ON SCIENTIFIC METHOD AS A METHOD FOR TESTING THE LEGITIMACY OF CONCEPTS

ABRAHAM D. STONE
University of California
abestone@ucsc.edu

ABSTRACT. Traditional attempts to delineate the distinctive rationality of modern science have taken it for granted that the purpose of empirical research is to test judgments. The choice of concepts to use in those judgments is therefore seen either a matter of indifference (Popper) or as important choice which must be made, so to speak, in advance of all empirical research (Carnap). I argue that scientific method aims precisely at empirical testing of concepts, and that even the simplest scientific experiment or observation results in conceptual change.

There are two classic approaches, within analytic philosophy, to explaining the distinctive rationality of modern science. One of these—due, in its most influential form, to Rudolf Carnap—begins with the idea that science deals in legitimate concepts (or, as Carnap preferred to put things, uses meaningful terms). The other, due to Karl Popper, is that science follows a method suitable for seeking objective knowledge: a method, namely, in which one's position is always exposed to tests which may in principle undermine it.

Both of these lines of thought are now regarded, in many quarters, as more or less thoroughly discredited. And so they are: it is no more open to us now to go "back to Carnap" or "back to Popper" than it would be for us to go back to Kant, or, for that matter, back to Aristotle. As the latter two examples might suggest, however, this kind of discrediting of a philosophical tradition does not necessarily free us from the duty of, in some way, continuing it. The present paper may be regarded as a contribution to that task.[1]

The first, Carnapian line of thought hoped to proceed as follows. We begin with a kind of concepts (or terms)[2] which beings like ourselves evidently have a right to use—say, with those which apply directly to the primitive contents of our consciousness, or are applied in direct observation of the sensible world—and with kinds of judgments in which it is evidently legitimate to use them (i.e., with "logic").[3] We then show how it would be possible, in principle, to introduce new concepts, including all those needed by modern science, in such a way that it would remain clear, at every step, that we had the right to them, and in what judgments they could legitimately be used. The initial hope, moreover, was that this could be done by defining the new concepts in terms of the old, which for Carnap and his associates meant: giving explicit rules by which all and only legitimate judgments involving the new concepts could be replaced by equivalent, evidently legitimate judgments involving the old ones.

Success in this project would not, of course, amount to showing that the judgments of current scientists are true, or to explaining why they are true. (Despite what one sometimes reads, the logical positivists were well aware that the former is the task of science itself, in general, whereas the latter is a task, in particular, for the science of psychology.) Less obviously: it would not, at least from Carnap's point of view, amount even to showing that scientific judgments are either true or false. That semantic property—the property of being assigned either true or false under each valuation—can be assigned by stipulation when we choose a conceptual (in the formal mode: linguistic) framework. The question of whether our judgments are legitimate is rather a question of their *empirical* meaningfulness. To show that all the judgments of science are legitimate in this sense would be to show that it is rational to entertain such judgments, in the presence of confirming evidence.

This project ran into various difficulties and came under various criticisms, but most of them can be traced back to a single problem, namely what Hempel and Carnap sometimes call the *experiential import* of scientific concepts. The legitimacy of such concepts themselves, that is, as opposed to the truth or falsehood of judgments containing them, depends on the occurrence of some (at least epistemically) contingent state of affairs. This causes problems for the allegedly possible-in-principle process of introducing new concepts, because it seems to show that there cannot be a procedure for replacing all and only legitimate judgments involving the new concepts with legitimate judgments involving the old ones. Those old concepts will not, in general, have the same experiential import as the new ones. It seems, therefore, that the pattern by which judgments involving the new concepts become illegitimate (as opposed to false) on the occurrence of a certain state of affairs will not, in general, be replicable using judgments that involve the old ones.

If the above statement of the problem sounds unfamiliar, and unfamiliarly complicated, that is because discussions normally assumed that the old concepts are those applied in direct observation, whereas the new ones are "theoretical." Under that assumption, the problem can be stated more simply. Observational concepts would, among other things, have no experiential import at all; observational judgments would never be illegitimated (as opposed to falsified) by any empirical state of affairs. So we can simply say: theoretical judgments cannot be replaced by observational ones because they may be illegitimate, i.e. lack an empirical content (whereas observational judgments cannot)—that is, because they can only be "partially interpreted."[4] But this brings us to a second problem: since scientific concepts in general have experiential import, it seems that *none* of them are observational in the required sense. This problem was usually (if somewhat misleadingly) called "the theory-ladenness of observation."[5]

