Danish Yearbook of Philosophy, Vol. 34 (1999), 25-42

SCIENTIFIC REVOLUTIONS AND INFERENCE TO THE BEST EXPLANATION

ALEXANDER BIRD
University of Edinburgh

Are scientific revolutions rational?

Thomas Kuhn gave us a description of scientific change in which periods of normal science are interrupted by scientific revolutions, where the paradigms governing normal science are overthrown and new ones instituted. In this paper I shall assume that something like this does indeed happen. The question I shall seek to answer is this: Is such a picture consistent with regarding science as rational? By and large, the standard answer to the question has been, No. There are a number of reasons for this. First, both opponents, notably Popper and Lakatos, and (some) defenders, such as the proponents of the Strong Program, have taken it to be so. Secondly, Kuhn himself does give the impression that during normal science rationality is relative to the governing paradigm and that during a revolution the choice of a new paradigm is governed by non-rational forces (politics, propaganda, personalities, power etc.). Thirdly, it has generally been assumed that if science were rational, then its development would be smooth and regular, not jerky as the normal science – revolution model suggests. Rationalism was regarded as accounting for the supposed fact that scientific knowledge is cumulative, always building on and improving what is there before. Since Kuhn was explicitly concerned to undermine the latter (mis)conception of science, it seemed to follow that his view must therefore undermine the view of science as rational.

In this paper I shall suggest that these suppositions are mistaken. Kuhn's description of scientific change is consistent with that change being by and large rational. Indeed, I shall suggest that Kuhn's model might even be expected on a certain account of theory choice – that account might *generate* the sort of cycle Kuhn describes. This in turn might give us some insight into other aspects of the Kuhnian model – in particular incommensurability, paradigms and the question of the scale of scientific revolutions

Inference to the best explanation

The account of theory choice with which I shall primarily be concerned is *Inference to the Best Explanation* (IBE). Although the first modern articulation of the nature of IBE is due to Peirce, (as abduction), IBE has nonetheless been largely ignored in the philosophy of science until recently. In particular Harman and Lipton have given detailed accounts of the concept and defences of the view that it is the most general description of our inductive practices. In my view there is nothing we can usefully call the *rule* of Inference to the Best Explanation, but we may spell out the idea in general terms. It is this: given adequate evidence, if *h* is clearly the best potential explanation of that evidence, then it is rational to infer that *h* is the actual explanation of that evidence, i.e. *h* is true (or is very close to the actual explanation and so is nearly true). Furthermore, we may add that any general inductive practice should conform to the *requirement of total evidence*. (The fact that we cannot say *a priori* when the quantity of evidence is sufficient, or when one hypothesis is clearly a better explanation than another, or provide a system for generating potential hypotheses, is why IBE cannot be stated as an inductive rule.)

In a little more detail, the conditions on satisfactory inference using IBE are
1. The invention of explanations Inference to the Best Explanation is a matter of selecting the best among the potential explanations of the evidence. IBE won't work if the only explanation we consider is the first to come to mind. To yield knowledge IBE requires that the inferred explanation is better than *all* the other plausible potential explanations of the evidence.

2. Goodness of explanations Once one has generated the plausible explanations, the next task is to rank them according to how good they are, so as to be able to select the best. Furthermore, one will require some measure of goodness (not merely a ranking), since we will want to infer the best explanation only if (i) it is itself good enough, and (ii) it is clearly better than the next best explanation.

3. Sufficiency of evidence IBE is applicable only once sufficient relevant evidence has been gathered and taken into consideration. However good an explanation is, if it is an explanation of only limited evidence, it won't be rationally inferable.

4. The requirement of total evidence The requirement of total evidence says that we should consider *all relevant evidence* in making inductive inferences. In the light of IBE, the requirement says that we are not entitled to infer *h* even if *h* is the best explanation of some proper subset *s* of the relevant evidence, if we ignore some other relevant evidence which is not explained by *h*, or which is better explained by some competitor to *h*.

Inferential obstacles

Meeting these conditions is clearly a considerable task in any sophisticated scientific sphere. Consider condition 1. The number of potential explanations might be very large. The chances of thinking of the right one might be quite low. Maximising the chances of meeting condition 1. would require thinking of as many potential explanations as possible. (The distinction between the contexts of discovery and of justification is often made, but in the light of IBE the two cannot so easily be separated. For according to IBE, justification and knowledge require reason to believe that the proposed explanation is better than any alternative. But typically that will require having considered at least the plausible alternatives. That in turn will require a capacity usually attributed to a thinker in the process of discovery, the power to imagine potential explanations.)

