through it (background knowledge again entailing that, to within observational error, the planets' motions are accounted for simply by gravitational effects), Fresnel (like all his predecessors) naturally took it that the ether is a (very highly attenuated) *fluid* and hence that light waves are longitudinal. However this specific assumption was refuted (as, of course, part of a theoretical system involving further assumptions) by the results of Fresnel and Arago's modified version of the famous two slit experiment.

Fresnel and Arago found that when plates of polarizing material are placed over the two slits in this experiment in such a way that the beams emanating from the two slits are oppositely polarized, the previously visible interference pattern is destroyed. Oppositely polarized light beams fail to produce interference fringes and in particular fail to interfere destructively for any values of the path-difference. But near the centre of the observation screen the two beams are nearly parallel, and so the vibrations making them up would also be parallel, if the waves were longitudinal. This in turn implies that they could not fail to interfere destructively at those places of the observation screen corresponding to path-differences of odd numbers of half-wavelengths (assuming, of course, that the general view is correct, that is, that the beams *do* consist of some sort of waves in a mechanical medium).

One possibility in view of this difficulty would be to give up the general theory, but that general theory had been impressively successful elsewhere, and there was not the slightest indication of how to look for a different general account that could do anywhere near as well. Hence Fresnel's attitude was essentially that, given that the general theory had to be correct, the only serious question was what this experiment was telling us about the vibrations that were known to exist. And the answer was clearly that the waves (or at any rate that part of the waves responsible in general for interference effects) cannot be longitudinal. Background knowledge tells us that the other possibility is transverse waves and so Fresnel inferred (and, given his general assumptions and background knowledge, it can be reconstructed as a genuine *deduction*) that the light vibrations are orthogonal to the direction of light-propagation and, in the case of oppositely polarized rays, orthogonal to one another.

Even this rather sketchy account shows that there is nothing in this process that even remotely resembles the 'bold conjectures', the constant ferment of ideas, the 'revolution in permanence' so beloved by Popper. Refutation plays an important part in this second case but Fresnel's response to the refutation in his initial theory can hardly be described as producing a bold conjecture that then just happened to survive criticism for a while. Instead Fresnel arrived at his new, transverse theory by a systematic, deductive process, using background knowledge and the results that had refuted his earlier theoretical system. Scientists use claims that they regard as relatively well-entrenched in order to deduce specific theories from experimental discoveries. Of course if *any* step from one detailed theory to another inconsistent with it counts as a revolution (as some of Popper's remarks about 'little' revolutions suggest), then science no doubt *would* count as 'revolution in permanence', but this phrase gives all the wrong signals. The changes at issue are produced by *depending* on background assumptions of various degrees of generality. These background assumptions are used (however temporarily) as 'givens' or premises. It is the relative fixity of its theories, not their constant revolutionary change that accounts for the success of science. This is not of course, as Popper suggested in reply to Kuhn, a case of unmotivated dogmatic attachment to theories. There is any event nothing immutable about the premised background assumptions (the general wave theory was itself eventually given up), but they are *relatively* speaking permanent, and this higher degree of relative entrenchment has played a big role in the success of science.

4. The 'Darwinian' Analogy: what was Popper's *real* view about 'conjectures'?

Is there any inkling of a more sophisticated view about rationally analysable theory-construction in Popper's work? Or is the impression justified that he flatly denies the possibility of any such rational analysis? Certainly it is easy to point to passage after passage that seems unambiguously to support the 'flat denial' interpretation.

There is, of course, the famous remark in *Logik der Forschung* that

> The initial stage, the act of conceiving or inventing a theory, seems to me neither to call for logical analysis nor to be susceptible of it. The question how a new idea occurs to a man—whether it is a musical theme, a dramatic conflict, or a scientific theory—may be of great interest to empirical psychology; but it is irrelevant to the logical analysis of scientific knowledge. (*LSD*, p. 31).

But even earlier, in his *Die Beiden Grundprobleme der Erkenntnistheorie*, Popper had espoused an explicitly 'Darwinian' account of theory-production, a view which seems directly at odds with what might be called the rational heuristic view outlined in section 3. In that early work, Popper asserted:

> there exists no law-like dependence between receptions,

> between new objective conditions and the emergence of reactions (or rather there is only one form of dependence, namely the selective one, which renders non-adaptive reactions worthless...)[11]

(Popper here clearly means by 'receptions' the evaluation of a theory once it has been articulated, and by 'reactions' the articulation of a new theory.) Popper repeated this claim at various points throughout his career. As we have seen, he explicitly asserted, for example (*OK*, p. 144), that

> The growth of knowledge—or the learning process—is not ... a cumulative process but one of error-elimination. It is Darwinian selection, rather than Lamarckian instruction.