If, moreover, one were tempted to look at this as a correctable lapse on our part—i.e., to propose that such observational concepts be adopted—there are reasons to think they would not be usable. For if a concept is legitimate in all empirical circumstances, it must transmit that legitimacy to all its logical derivatives—that, recall, is the very function of logic in the whole project. Among the logical derivatives of (presumptively) legitimate concepts, however, there are some (such as *non-raven*) whose use is problematic, and others (such as *grue*) which seem downright unacceptable: unacceptable, that is, in that we cannot, under present circumstances, gather evidence for the truth of judgments in which they are used. There is not, for us, any evidence which serves to confirm such judgments. As a (formal) logical matter, of course, they must have a truth value: we can even imagine circumstances in which it would be possible to confirm them. But that merely serves to emphasize the point. Logical derivation, which preserves the logical property of possessing a truth value, fails, in general, to preserve meaningfulness in the sense of confirmability, and it fails because both the legitimacy of the underlying concepts (in this case, *raven*, *green* and *blue*) and the illegitimacy of their monstrous progeny are due to contingent, non-logical—presumably, empirical—factors. This means, however, that the concepts in question cannot be purely observational in the required sense. One might (again, somewhat misleadingly) say that they are laden with a theory: a theory of the world which supports our "projection" of them, rather than of the alternatives.

The second line of thought also faced difficulties. Here the idea was that science is rational because its theories—that is, in Popper's view, judgments, or systems of judgments—are empirically falsifiable. This must, as Popper recognized, be a matter of the method scientists are resolved to follow. He did always maintain that certain systems of judgments are intrinsically *un*falsifiable, insofar as they do not under any circumstances (even with added "boundary conditions") rule out anything we would be willing to accept as an observational judgment (any "basic statement"). But the methodological rule is primary: it is *because* we are not prepared to accept anything as an empirical falsification of such theories that science ought (for the most part) to avoid them. And, on the other hand, the avoidance of such theories is not in itself enough to satisfy the methodological rule. For (a) we must decide which (possible) judgments to accept as (possible) observational results, and (b) any system of judgments can be saved from empirical falsification by one or another type of ad hoc assumption. The falsifiability of scientific theories, and hence the distinctive rationality of science, is thus due, in Popper's eyes, to the fact that scientists have resolved not to adopt "stratagems" or "twists" for saving their theories from refutation (e.g. ad hoc auxiliary hypotheses or boundary conditions, or refusal to accept falsifying observations) (1994, §9, pp. 22–3; §§19–20, pp. 47–52; 1974, 983–4). What they do with their theories, rather, is try as hard as possible (within reason) to falsify them.

Having adopted this criterion of scientific rationality, however, Popper opened himself to the objection that actual scientists do not follow such a method. This would not, if true, constitute a refutation of his methodological theory, since it is not "naturalistic," i.e. not itself an empirical (scientific) theory about the method which scientists actually follow (1994, §10, pp. 23–5). Faced with particular instances in which (alleged) scientists failed to adhere to his strictures, he could and would always say that they were not behaving as they ought to. It remains the case, however, that the whole purpose of his theory is to explain why a certain actual human activity is, at least in its better moments, distinctively rational compared to others. As Popper himself admits, a *general* demonstration that scientific practice, in its most characteristic episodes, is out of step with his method would rob his proposal of any point (e.g., 1974, 1005). Thus when philosophers such as Putnam, Lakatos, and (especially) Kuhn claimed to have produced just such a demonstration, they posed a serious challenge to his view.[6]

I do not claim that the difficulties listed above are insuperable. Not being a sociologist or a historian, I cannot even say with any certainty whether the alleged difficulty for Popper's theory is based on a correct appraisal of the facts.[7] In any case, as David Lewis has emphasized (1983b, x), philosophical theories, let alone whole philosophical traditions, are rarely if ever refuted in (even) the sense that scientific ones are. But philosophical theories and traditions can be, as I put it above, discredited. Faced with such an outcome, on the one hand, and with a continuing duty to the traditions in question, on the other, one looks for a way forward. Lewis himself, in fact, can be seen as attempting just that, with respect to both of the traditions here considered. His attempt to continue the Carnapian one is explicit and self-conscious, most clearly in his explanation of how a definition can bestow experiential import (1970). The relationship to Popper is less obvious, less explicit, and perhaps partly or wholly unintended. But note that Lewis, like Popper, gives a negative solution to the problem of induction, and also that, rather than offering a proof that science will probably succeed, at our world, in identifying the absolute truth, he contents himself with a (metaphysical) demonstration that *we may reasonably hope* it will find theories with a greater degree of verisimilitude.[8]