Considering condition 2., ranking the potential explanations correctly according to goodness would be easy if there were some algorithm for measuring goodness. But there is not. There is little enough agreement on what counts as the goodness of an explanation. Even if we take specific proposals as to what goodness is: simplicity, explanatory unification, explanatory loveliness, assessing to what degree these criteria are satisfied is scarcely any more mechanically decidable.

If the number of potential explanations is high, then the quantity of evidence required to provide a satisfactory ranking of them will also be high, making it correspondingly difficult to meet condition 3. But maximising the chances of having sufficient evidence (condition 3.) is at least simple – just collect more data. Nonetheless there are pitfalls and costs here too. The main cost is obvious. Collecting more data once one has enough is wasted effort, effort which might have been better put into some other enquiry (it will probably be wasted money too). There is a potential pitfall, that in collecting too much data one might include misleading data or mistake one's evidence, taking to be evi-

dence that which is false. We do not always know what our evidence is. (One's evidence is what one knows, and one does not always know what one knows.[1]) So there is a chance of including among one's evidence that which isn't evidence, and thereby undermining one's inference. The chances of this are increased as one seeks ever more evidence.

Once one has enough evidence, one must not, in making one's inferences, ignore relevant evidence. This fourth condition might appear easier to meet. After all, if in doubt, one might consider all one's evidence and so trivially meet the condition. Matters are not, however, so straightforward. As discussed, one does not always know what one's evidence is. Hence there is a chance of overlooking evidence (relevant or not). There is also, as already mentioned, the chance of considering something which isn't evidence. Furthermore, 'considering' and 'taking into account' here must mean more than being aware of'. Rather, it here means 'explaining' or 'accounting for'. This is because what 'relevant evidence' here means is precisely 'evidence which must be accounted for in a satisfactory explanation.' Consider the following example. Detectives investigating a suspicious death may entertain the hypothesis that the death is suicide. They know (i) that the gun was in the dead person's right hand, and (ii) that the dead person was left handed, and (iii) that the following day a neighbour had a win on the lottery. Fact (iii) is irrelevant. It is no requirement that a satisfactory explanation of the death account for this fact. But fact (i) clearly is relevant, since a failure to account for it would be a deficiency in any explanation of the death (for instance, given (ii) also, it seems to rule out the favoured hypothesis). Hence, if irrelevant evidence is taken into account, i.e. is required to be accounted for by any satisfactory explanation, then that may undermine one's ability to make an inference. Simple as this example may be, it is worth pointing out that it suggests that the relationship of evidential relevancy need not be obvious and is not *a priori*. For example, fact *f* may be evidentially relevant to hypothesis *h* by shedding light on some alternative hypothesis h^* (by making h^* unlikely or a poor explanation). So, seeing that *f* is evidentially relevant will depend upon our ability to spot what the plausible alternative hypotheses are. But knowledge of the full range of plausible alternative hypotheses cannot be *a priori*. Thus we may adopt the traditional requirement of total evidence, on the understanding that what is relevant evidence is an *a posteriori* matter.

Inferring reflectively

The forgoing section argued that there are pitfalls and costs associated with attempting to satisfy the conditions on IBE. For that reason we may wish to infer in a more reflective manner, so as to avoid the pitfalls and reduce the costs. For instance, reflection might tell us whether we have sufficient evidence to sanction an inference (condition 3).

It is never the case that knowing *p* requires knowing that one knows that *p* or even having a justified belief that one knows that *p*. So, in this case, to yield knowledge it is required only than an application of IBE actually meet the conditions listed, not that this be known. On the other hand, we do have a general interest in reflecting on our methods of acquiring beliefs, for two reasons. First, as mentioned, there may be costs and pitfalls we wish to avoid. Reflection on our methods may allow us jettison unreliable methods, improve partially reliable methods, assess the reliability and limits of new methods and so on. Secondly, we may wish to convince someone of a conclusion who retains some doubt as to its truth. If the method and its reliability are made clear to the audience then an appreciation of the merits of the method will help lead them to accept the conclusion it leads to. In the case of belief gained by direct application of sense perception or by simple deductive inference, there is little to be shown on this score. The reliability of modus ponens is almost universally accepted (even if unconsciously) and so as a method of inference needs neither to be explicit nor defended. A person's eyesight is not always perfect. Nonetheless, its unreliability (or the limits of its reliability) will often be quickly revealed to the individual (and often quite quickly to others), since perceptual judgments are so very frequently verified or falsified almost immediately, by use of another sense, or by the same sense within the limits of reliability (i.e. closer inspection) or by the reports of others.