In his 1973 Herbert Spencer Lecture, revealingly entitled 'The Rationality of Scientific Revolutions. Selection *versus* Instruction', he again argued what appears at least to be an unequivocal selectionist, anti-instructionist line. Popper was there concerned to draw parallels between biological adaptation, behavioural adaptation, and the 'adaptation' of scientific theories to their epistemic environment. He asserted that:

> On all three levels—genetic adaptation, adaptive behaviour, and scientific discovery—the mechanism of adaptation is fundamentally the same... If mutations or variations or errors occur, then there are new instructions, which also arise *from within the structure*, rather than *from without*, from the environment.
>
> These inherited structures are exposed to certain pressures, or challenges, or problems: to selection pressures; to environmental challenges; to theoretical problems. In response, variations of the genetically or traditionally inherited *instructions* are produced, by methods that are at least partly *random*. On the genetic level, these are mutations and recombinations of the coded instructions...; on the scientific level, they are new and revolutionary tentative theories... It is important that these tentative trials are changes that originate *within* the individual structure in a more of less random fashion—on all three levels. The view that they are not due to instruction from without, from the environment, is supported (if only weakly) by the fact that very similar organisms may sometime respond in very different ways to the same new environmental challenge. (*MF*, pp. 78–79).

[11] Popper, *BG*. This was written before *Logik der Forschung* but finally published only in 1979. The quotation is from p. 27 and the translation is due to Elie Zahar.

It is true that Popper here stressed that genetic mutations occur against the background of an otherwise non-mutating genome which is systematically transmitted by inheritance. And it might be claimed on his behalf that this provides the analogue for the material that gets transmitted from older to newer theory in a systematic way. It is unclear that this analogy can be made to work in any even remotely precise sense, but it is surely abundantly clear that it is not generally by 'random variation', *even against a fixed background*, that scientists produce a multitude of theories which are then let out into the critical jungle to see if they survive.[12] The way for example that Fresnel produced the transverse wave theory of light is, as we saw, *clearly* 'instructed by the [epistemic] environment'. Fresnel made a deliberate and conscious attempt to produce a theory that would fill an existing epistemic environmental niche—that is, a theory that would preserve the successes of its predecessor while solving the empirical problems faced by its predecessor. Nothing could be less Darwinian.

Further evidence that Popper was committed to denying any role for rationally analysable heuristics can be gleaned from his reaction to Kuhn's idea of 'normal science'. Popper interpreted Kuhn's emphasis on the importance of normal science as amounting to the advocacy of some form of rationally unmotivated dogmatism. Popper admitted that 'Kuhn has discovered something which I failed to see ... what he has called "normal science" and the "normal scientist"' (*MF*, p. 57). But what Kuhn discovered, though real, Popper found deeply disturbing: 'In my view, the "normal" scientist, as Kuhn describes him, is person one ought to be sorry for' ('NSD', p. 52). And Popper suggested that, were all scientists to become 'normal scientists', this would be 'the end of science as we know it' ('NSD', p. 57).

Popper's remark stems from his horror of dogmatism. But, while there no doubt are passages where Kuhn overdoes the need for scientists to be 'committed' to their paradigm-constituting general theories, and while (as Kuhn in effect himself later admitted) there is no doubt that he greatly overdoes the 'paradigm monopoly' view, the point underlying his analysis is of course *not* that scientists should be 'dogmatic' in any obviously unacceptable sense. Instead he is arguing that the history of science shows that scientific progress is best made (perhaps only made) *not* by holding every assumption equally open to criticism in the sort of critical free-for-all that Popper *seemed* to advocate, but by a process in

[12] I should not want to assert that scientists are never reduced to 'random conjecture' but this is both unusual and very much scraping the heuristic barrel.

which background theories are taken as relatively well-entrenched, and are systematically used along with previous successes in science *in the construction of new specific theories.* (Of course, even these relatively well-entrenched background principles may *eventually* themselves be rejected—in so-called revolutions.)

A scientist who holds on to Newton's theory in the light of the Uranian 'anomalies' is being no more 'dogmatic' than his colleague who regards those anomalies as pointing to the falsity of Newton's theory. The two are simply placing the 'blame' for the refutation of the same *overall* system in different places: each accepts certain theories (in the latter case, the auxiliary and instrumental theories rather than the core) and therefore finds the evidence a reason to reject others. Kuhn, when properly construed, is simply recording the fact about the history of science that, once a theory such as Newton's has proved successful, it has generally proved fruitful to regard it as relatively well-entrenched and to look to deal with anomalies for the overall system by taking that 'central' theory—*relatively*—for granted. This, in turn, is no unmotivated choice, instead it allows the scientist to take advantage of various heuristic ideas based on the central theory.[13]