My own attempt is not the same as Lewis's, though it does bear some relation to it (as I will explain below). What I suggest here is as follows. The Carnapian tradition was right to focus on the concepts used in science: modern scientific rationality, I will claim, has distinctively to do with getting the right concepts, rather than the right judgments. This sounds as anti-Popperian as could be: it sounds like a version of what he calls "essentialism." And indeed I will have some things to say below, following Lewis, about science as discovering essences. The conflict, however, is not as great as it seems. Popper's objection (when he makes it carefully) is to a doctrine of *ultimate* essences—that is, to the establishment or attempted establishment of a set of concepts which are unquestionably the right ones (e.g., 1989, 104–7). Such a doctrine he rightly condemns as obscurantist (that is, relatively irrational). But the search for the correct concepts need not imply such a doctrine any more than the search for truth need imply dogmatism. What I will claim to be distinctive of modern scientific rationality is, rather, precisely that it exposes its choice of concepts to severe empirical tests, and thus to possible undermining by experience. In this way it gains, not certainty in its possession of absolutely correct concepts, but reasons for preferring one set to another as relatively correct—or, as Lewis puts it, relatively "natural" (1983a, 13).

It follows, first of all, that modern science must always operate with concepts which have experiential import—for the exact same reason that, according to Popper, it must operate with judgments which are in principle falsifiable. Thus I explain the difficulties of the Carnapian line of thought: what these philosophers hoped could in principle (though not in actual fact!) be done once and for all, namely to establish which concepts are legitimate for scientific use,[9] is instead the continuing business of empirical science itself. Like Popper, however, I must add that the vulnerability of concepts to test is primarily a question of method.[10]

I have already said that I am not a historian or a sociologist. Nor (unlike certain recent philosophers of science) do I pretend to be one. I am not competent to show in historical detail, therefore, how my proposal can meet Kuhnian and related objections to Popper's view. What I will do here, instead, is more like what Carnap and Popper themselves did: I will explore some simplified and more or less heavily fictionalized examples. I will assert, but not try to show, that these examples exhibit characteristically modern-scientific methods of empirical testing.[11] What I *will* try to show is that (a) the methods thus exhibited are best understood as methods for testing, not judgments, but concepts and (b) that they are methods which it would be (relatively) rational to adopt. On this basis, one might further hope to show (c) that an enterprise employing such methods would produce something like Kuhn's (alleged) historical data. That is, one might hope to show that scientists who aimed primarily at testing concepts would appear to be engaged in something like Kuhn's alternating normal, extraordinary, and revolutionary science—i.e., not to be genuinely testing anything at all—if one came with the expectation that the point of empirical work is to test judgments. But I will leave for another occasion point (c) and its attendant difficulties (namely, the difficulty of sorting out the types and phases of revolutionary conceptual change, and the difficulty of interpreting Kuhn).

1. "Induction": Two Examples

The topic of this section is the way in which modern science tests universal judgments such as

(i) All ravens are black.

This kind of generalization is traditionally supposed to be confirmed (or disconfirmed) by induction, where induction is a process in which, after observing some state of affairs over and over (in this case: after observing many black ravens, and none that are not black), one concludes that it (probably) occurs universally. Carnap offers a version of that traditional view, as discussed above. I will argue, in contrast, that it is an inaccurate description of modern scientific methodology, and one which leaves out precisely its distinctive feature. Thus far I agree with Popper. Popper, however, wants to correct the view in question by substituting a different kind of test: one in which the judgment is corroborated or falsified by exposure to empirical evidence. I will argue, in contrast, that the severe testing characteristic of modern science concerns, not the judgment, but the concepts deployed in it. As for the judgment itself, it may indeed, as a byproduct, be either confirmed or disconfirmed. Should I say, therefore, that science does not proceed by induction, or rather that traditional accounts of induction were incorrect? Unlike Popper, I do not regard such terminological questions as trivial or unimportant. Hence the scare quotes in the title of this section.

The example (i) is traditional,[12] and it is good to use traditional examples when possible. But though it *will* prove possible discuss a hypothetical scientific project of testing (i), it will not be easy. The problem is that (i) does not, in our current situation, belong to any scientific theory, but at best to the data for which some theory (perhaps a theory of evolutionary or of molecular biology) might have to account. It is therefore not easy to see how scientists would go about testing (i), if there were or had ever been such a research program.

Let me make clear what I do and do not mean by this complaint. It is not that the truth of (i) could not be doubted. That something is dubitable, however, is not in itself enough to show whether or how it could be scientifically tested. For there is in practice a large and crucial difference between the things that a theory proposes and those which it need only explain or, at the least, not contradict. Now, it may be that our ancestors carried out some crude or unsystematic or unconscious version of the scientific testing of (i), and that our (alleged) knowledge of its (partial) truth is thanks to that process—that is, that it was at some point a crudely or unsystematically or unconsciously established scientific theory. No doubt science is, in a *certain* way, an extension of common sense. The quasi-scientific theory in question, if there ever was one, has long since been superseded. In any case, however, (i) certainly is not now and never has been the object of explicit testing in modern scientific research. It is difficult even to imagine its being so. Yet we will have to imagine just that, if we are to see what kind of evidence would then be brought to bear, and how.