Belief formation in science is not always like this. The reliability of methods used in science is neither always universally known nor subject to immediate test. The beliefs generated may not be independently verifiable; our only test of them may be the method in question, which will be itself a product of science. Its reliability will depend on the truth of some underlying theory. And so it may be required that one demonstrate the truth of that theory in order to persuade someone to accept the beliefs produced by this method. The aim of science is the expansion of knowledge. This requires adding to our methods for generating beliefs, but it also requires the avoidance of falsehood. Hence new methods and procedures for expanding our beliefs should be subject to our scrutiny.

An application of IBE presents particular problems in this respect. The validity of an application of modus ponens can be checked for mechanically and correspondingly the human recognition of a valid inference using MP is usually unconscious. But IBE is not like this since whether an application of IBE meets conditions 1.-4. cannot be decided mechanically. Consider just condition 1.: Is the proposed explanation better than any alternative? That cannot be decided mechanically, and when the field of investigation is a new one (what explains the anomalous motion of Uranus?) we cannot make reliable unconscious judgments of the matter. Conscious reflection is required. Similar remarks may be made on the evaluation of the goodness of explanations and the meeting of the requirement of total evidence. Hence, especially when applied to a new or poorly understood field, a new application of IBE must be conceived of as akin to the application of a *new* and untried method, not the instantiation of an old method whose reliability has been independently tested.

Say we were to try and formulate IBE in such a way that the question of whether something is indeed an application of IBE can be decided in a straightforward manner. Then it would have to be that the issues of whether the explanation is good enough and so on would have to be made independently of the application of IBE. In which case the fact that the explanation is good enough would have to be stated in a premise, i.e. as part of the evidence in addition to the experimental data. Like all evidence, such facts would have to be known. So, either way, it looks as if in practice we must be in a position to provide at least some justification for thinking that conditions 1.-4. are met when we employ IBE.

We have reason therefore to want to look at an application of IBE to see whether the conditions have been met. But it turns out that there is little additional guarantee that one can easily provide, especially when IBE is applied to a new field. Can we easily know that condition 1. is fulfilled? Often yes: the detective investigating the suspicious death knows that there are only so many plausible explanations of death, and so knows whether he has considered them all. But a scientist investigating a newly discovered phenomenon in a little researched field may have little idea whether she has considered all the plausible explanations. In seeking to satisfy condition 1. she will do her best to think of all the explanations she can. In trying to satisfy herself and us that all the plausible explanations have been thought of, the best she may be able to say is that she has thought hard about the matter and nothing else has occurred to her. No doubt, if colleagues and other researchers are unable to supply alternative explanations, that will give us greater confidence. But nonetheless we may collectively be in a boat similar to that in which the scientist finds herself individually: there may be no strong case for asserting that all the explanations have been thought of. This is one reason perhaps why scientists tend to regard to advanced theories as tentative. They say things like 'this is the best model we have' or 'this is what I believe is the likely explanation but I am not sure we know it to be so yet.' If the condition on being entitled to assert that *p* is that one knows that *p*, then the entitlement condition on asserting that one knows that *p* is that one knows that one knows that *p*.[2] In this case, even if the conditions on IBE are met, we may not know that they are met. Hence one might know *p* but not know that one knows *p* – hence one would not be entitled to assert that one knows that *p*.

Similar obstacles face knowing that conditions 2.-4. are met. Regarding condition 1., we saw that we may have little reason to be confident that the true explanation is among the explanations considered. Sometimes we may be more confident, but this confidence is frequently intuitive rather than justifiable in terms of reason. One's intuition may well be reliable, a scientific sixth sense or a nose for the truth, honed like a wine-taster's palate from experience: a justification but not one easily transferable to others who lack that sense. With respect to condition 2. the same may be said about judgments of goodness, of simplicity, explanatory loveliness and so on. Like the flavours in a complex wine they may be objective and knowable by trained intuition, but not quantifiable. Similarly remarks may be said about judgments of sufficiency of evidence. If judgments of taste are objective in the way Hume conceived them to be, then there may be more in common between scientific and aesthetic judgment than is commonly supposed. Knowing that conditions 3. and 4. have been met require knowing what one's evidence is and which of it is relevant. As already remarked, one does not always know what one's evidence is, and knowing of it whether it is relevant is not *a priori* and need not be straightforward.