It was in connection with Kuhn's vision of long periods of essentially cumulative 'normal' development, punctuated by occasional 'revolutions' that Popper countered that 'science is revolution in permanence'. I believe that, ironically, the correct criticism of Kuhn's account is not that he underestimated the revolutionary nature of normal science but that he underestimated the normality of so-called revolutions. It would be easy to get involved here in essentially semantic squabbles about what counts as a revolutionary change—especially if, like Popper (and also—though for different reasons—like Kuhn in the *Postscript*), we allow ourselves to talk in effect about 'mini-revolutions'. I have already given one example, however, of a seemingly quite radical change in theory—Fresnel's switch from the fluid to the solid theory of the ether—that certainly felt like a big change to the protagonist and his contemporaries. It seemed strongly counterintuitive that the ether could be a solid and yet let the planets move through it with no perceptible effect. Yet this shift, as we saw, involved nothing like a 'conversion experience' and nothing like a 'bold conjecture', but

[13] This, incidentally, is why Popper's concession that 'dogmatism' may occasionally have *some* value (see 'NSD', p. 55 and *MF*, p. 16) is off-beam. There is *never* any need for 'dogmatism', only a need for good ideas about which particular parts of large theoretical systems need to be amended in view of experimental difficulties.

instead a systematic argument on the basis of what was then considered known *plus* Fresnel and Arago's new phenomena.

A defender of Kuhn would no doubt argue that this particular change, though perhaps more radical than it might initially seem, is nonetheless not radical *enough* to count as revolutionary in his terms—that the shift from longitudinal to transverse waves is part of 'normal science'. But it is easy enough to point to cases of theory-shifts that Kuhn explicitly counts as revolutionary, and that fit exactly this same model—the only difference being that the background 'premises' involved were of a still more general kind. Kuhn explicitly includes, for example, the shift from the Newtonian particulate theory of light (which, roughly speaking, was the most widely accepted fundamental theory of light in the eighteenth century) to Fresnel's wave theory of light (which had certainly become the most widely accepted such theory by the mid-1830s) as a scientific revolution. Yet the basic wave theory too was not arrived at by anything remotely describable as a 'religious conversion' or as a 'blind conjecture' following the refutation of the earlier theory. Rather the wave theory itself could be argued for by something like a 'deduction from the phenomena'—an argument which Fresnel himself hinted at more than once. (The argument is found explicitly in Huygens's *Treatise on Light*.)

A more fundamental part of background knowledge in the early nineteenth century than any claim about light was what might be called the 'classical world view'—the theory that all physical processes basically involve matter in motion under the action of forces. What, given this world view, might light sources emit? The general 'classical' theory (already incorporated into a range of successful specific theories) permitted only two possibilities: light sources like the sun either emit matter or they, so to speak, emit motion (or, of course, a combination). Basic ideas about matter split the first possibility into two—light might consist of a continuous stream of matter or of a discontinuous stream of particles. The former ran into a range of empirical difficulties and so, by Fresnel's time, had the particulate theory: theoretical systems based on it had done no more than accommodate various known phenomena and even that accommodation had been achieved only at the expense of a range of assumptions of extreme implausibility. But if matter was out, background knowledge also implied that disembodied motion was a nonsense: the motion had to be held by some medium in the finite interval (again the finite velocity of light was a 'given' established part of background knowledge) between leaving a source and meeting a receptor. Hence there must exist some material medium between the source and the receptor which carries the disturbances constituting light. Since light is known to be transmitted freely through a vacuum, that medium cannot be the air. Finally it was again part of accepted background knowledge that light, whatever it might precisely be, must be fundamentally periodic: a monochromatic light beam must, somehow or other, re-exhibit the same property at regular intervals. Thus we finally have the theory of light as periodic disturbances in an all-pervading, intangible, material medium—that is, we have the fundamental general idea of the classical wave theory.

What otherwise might appear as a bold conjecture—that there exists an invisible, intangible medium filling the whole of space, vibrations in which constituted light—is thus shown to be the entirely rationally reconstructible result of plugging new data and judgments based on data about other possible theories into background knowledge.

Elie Zahar has argued that, once we take into account the heuristic use of the correspondence principle, then even what seem to be *very* 'revolutionary' revolutions (such as the relativistic one) can be given gradualistic explanations along the above lines.[14] The 'correspondence principle' in this sense is the requirement that, in cases of theory change, the new theory, if it is to gain scientific acceptance, must always share the empirical successes of the old theory—a feat that the new theory generally achieves by 'going over' (*via* some mathematically characterized limiting process) to the old in the empirical domain in which the old was successful. As a requirement for the acceptance of a theory 'already on the table', this principle has often been stressed—notably by Popper himself in several places. What Zahar has shown, I believe, is that the principle is often quite deliberately and consciously used in the *construction* of the new theory. That is, it is used as a heuristic principle rather than merely as an *ex post* appraisal criterion.