So (i), in the way we would need it, is a fictional example. I have already said, however, that I will rely on such examples. The problem with (i), from my point of view, is only the very great difficulty—the very great leap of the imagination—involved in treating a case so *far* from reality. Let us therefore begin by taking on something a bit easier: a judgment which, more or less, actually does belong to a modern scientific theory. I have chosen the following, in part because it is similar to (i):

(ii) All post-main-sequence stars are red.

Like (i), (ii) is not precisely correct, both in that there are exceptions and in that it is not completely or absolutely true even in the normal cases. But, like (i), (ii) is relatively true in almost all cases. Or so we think. In the case of (ii), however, we think so because it has been tested in the process of modern scientific research.

What is the empirical evidence for (ii)? In a way, that is a trick question. The judgment (ii)—and this should be a familiar point—belongs to a complicated scientific theory, the theory of stellar structure and evolution, which we believe to be (more or less) correct for any number of reasons. In a sense, there is no empirical evidence whatsoever simply for (ii) as such. Still, there is a kind of empirical evidence which constitutes a *test* of (ii). Consider the (imaginary and highly simplified) data plotted in fig. 1, which is known as a Hertzsprung-Russell diagram.

The vertical axis represents the absolute visual magnitude of a star—a measure, roughly speaking, of its luminosity. The horizontal axis represents a star's color, measured by taking the difference between its visual magnitude as seen through two different colored filters. The stars in the narrow band running from the upper left to the lower right are (mostly) on the main sequence; the more diffuse group of stars on the upper right are post-main-sequence.

From the diagram one can see part of the reason for saying that (ii) is only relatively true. What is (mostly) true is not that post-main-sequence stars are extremely red absolutely speaking, but that they are redder than main-sequence stars of the same luminosity—or, to put it differently, that they are brighter than main-sequence stars of the same color. Under certain reasonable assumptions, the color of a star is associated with its surface temperature and hence with its luminosity per unit surface area; two stars of the same color which differ in luminosity must therefore differ in total surface area, i.e. in size. Hence it would be perhaps less misleading, in place of (ii), to say that all post-main-sequence stars are bright, or, equivalently, that all post-main-sequence stars are large. But (ii) is more or less accurate, properly understood, and has in its favor also its similarity to the traditional (i). Let us therefore continue with (ii).

In the debate between Popper and his critics, interest centered on the question of how singular instances relate to universal judgments such as (ii). Popper maintained that no singular observation could ever serve to confirm or verify a universal theory, even partially; his critics claimed that no singular observation could serve to falsify one. We should note right away that, at least as far as (ii) is concerned, and in relation to the data shown in fig. 1, both sides are correct. No one of the observations recorded in fig. 1 can be said to confirm (ii), contribute to its verification, or make it more probable. No such activity is represented here as examining many stars, finding that all the post-main-sequence ones are red, and then generalizing to the universal judgment (ii). For the same reason, moreover, there is no single observation we could add to fig. 1 which would serve to falsify (ii). That is: fig. 1 also does not record a process of examining many stars in search of a post-main-sequence one which is *not* red. On the contrary: we can only know (at least, from the data presented here) that a given star is post-main-sequence by the very fact that it is (relatively) red.[13]

And yet fig. 1 does provide a striking confirmation of (ii), of the theory of stellar evolution of which (ii) is a part. And there could well be data like those of fig. 1 which would serve to undermine (ii) and thereby to falsify that theory. Not that it could tolerate no departures whatsoever from the pattern shown. Observation of particular samples of stars (for example, from very young or old clusters) would yield different patterns, and the theory can account for that. But if, in the long run, nothing at all similar to fig. 1 were obtained—if, for example, we always got something like fig. 2—then that theory would be in hot water. As for (ii), it is perhaps not quite right to say that it would be "falsified" or "disconfirmed"—that is why I instead used the vaguer term "undermined." What would emerge, in such a case, is not that (ii) is false, in the sense that, say, some post-main-sequence stars are blue; rather, (ii) would be seen to rest on a classification of stars, into main-sequence and post-main-sequence, which has no basis in the data. To put it slightly differently: it would emerge that (ii) implies a question ("What color are post-main-sequence stars?") which has—on the basis of the available data—no objective answer.