Hence, the difficulties in effectively and efficiently meeting the conditions on IBE are not easily remedied by reflecting on what one is doing. IBE can and does yield knowledge, but there are also pitfalls which may cause it to fail to do so, and seeking to maximise the chances of meeting the conditions is an intellectually and perhaps scientifically costly exercise. This is above all true in new fields of research, where IBE is to some extent casting in the dark. Unlike perceptual methods of belief formation there is often no independent check on

the correctness of the beliefs formed. But that is not to say that there is never any check. For experimental predictions provide some check, for they provide potential falsifiers for the inferred theory. Continued experimental success may provide some additional grounds for crediting one's inferences with truth.

In the light of the difficulty in satisfying the conditions on IBE and in independently justifying or checking one's use of IBE, there is reason to ask, is there some other inference procedure which has a high chance of yielding the same results as IBE without the same pitfalls and costs? There is no such strategy which is fully independent of IBE. If there were a strategy which gave the same results but without the disadvantages of IBE then it would be at least as good as IBE (which I regard as the optimal strategy). However, I do think that there is a strategy which is dependent on some use of IBE but which obviates the need for repeated uses of IBE. The strategy is sub-optimal, but is a sufficient approximation to IBE for it in certain circumstances to yield the same results. The cases concern our treatment of an established theory in the light of new evidence.

Three types of theory adjustment

Let us imagine that at some time we have a theory ('current theory') which is the best explanation of the relevant portion of our total current evidence. We now acquire some new evidence, and so we must consider whether we are right to change our theory and if so how. I'll consider three kinds of case.

<u>Type 1. cases</u> New evidence is such that it will be explained by a straightforward application of current theory.

<u>Type 2. cases</u> New evidence is not anomalous, but neither can it be explained by current theory.

<u>Type 3. cases</u> New evidence is anomalous – that is, it is inconsistent with current theory.

In all these kinds of case the optimal strategy to apply is IBE subject to the requirement of total evidence. But as suggested, there are pitfalls and costs associated with IBE, and so the question raised at the end of the last section becomes pertinent, is there any other strategy which has a good chance of yielding the same result as our best effort at applying IBE would have done? I will consider each kind of case in turn.

<u>Type 1. cases</u> In the simplest cases there may be simple answers, for instance where there is a clear and straightforward application of current theory to explain the evidence in question. The solution is a pure extension of existing knowledge. If current theory is the best explanation of the relevant old evidence, then it will be the best explanation of the relevant old evidence plus this new piece of evidence. There is no reason to think in the light of the new evidence that any of the old but now irrelevant evidence has now become relevant. If conditions 1-4 were met before (as we are assuming) they are still met now.

<u>Type 2. cases</u> Anomalous evidence is evidence such that no reasonable explanation of it appears to be consistent with accepted theory. Sometimes new evidence cannot be explained by the old theory but neither is it anomalous. Let us first assume that it is clearly relevant or not. If irrelevant, it can be ignored and we have no reason to change our attitude to current theory. If relevant, then the normal response will be to seek a new explanatory hypothesis to add to the old theory. If we can do so, then current theory plays a part in explaining the new evidence when conjoined with the new hypothesis, and so, as in type 1 cases, that provides reason for thinking that current theory is the best explanation of old plus new relevant evidence. If the evidence is relevant but no additional hypothesis may help explain it, then the evidence should be treated as anomalous (see type 3. cases). If it is not clear whether the new evidence is relevant or not, then we may attempt to integrate it with current theory, as if it were relevant. If we succeed, then that is reason to think that the evidence is relevant, and that current theory is the best explanation of all relevant evidence. Failure to account for the new evidence in this way may itself be taken as reason to think the evidence is irrelevant, and ipso fact not requiring any change to current theory.

<u>Type 3. cases</u> Matters are more complicated where the evidence is genuinely anomalous. A satisfactory solution seems to require some modification to what is already believed. What we seek is a strategy which will give us a modified theory which best explains the available evidence, but avoids the problems that an application of IBE *de novo* would bring with it.

Let our current theory be T, and the body of available evidence relevant to it be E, and let the newly available anomalous evidence be *e*. What we know is (i) that we must make a change to T in order to accommodate *e*, and (ii) that T is well supported by a large body of evidence, E. Given (ii) we may reasonably assume that the more radical a change we make to T, the less well will the resulting theory, T*, be supported by evidence E. T* will be best supported by E (and *e*) when T* is as near as possible to T while also satisfying (i). If T is the best explanation of E, and *e* is some small addition to E, then T* is likely to be the best explanation of (E+*e*) if T* as explanatorily similar to T as can be. This principle I call *Least Disruption*. (Least Disruption is very close to Quine's maxim of minimum mutilation.[3] They are distinguished not so much by their content as their justification, which in Quine's case originates in his holism.)