So far, then, so unambiguous: the articulation of promising new theories is *not* (unsurprisingly) a question of throwing out possible conjectures 'at random' and then subjecting them to rigorous selection pressure; the process is not (even approximately) analogous to Darwinian natural selection; and yet Popper time and again emphasized this alleged analogy. He seems, therefore, consistently to have held the wrong view about theory-production and to have misinterpreted accounts that might have pointed him in the right direction. As in the case of refutations, however, it is not too difficult to find remarks that might be used to support the claim that he held a more sophisticated view.

[14] See his 'Logic of Discovery or Psychology of Invention?' *British Journal for the Philosophy of Science* (1984), pp. 243–261.

For example, he more than once insisted that in order for a scientist to produce a worthwhile conjecture, he must be 'fully immersed in the problem-background.' It would, of course, be quite mysterious why this should be necessary if 'the only form of dependence between [the production of a new theory and the epistemic situation] were the selective one...' But Popper seems never to have followed up this hint and seems to be have remained unaware of the mystery.

Secondly, Popper often stressed that metaphysics could exert a beneficial influence on the development of science. This position is one that, he repeatedly insisted, most clearly marked him out from the logical positivists. He claimed to have invented the idea of a 'metaphysical research programme' (an idea subsequently appropriated, but misinterpreted by Lakatos). Such metaphysical programmes may 'play a crucial role in the development of science'. General background claims such as those of mechanism or determinism that, I have argued, have figured as 'premises' in the deduction of theories 'from the phenomena' are of course reasonably regarded as metaphysical. However, Popper, whose attitude to such 'programmes' is by no means unambiguously positive, never seems to have developed this idea and certainly never spelled out how *exactly* metaphysical principles play a 'crucial role' in the development of fully scientific theories. Although he complained that Lakatos took the idea from him, there is no real hint that Popper was really aware of the central idea in Lakatos's conception: that of an articulable 'positive heuristic' providing definite guidance for the construction of theories within a programme.

The *third* sort of reason why Popper's view of theory-production is less clear (and therefore less clearly wrong) than might at first be supposed is that his later developments of the 'Darwinian' account of theory-production are cast around with qualifications and asides that often do not seem to cohere with the central message. It seems difficult indeed to hold the view that theorizing has no element of goal-directedness or to hold that the superabundance of rival theories ('mutations') really exists that would be necessary to give any plausibility to a properly selectionist account. (Scientists of course generally find it extremely difficult to produce *one* theory that solves the problems a fully acceptable theory in a given area would need to solve, let alone a superabundance of them.) Not surprisingly, then, there are hints that Popper saw some of the difficulties here. But rather than simply give up the Darwinian analogy on account of those difficulties, he seems to have reacted by asserting it all the more forcefully as the *fundamentally* correct view, while adding what appear to be details that, taken together, and in so far as they are clear at all, seem almost to amount to surrendering the view. The basic message, as we saw, is that Popper is on the side of 'selection' rather than 'instruction' both in biology and in the case of theory-production in science. But then certain detailed remarks suggest that he interprets 'selection' in a way that is, to say the least, rather unorthodox. So, for example, Popper introduced an alleged distinction between 'blind' and 'random' variations, and allowed that while variations in scientific discovery are 'more or less' (sic) random, they are 'not completely blind' (*MF*, p. 81). And he finally seemed to blur the whole distinction between Darwinian and Lamarckian approaches by claiming that *even in the properly biological case* 'if there were no variations, there could not be adaptation or evolution; and so we can say that the occurrence of mutations is either partly controlled by the need for them, or functions as if it was' (*MF*, p. 79).

I have no doubt that, given sufficient motivation, a case could be constructed on the basis of such remarks that Popper had a more sophisticated view of theory-production than the bold conjectures view his more general remarks seem to recommend. But, as Popper himself has argued in other connections, it may be better to hold a view that is clearly wrong, rather than one that escapes being clearly wrong only by virtue of not being clear.

Again, I thankfully leave further exegetical research to others and settle for a qualified judgment: *if* there are hints of a more sophisticated account of theory-production in Popper's work then, unlike in Kuhn and Lakatos, there is no attempt to develop them.

Popper is famous for his strikingly simple view of science. Unfortunately, despite its undoubted charms, the view is much too simple to be true. No doubt at a sufficiently high level of generality, science can be described as a process of trial and error, conjecture and refutation. But if it is to be so described, then both the process of 'conjecturing' and that of identifying 'error' need to be understood in quite sophisticated ways. Sometimes Popper seems to be aware of this and sometimes he seems vigorously to deny it; but in any event others have certainly managed to describe the processes with much greater clarity, accuracy and detail.