Note that the point here is not that (relative) redness is an essential attribute of post-main-sequence stars, or that they are red by definition. To determine how we should today apply these (quasi-) Aristotelian terms, "essence" and "definition," is a worthwhile, though very difficult, project: as I hinted above, my thesis here can be understood to mean precisely that science searches for essences or for (real) definitions. But, in advance of a successful outcome to that project, we can say at least that, if there are attributes essential to post-main-sequencehood, they must be far less superficial than brightness and color—more like: exhaustion of core hydrogen and beginning to shine by some new energy source. It is not that far fetched, let alone inconceivable, that we should someday find that many such stars are not (relatively) red. Even as things stand, not *all* of them are: the big empty spaces in fig. 1 are mostly there, not because no star ever occupies such positions, but because no star remains there for more than a small fraction of its lifetime. Hence, far from being a necessary truth, (ii) is not even universally true, only approximately so. The point so far is thus not logical or metaphysical, but epistemological. The evidence which, in our current situation (at least, in this simplified fictional version of it), serves to confirm the (approximate) truth of (ii), is not separable from the evidence by which we establish that the classification main-sequence/post-main-sequence is a good one. Evidence which served to undermine the former would also have to be such as to strike a blow against the latter.

Note also that the point is not about whether sentences face the tribunal of experience alone or as a corporate body. It is true, as I have already pointed out, that (ii) could not, or could not easily, stand or fall alone. It is true that under some circumstances in which a judgment like (ii) is faced with data like those in fig. 2 we might be inclined (for whatever reason) to say that—in terms of this example—post-main-sequence stars are blue, that all stars are main-sequence, or even that all post-main-sequence stars are surrounded by clouds of dust and hence undetectable at visible wavelengths. And it is also true that, if such moves were made in a methodologically incorrect way, they could turn into the kind of "twist" which might save any theory in the face of any data. But although all these things are true and important, they are not the point of my example. The point is about how, under circumstances such that we *would* allow the data to confirm or (roughly speaking) disconfirm (ii), we would go about allowing them to do that. It should be clear, at least in this example, that the overall pattern of observations, rather than any particular observation as such, could then contribute to the confirmation or disconfirmation of (ii). One might say that the data, as judges in the tribunal of experience, sit not individually but en banc.

Note also, finally, that the point is not about the theory-ladenness of observation. The concepts *red* and *post-main-sequence* (and *star*) deployed in (ii) are indeed laden with theory, i.e. have experiential import. But, once again, although that is true, it is not the point of this example. In fact, what we can see here is almost the opposite: the way the observations themselves can, in the proper circumstances, throw off a piece of theory with which they are laden. This is clearest in the undermining case. If we approach fig. 2 on the theory that there are two kinds of stars, main-sequence and post-main-sequence, which differ, among other ways, in color, we will be frustrated: the data shown there do not allow themselves to be so interpreted. That is precisely why they serve to undermine the theory-laden (ii). A subtler and more important kind of theoryunloading takes place, however, even in the case of confirmation. We enter fig. 1, so to speak, with the theory that color and luminosity are relatively natural properties of stars—certainly if we employed concepts like *grue* we would end up with very different observations. The data shown there do not (and could not) reverse that prior determination. But this does not mean that such an initial determination as to what is most natural is unchallengeable ("foundational"). On the contrary: what the data of fig. 1 *do* show—again, under circumstances such that we are prepared to accept their verdict at all—is that there is a still more natural way of classifying stars than by color and luminosity: that these are, as I put it above, relatively superficial characteristics, best understood as a sign of something else.[14] The confirmation of (ii), in other words, involves something like what is usually called "reduction."[15] In this case too, then, the function of the data, in their overall pattern, is not simply to determine whether (ii) is (probably) true or false. That, to the extent that it happens, is a mere byproduct. What the data serve to test is the relative legitimacy of the concepts we have deployed. It remains to be established how and whether this conclusion can be extended beyond our one case.

2. Lawlikeness and the Introduction of Theoretical Terms

Let us, first of all, return to (i), the judgment that all ravens are black. Although it is formally similar to (ii), it looks at first blush as if the process of confirming it is or would be different. Wouldn't we simply go out and examine many ravens to see if they were black? Even Popper agrees that that would be the procedure, although he disagrees with the traditional inductivist's account of its purpose: namely, that, by finding many black ravens, but no (or not very many) non-black ones, we could gradually build up evidence for (i).

I would not want to deny that we do often go through a procedure very like what the inductivist imagines, and with the very purpose the inductivist imagines it as having—nor, indeed, would Popper. But I agree with Popper, first of all, that the attempt to confirm a theory by such a procedure is not particularly scientific or even rational. Later I will argue that this kind of "induction" is in itself nothing more than superstition. And I would point out, secondly, that the imagined process is only possible at all because the question of what ravens are—of how ravens are to be distinguished from other things—is regarded by us as unproblematic. Here already I call attention to a problem which Popper does not regard as important (though he is aware of it[16]), but which does receive substantial attention from both Kuhn and Lakatos.