Note that in cases of both types 1. and 2. the principle of Least Disruption is obeyed, as well as in 3. In both cases existing theory is maintained. And unlike in type 3. cases, existing theory is maintained unchanged, augmented in type 1. cases at most by auxiliary hypotheses (so T*=T), and in type 2. cases by additional explanatory hypotheses ($T^* \supset T$). The three cases are not always clearly distinct. When an anomaly is dealt with by the use of ad hoc hypotheses, then types 2. and 3. coincide. To the extent that there is no clear boundary between an auxiliary hypothesis, which for instance merely attributes a value to a variable, and a new explanatory hypothesis, types 1. and 2. overlap. This does not matter, what is shared by all the cases is that the theory changing moves all obey the principle of Least Disruption.

An analogy which allows one to comprehend Least Disruption involves considering how one would react to a piece of information which is in conflict with a belief held on the basis of an assertion from a highly trusted authority. In the absence of the evidence employed by the authority, we are not in the position of being able to try and find an overall explanation of his evidence plus our new anomaly. Knowing that the old belief was well-supported by evidence, we look for an account of the new evidence which is as close as possible to the old belief so as to maximise the likelihood that the unknown evidence for the old belief will also support the revised belief. Of course, in science we not typically in a position of complete ignorance about old evidence. But it is nonetheless the case that we do not seek to incorporate new evidence along with old evidence but along with old theories. Trying to review all the old evidence is intellectually too difficult. Old evidence exerts its influence not directly but indirectly via the old theory. The old theory is the representative of the old evidence at the negotiating table of theory updating. As we shall see, the old evidence may seek to be its own representative if the old theory fails in its negotiations.

Theory and evidence

I should emphasize that it is not *simply* intellectual weakness which inclines us to deal with the theory rather than its evidence. That view reminds us of the old positivist picture of theories as generalised summaries of their evidence where evidence is what has some special feature – being observational, being sense-data, being a protocol-sentence. For the positivist, theories themselves cannot be evidence. But on the view that evidence is what is known, it is perfectly legitimate to say things like "the existence of isomers is evidence for the atomic hypothesis" even though the evidence here is a highly theoretical proposition. If this is right, then in many inferential negotiations theories may appear as evidence in their own right – if they are known.

As remarked above we do not always know what we do not know, and so we may not know that something we think is evidence is not evidence. This situation of not knowing whether p is known or not will be most likely when p is known but close to not being known, or similarly when p is not known but is close to being known. New evidence may change both what our evidence actually is and what we know our evidence to be. So, for instance, we might have thought that T is known, and so evidence, until an anomaly suggests that it is not. Even so, if T was well supported by (other) evidence, some relative of T might still be known, for instance 'T is approximately true'. The latter therefore would be evidence and may be used in an inference employing the new evidence. Least Disruption trades on the hope that approximate-T is known and hopes to infer some more precise theory from it. That hope itself might in the best situations amount to knowledge, in others it might be a reasonable supposition, and in yet others it might be badly mistaken.

Limitations of Least Disruption

The principle of Least Disruption provides a heuristic for theory change during periods of normal science. It helps explain the behaviour and attitude of scientists. If an anomalous experimental result arises, then instead of dropping or radically altering current theory, less disruptive moves would include denying

the result (blame it on the incompetence of the experimenter), introducing an *ad hoc* hypothesis, or adjusting an auxiliary hypothesis. If changes are needed to the core of current theory, then this should be done in as conservative a way as possible. For instance it would be more conservative to adjust the value of some constant than to change the structure of the theory.

Least Disruption is inherently conservative, because conservative moves are more likely to maintain the explanatory virtues of earlier theory. If IBE applied to total relevant evidence is our optimal inductive procedure, then Least Disruption is a strategy designed to emulate that procedure, a strategy which aims to give similar results. It is a strategy that is adopted in lieu of the optimal procedure because of the practical difficulties in the way of obeying the optimal procedure every time we make an inference. Although Least Disruption does not involve full consideration of all relevant evidence, it aims at inferential stability.

However, since we are not guaranteed that a least disruptive change is in accord with the principle of total evidence, there is correspondingly no guarantee that the least disruptive T* will be the best explanation of (E+e), although we might have, as I have argued, every reason to think that it is highly likely be. If in normal science problems continue to be solved without too many anomalies arising, then we may be all the more confident that the strategy of Least Disruption is working. However, it is possible that each adjustment made in accordance with Least Disruption is taking us further away from what we would have come up with if we had employed IBE in accord with total evidence. A series of changes, each of which is rational and locally optimal may lead to a conclusion which is not globally optimal. In this situation we will not have inferential stability. A useful and well-known analogy is the hill climbing problem. In a fog, the strategy of following the greatest uphill gradient at any point is designed to emulate the strategy (in clear conditions) of walking towards the peak one is aiming to climb. But the emulation is not perfect and may on occasion lead one to be stranded on a low hillock. Another analogy is that of a person trying to reach a church whose steeple she can see in the distance. She has no map and so adopts this sensible policy: at any junction take that road whose direction is nearest to the direction of the church. Such a strategy, though rational, may in fact lead one further away from one's objective.

Least Disruption may on occasion do just this. Let us return to the four conditions on IBE. Least Disruption may fail to yield a theory which we would have inferred had we obeyed those conditions. In this regard condition 1. is the most safe – if we had considered all the plausible hypotheses when inferring the current theory, then the new evidence will not typically add to them. There are some circumstances, nonetheless, in which this may occur. Our evidence includes all that we know. And so it will include knowledge of developments in distinct but parallel fields. Hence a development in one (new evidence) may suggest a new hypothesis hitherto not considered in the other. More generally, surprising new evidence might prompt us to think of new hypotheses in the context of IBE. But in the context of applying Least Disruption, we are not bidden to think of new hypotheses, only to adjust the ones we already have.

Least Disruption assumes that the adjusted theory, T*, while better than T, will be a better explanation than the other competitors to T. The closer T* is to T, the more likely this will be. But there is no guarantee. The very point of Least Disruption is to avoid full investigation of the range of explanations and evidence, and *a fortiori* of the basis for judgments of comparative goodness. Hence it need not be clear whether the adjustment to T changes some feature of T which was important in deciding that it is better than its competitors, or some less important component. If the former, then it may be that in moving to T* we lose, unawares, much of what made T the best theory. Even if T* is still the best theory, the change may mean it is no longer the best theory by a long margin. Least Disruption will not recognise this.

Where all the evidence points in the same direction, we may not need much of it to come to a reasoned conclusion. When, however, the evidence points in divergent directions, more evidence will be required to decide on the best theory. Necessarily, an anomaly is a piece of evidence which points away (initially at least) from current theory. Hence it may be that in the presence of a new anomalous piece of evidence, a proper application of IBE would now desist from making a judgment until yet further evidence is gathered, while Least Disruption will promote belief in a new theory.

Lastly, the requirement of total evidence may no longer be met. I have argued that it is not *a priori* that evidence *e* is relevant to hypothesis *h*. *e* may be a piece of old evidence deemed irrelevant to T. But the new anomalous evidence may make *e* relevant to T and/or T* (for instance by casting light on auxiliary hypotheses which may help explain the anomaly). Least Disruption does not allow for this.

These are reasons for thinking that Least Disruption is indeed sub-optimal compared to IBE in that in *some* circumstances, Least Disruption may lead us astray.

That science has been lead astray by Least Disruption would be suggested by a decreasing success in solving problems, by mounting anomalies and increasing need for unsupported ad hoc hypotheses. All these suggest that current theory, as revised, is no longer as well-supported by total evidence as its predecessors were.

In such circumstances it is rational to ignore the principle of Least Disruption. As a strategy it is no longer successful. What made Least Disruption a rational principle is that it would normally be a good approximation to IBE. The failure of Least Disruption suggests that what is required is not the approximation to IBE but the real thing. This will involve ignoring some of the recent modifications to theory made in accordance with Least Disruption. So instead of looking for a solution to encompass just the current piece of recalcitrant data, one will reconsider evidence relevant to the last change to the theory, and to the change before that and so on. Returning to the analogies, it might be best to go downhill for a while in order to go up again to the peak. Or, once one realises that the road one is on is leading away from one's destination, it will be time to turn around and take a different turning at some earlier junction.

Now one is considering a broad range of evidence and looking for the overall best explanation of it, one will be in a position to contemplate quite radical changes to current theory, not just conservative changes as called for by Least Disruption. Large scale readjustments of theory may be the outcome – a scientific revolution.