We may begin with Lakatos's assertion that a judgment like (i), when interpreted in such a way as to be easily falsifiable by a single instance, has no "scientific value" and is "a mere curiosity" (1970, 18–19). Since (i) evidently is falsifiable in more or less that way, he thus agrees, implicitly, with my initial complaint that it does not now belong to any scientific theory. But what could give science an interest in (i)? Lakatos answers that we would need to take (i) as asserting a causal relationship between (in terms of our example) ravenness and blackness, and adds that, in that case, there would be an implicit *ceteris paribus* clause—i.e. that an apparent falsification could always be explained by the interference of other causes (ibid.). There is, I think, something correct about this answer, correctly understood. But Popper, with good reason, rejects Lakatos's version of it as "essentialist," hence (by implication) as obscurantist (1974, 1187 n. 79). If scientists were to take for granted that there are such natural concepts (essences) as ravenness and blackness, and interpret (i) as asserting (*ceteris paribus*) a causal connection between the two, then (i), far from becoming an interesting, scientifically testable judgment, would instead become irrefutable—as Lakatos himself points out.

The true situation, however, is almost the opposite: only the scientifically interesting version can be used to test a scientific theory. The fact that we can, in our current situation, simply "go out and examine many ravens" in order "to see if they are black" means precisely that we take the concept *raven* (as well as the concept *black*) for granted. This is why (i), as we mean it, is more or less easily falsified. But it is also why (i), rather than belonging to some scientific theory, is only a fact for which such theories may need to account—i.e., why the falsification in question would be no instance of scientific testing. We simply *present* our scientific theories (e.g. of evolutionary or molecular biology) with a certain thing, and say, in effect, "This is a raven. Note that it is black. Explain." We will be pleased if they can do so, or can at least explain why it is not impossible. But the discovery of a white raven would not, as such, be any problem for those theories. That would simply be a new challenge: explain why *this* raven is white. Of course, if the white ravens turned out to be identical (genetically, chemically, ecologically) to the black ones, then we might have a problem. But if we did it would be because two things which, according to our theories, ought to be the same color, are not; it would have nothing to do with the whiteness of ravens as such.

If, therefore, science were to take an interest in the truth or falsehood of (i)—if, that is, (i) were to be an actual judgment of scientific theory—then it would have to matter to the theory which things are ravens, not just whether those things are black. That is: the concept *raven* would have to be a concept of the theory in question—a concept in which the theory, so to speak, has a stake.[17] In that situation, it would, as Lakatos rightly notes, no longer be such a simple matter to falsify, or, in general, to apply evidence for and against (i). But Lakatos is wrong to suppose that (i) would thereby be immunized against testing. On the contrary: as our discussion of (ii) should already suggest, it would be precisely then that a new, characteristically scientific way of testing (i) would become possible.

What I have called "having a stake" is in reality a somewhat vague matter which admits of degrees—that, in part, is the moral of Quinean holism. But simply by talking about "theories," about judgments included in or entailed by them, and so forth, we are already engaged in an extreme idealization. And, continuing in the spirit of that idealization, we can understand how it is that a theory can have a stake in the nature of some (kind of) thing—how, that is, a concept can be theoretical—by means of an idealized (fictionalized) account of the way in which one might introduce a new theory from scratch, and of the way that new theory might bring new concepts with it.

Here we can connect once more with the Carnapian tradition. From Popper's perspective, that tradition is characterized by its interest in justifying induction. But, as I noted above, from the point of view of Carnap and his associates themselves, the most fundamental process was the legitimation of new, theoretical concepts via an imagined (fictional) introduction of them on the basis of old, evidently legitimate, observational ones. Now, Carnap, as we all know, was forced to concede early on that it was impossible to insist on this always being done by explicit definition. But David Lewis (1970), surprisingly, shows how to insist on that, after all. How can that be? By a trick, of course: he solves the problem by redefining it, in two different ways. First (following a suggestion of Putnam's) he gives up on the idea of observational terms, and instead simply discusses a procedure for introducing new theoretical terms on the basis of "old" ones.[18] Second, he allows definitions to be definite descriptions. Carnap himself uses this technique in his 1974,[19] but later drops it, presumably just because a term defined in this way will normally have experiential import beyond that of any term in the definiens.[20] Lewis, noticing that this is just what we need (1970, 89), proposes to bring the technique back and apply it systematically.