A simple illustration of this process is the history of physics leading to the revolution of special relativity. Maxwell's equations require that the speed of light be constant. In the light of the background of Newtonian physics, the least disruptive way of accommodating this requirement is to infer the existence of a pervasive medium, so that the speed of light is constant in this medium – the aether. But the Michelson-Morley experiment fails to reveal an aether (or, more precisely a fixed aether). A proposal which disrupts extant theory as little as possible is Fitzgerald's hypothesis, quantified and explained by Lorentz's theory, that objects contract in the direction of motion through the aether. The next move, by Einstein, is not of the same sort, a conservative least disruptive move, but is a revolutionary change to the core theory which obviates the moves made by Maxwell, Fitzgerald and Lorentz. Once we look for the overall best explanation we see that special relativity is a better explanation than aether plus Fitzgerald contraction.

The approach to theory change I have outlined suggests that during normal science scientific inferences are governed by limited applications of or approximations to IBE constrained or governed by Least Disruption. The changes involved in a scientific revolution are less restricted applications of IBE, not constrained by Least Disruption. This view coincides with and indeed helps explain several aspects of the scientific process as described by Kuhn, while illuminating other aspects not so strongly emphasized by him. Primarily, as just noted, the approach explains Kuhn's distinction between normal and revolutionary science. Furthermore it explains why anomalies might build up under normal science – the heuristic of Least Disruption is an *approximation* to IBE – and why a scientific revolution may resolve the anomalies – because it is a much closer approximation to the optimal inference procedure.

It is often supposed, and Kuhn gives this impression, that paradigms are unchanging – until they are overthrown in a revolution. This is a point of agreement with Lakatos, for whom the theoretical hard core of a scientific research program is unconditionally protected by the negative heuristic. It seems however that this is not true to the history of science. Paradigms and theoretical cores are not immune to change. Darwin's account of evolution by natural selection is now very different from what it was in his day and its true and proper nature is hotly disputed among theorists all of whom call them Darwinians. Even Newton's theory of gravitation was considered for modification by committed Newtonians – Clairaut and Euler – when a modification would have better accounted for the evidence. Kuhn and Lakatos are willing to contemplate for normal science only changes or additional to auxiliary hypotheses. On the one hand, my account suggests why Kuhn and Lakatos are right that scientists are reluctant to change the core theory. This is because changing the core theory will typically be more disruptive than adjusting an auxiliary hypothesis. So Least Disruption requires something akin to Lakatos' negative heuristic. But my account also suggests that this disinclination need not be absolute. Least Disruption does allow scientists to make conservative changes to core theory as well. And this seems to be borne out by the historical cases. It is not that paradigms are dogmatically unchangeable nor that there is a conventional ban on changing the theoretical core; rather there is a rationally justifiable preference for conservative change in accord with Least Disruption.

The fact that Kuhn regards paradigms as incommensurable together with the examples he gives of scientific revolutions promote the impression that scientific revolutions are always large-scale events, encompassing the whole

of a discipline. Nonetheless, at several places Kuhn does also say that the cycle of: normal science – crisis – revolution – new paradigm – normal science can occur at a lower level in science, not only at the most general level.[4] Thus sub-disciplines and specific research areas can undergo scientific revolutions. My explanation of scientific revolutions allows them to differ in scale and to occur at different levels in the scientific hierarchy. The abandonment of Least Disruption does not require the recasting of a whole science. Rather it involves the reappraisal of evidence and the application of IBE, leading to the rejection of some previously held hypothesis. This process can be limited to a certain domain, for the following reasons. First, not all evidence is relevant, and correspondingly, not every theory needs to be in doubt in the face of an anomaly. Second, as in type 2. cases, Least Disruption can go hand in hand with IBE. Least Disruption tell us to keep as much theory as possible fixed, while IBE is applied in a limited fashion, to any evidence which remains unexplained by the retained quantity of theory. The amount we decide to retain might vary. So at first a larger quantity of theory might remain fixed and IBE applied to a small quantity of evidence. If this does not succeed some of the hitherto fixed theory might be abandoned, and the evidence relevant to it will now be in the pool for explanation by some new hypothesis to be chosen in accordance with IBE. Thus a scientific revolution need not encompass everything we believe, nor even everything within a given science. A small corner of a science may undergo a revolutionary change in this sense.