Briefly and roughly put, Lewis imagines that a theory *T* claims something using some new terms, $\tau_1, \ldots \tau_n$. The terms are new in the sense that we plan to act as if everything we know about their meaning is contained in what the theory claims. Or almost everything, rather. For such a term-introducing theory, according to Lewis, is always accompanied by a further implicit claim. Say that an n-tuple of entities $(a_1, \ldots a_n)$ *realizes* T if substituting a_1 for τ_1, … and a_n for τ_n makes T come out true. The implicit claim is: that there is a unique n-tuple of entities which realizes T. Given this claim, we can define each τ_i with a definite description: τ_i denotes the ith member of that unique n-tuple. If the implicit claim turns out to be false, then (given the right treatment of denotationless terms, which Lewis takes from Dana Scott) these definitions will leave the τ_i denotationless, and theory T itself will be false. But if, conversely, T is true, then the implicit claim must be true, and this supplies the crucial missing piece which will make our terms fully interpreted.

Lewis allows that the new terms can name entities of any kind, but the interesting case for our purposes is that in which they serve to introduce new concepts—i.e., purport to name something like properties. We need, in particular, to imagine a theory which introduces the term "raven" in such a way as to have a stake in the concept *raven* and hence an interest in (i). According to Lewis, therefore, it must say something about (what we intend to call) ravens, and we must be prepared to treat that as (almost) everything we know about the meaning of the term. Now, it could be that what the theory claims about ravens includes or strictly entails (i)—i.e. that (i) "belongs to" the theory in a formal sense. For example, we might have:

(T) … & a raven is a black bird with a short neck & ...

In that case (i) would come out true by definition: if there is anything at all which is properly called a raven, it would have to be black. A more realistic case, however, would be that (i) "belongs to" the theory in the looser sense that, given suitable approximations and simplifications, it follows from that theory, together with reasonable background assumptions, that (i) is (mostly, relatively) true. Imagine, for example, that we have an evolutionary theory which (unlike our actual ones) describes particular ecological niches and predicts whether they will be filled. Then the theory might introduce the term "raven" as follows:

> (T') … & the ecological niche Δ is filled by the species of ravens & ...

Given what we know about Δ (and reasonable background assumptions), it might follow from *that* (under certain approximations, etc.), that ravens are (mostly, relatively) black and short-necked, and are birds, or at least flying animals (animals which mostly fly pretty often and relatively well).

Whatever the difference between them, both of these imagined theories have in common what we need: both require (i) in a way that our current theories do not. Neither can be allowed to explain with equanimity why some ravens (a significant number of them) turn out to be white, after all. For they both have a stake in the concept *raven*, i.e. in the question of which things are ravens and which are not. In the case of (T) this works directly: (i) is correctly judged if and only if *raven* is legitimate; if it is not, then the implicit claim which goes along with (T) must be false. (We might or might not then call (i) itself "false"; it would certainly be incorrectly judged.) In the case of (T'), the situation is more complicated, since it is strictly speaking possible for (T') to be true while (i) is not (even mostly and relatively). It is possible, for example, for some things to be ravens as defined by (T'), and yet be mostly white. I called this more realistic, and it is indeed similar to our (still more realistic) case of post-main-sequence stars, many or even all of which might conceivably be blue. But that could only happen if some of our background assumptions, approximations, etc., are incorrect (e.g., perhaps, if the initial helium content of many stars is much different than we think, or if our approximate treatments of convection are seriously off). Something analogous, we are imagining, holds in the case of (T'). Since, however, our background assumptions, etc., are by definition things we are inclined (at least for the time being) to take as basically right,[21] whatever kind of evidence tends to undermine (i) will be prima facie evidence that there is no such kind of things as ravens, and vice versa.

What kind of evidence would that be? In what circumstances might we be led to deny the truth of (T) or (T')? The hard problem, actually, is not to come up with such circumstances, but rather to think of any others. For, although we have so far been blithely assuming that the implicit claims of (T) and (T') might be true, in fact that is far from obvious.

The problem is *not* what Lewis (1984) calls "Putnam's Paradox"—that is, the fact (not regarded as a paradox by Putnam himself) that, for model-theoretic reasons, almost any theory whatsoever is multiply realized. That, as Lewis points out (60–61), is a problem for "global descriptivism": the attempt to introduce *all* (or, all non-logical) terms of our language at once as theoretical terms of one enormous theory.[22] In our case, on the other hand, where we already understand the "old" terms—a case of what Lewis calls "local descriptivism"—these model-theoretic problems do not arise. There is, to take Lewis's example, no general model-theoretic reason why more than one person must have committed all the crimes attributed to Jack the Ripper: that actually seems unlikely, but in any case is surely not a metalogically necessary truth.[23] Even, similarly, supposing that (T) says nothing about ravens other than what I've written out, there is no general model-theoretic reason why more than one property must realize it—e.g., why more than one kind of bird must be black and short-necked.