Since revolutions may differ in scale, it is quite possible for the first steps in a revolution to be partial or incomplete. The initial phase of a revolution may involve a partial readjustment of theory, while a later phase may involve a more full application of IBE, covering a greater range of evidence than the first phase. Thus special relativity may be seen as the first phase of a revolution whose culmination is general relativity. Since the first phase is a less than complete application of IBE, it may be unstable. Copernicus's revolution is, as described by Kuhn, an incomplete revolution, rejecting some key features of the Ptolemaic system (geocentric universe, the equant) while retaining others (circular orbits, epicycles, Aristotelian physics).[5] Kuhn describes how this combination is unstable. For some the resolution involves a retreat to something nearer the older system – Tycho Brahe exemplifies this approach, – while others, Kepler and Galileo took Copernicus's innovation to be the first step towards a new astronomy and physics encompassing changes far exceeding any-thing Copernicus could have envisaged. Thus the overall best explanation was achieved only stepwise, ultimately being put in place by Newton.

Given any state of evidence, there need not always be one clear best potential explanation of it. Two or more competing explanations may invite research, each centered around an explanatory hypothesis. In this case, if one research program runs into trouble a novel readjustment of theory may not be necessary, as one of the competing programs may be able to resolve those problems. Since the winning theory is already there, its general adoption will not be seen as revolutionary.

Once the principle of Least Disruption is ignored, it will be possible to see old evidence in a new way. What was once anomalous evidence which should be explained away in accord with the old theory might now be powerful confirming evidence for a new theory. (Such as the perihelion of Mercury.) The revolutionary recasting of theory might also allow for the reinterpretation of theoretical moves made before the revolution, made in accordance with Least Disruption. Since those moves were made to incorporate salient evidence, it is possible that some aspects of those moves will be preserved. But their relationship to the new theory may well be very different from their relationship to the old theory. Thus the Lorentz equations, which were supposed to explain Fitzgerald contraction, are retained by Einstein's special theory of relativity. But their role in relativity is quite different. Fitzgerald took contraction to be a physical effect of the aether, while in relativity the Lorentz equations quantify the effect of relative motion on the *observed* size of things. In this way many features of the world will be seen in a new light – a change which will have some of the features of the gestalt shift referred to by Kuhn, and which, I suggest, explains his earlier emphasis on incommensurability

Another, psychological, form of incommensurability may be explained by this approach. The large-scale readjustment of theory, the result of applying IBE to a broad range of evidence, is most likely to be achieved by a scientist of genius. Since this change is not proceeding in the normal (least disruptive) fashion, it may be difficult for other scientists to see the necessity for or benefits of the change. It may be psychologically difficult to make such a change, especially as the change suggests that some of the recent developments in the area have been in the wrong direction.

Conclusion

We use a variety of dispositions, strategies, methods and rules in forming and changing our beliefs. In this paper I have assumed that in much scientific inference the optimal inference pattern is captured by Inference to the Best Explanation. But IBE, I have argued, has high intellectual costs and suffers from potential pitfalls. And there is no simple additional method for avoiding those pitfalls. For these reasons we do not want to attempt to apply full-blown IBE to the totality of relevant evidence whenever we acquire new evidence. Better would be to adopt a strategy which would emulate IBE; that is, a strategy which we believe to be likely to render the same results as IBE. Least Disruption is, I suggest, such a strategy. However, an emulation of IBE which is not guaranteed to give the same results, may lead on occasion lead to results which IBE would not have given us. The fact that Least Disruption is giving us suboptimal results will usually be revealed by mounting anomalies. In such circumstances it will be rational to abandon the attempted emulation of IBE and to employ the real thing instead. While Least Disruption was designed to give us conservative change, IBE may well give us radical change – a largely new core theory. Now the process described starts again. Armed with this new theory which is the best explanation of current evidence, we will not want to apply full-blown IBE again as soon as new evidence arrives. We are in a strong position to start using Least Disruption again. Thus on this picture, which describes theory change in terms of (an explicable alternation of) rational principles, we would expect periods of conservative change to be followed, in the face of accumulating anomalies, by radical change, which is itself followed by another period of conservative change. This provides an explanation of Kuhn's cycle of normal science and revolution, but without the need for non-rational elements in theory change.

Notes

1. See T. Williamson 1997 "*Knowledge as Evidence*" Mind 106, and 1992 "*Inexact Knowledge*" Mind 101.
2. See T. Williamson 1996 "*Knowing and Asserting*" The Philosophical Review 105.
3. W.V. Quine 1992 *Pursuit of Truth* (revised edn) Cambridge, Mass: Harvard University Press.
4. T.S. Kuhn 1970 *The Structure of Scientific Revolutions* (2nd edn) Chicago: Chicago University Press; p.49.
5. T.S. Kuhn 1957 *The Copernican Revolution Cambridge*, Mass: Harvard University Press.