The problem arises, rather, because the names of properties are unlike the names of individuals.[24] Continue, for example, with the assumption that (T) says nothing more about ravens. Then (T) claims there is a unique property φ (the property of being a raven) such that the things possessing φ are are black, short-necked birds. This claim seems, on any simple interpretation, to be either trivial or false. For let β be the property of being a black, short-necked bird. If we first try taking (T) to assert that there is a unique φ such that all φ things are β, we find that (T) claims there is only one β thing[25]—surely not what's intended. We could, alternatively, take (T) to mean that there is a unique φ such that *all and only* the φ things are β. If we understand this metaphysically, to mean that there is a unique φ such that, *necessarily*, all and only φ things are β, then it is true: the unique φ in question is β itself. But (i) then becomes a tautology, and "raven" a mere abbreviation, rather than, as we wanted, a theoretical term of (T). So try (what is anyway more plausible) taking (T) to claim that there is a unique φ such that, *as a matter of fact*, all and only φ things are β.[26] But then (T) claims something false: we can easily cook up infinitely many such properties. This is easy to see, in fact, in Lewisian metaphysics, where a property is just a class of possibilia: starting with the class of actual black, short-necked birds, we can tack on any class we want of non-actual things and obtain a property which meets the test.

Part of the work Lewis has in mind, in his 1983a, for the concept of a natural (or "elite") property, is apparently to solve just such problems. He says there that "in putting forward as comprehensive theories that recognize only a limited range of natural properties, physics proposes inventories of the natural properties instantiated in our world" (1983a, 38).[27] So as not to pre-judge the issue of physicalism (whatever issue that might be), we may write "science" for "physics" here, and "relatively natural" for "natural." The proposal would therefore be that we understand (T) to mean: there is a unique relatively natural property φ instantiated in our world such that all φ things (and note: we need no longer say "all and only") are β. And now we can see that (T) and (T') are not so different as they perhaps seemed. Both say that there is a unique, relatively natural concept—i.e., a real essence—*raven*, instantiated in our world, such that, in the circumstances actually prevalent there, the things falling under that concept (possessing that essence) are black, short-necked birds.[28] Thus we arrive at something like Lakatos's understanding of (i), not because all scientifically interesting statements must be interpreted causally, nor because of the supposed all-pervasiveness of *ceteris paribus* clauses, but simply because a theory cannot have concepts of its own (to which it gives experiential import) unless it makes such claims about essences. If, therefore, we are not to follow Lakatos in taking (i) (and all other scientifically interesting claims) as irrefutable, we must understand how the data can bear on whether or not a certain relatively natural property exists.

Lewis himself offers few hints. His suggestions as to what, metaphysically speaking, might constitute naturalness in a property are not helpful in this respect, nor are they meant to be. He does mention, as an "excellent" empirical reason to think that certain natural properties are instantiated, that they are needed to formulate a satisfactory system of natural laws (1983a, 38). He does not explain why this reason is so excellent; he could not, without explaining what makes a system of laws satisfactory. But the main problem is, in any case, that this reason can lead only to the adoption of new natural properties, not to the rejection of old ones. This is in line with the whole Carnapian tradition, which focused always on concept formation, not concept destruction. And yet, as we have seen, and as Lewis recognizes (see 1984, 66), an account of the progress of modern science would have to be above all an account of the latter process.[29]

One person who, perhaps unwittingly, comes close to offering such an account is Kuhn (1974). There, uncharacteristically (he himself calls it a change in his "mode of discourse"), Kuhn considers an imaginary example. "Imagine," he says,

> that you have been shown and can remember ten birds which have authoritatively identified as swans; that you have a similar acquaintance with ducks, geese, pigeons, doves, gulls, etc.; and that you are informed that each of these types constitutes a natural family. (811)

Of what have you thereby been informed, and how could you learn that you were *mis*informed? To the first question Kuhn answers in a way compatible with Lewis's "excellent reason": "Seeing a bird much like the swans you already know, you may reasonably presume that it will require the same food as the others and will breed with them" (ibid.).[30] But, he adds, one's ability to draw such conclusions depends on there being "perceptual space" between the species: one must be able to assume that something "much like" a swan actually is one. Hence he answers the second question: the alleged information could be infirmed "by the discovery of a number of animals (note that more than one is required) whose characteristics bridge the gap between swans and, say, geese by barely perceptible intervals" (ibid.). This recalls fig. 2, which undermines (ii) precisely by showing that there is no "perceptual space" where (ii) requires it. In the context of (i), it suggests something like fig. 3 (in which each point represent an observed bird).

Such data would not logically compel us to give up on our concept *raven*: we could simply draw a circle in the lower left corner and call everything inside a raven, if we so desired. But they would show that *we* must draw this boundary (as Kuhn puts it: that it is "arbitrary" [819 n. 35]). And what better reason could there be for concluding that a concept is unnatural (not a real essence) than the recognition that we can take no guidance from the world in drawing its boundaries?[31]

There is a huge difference, however, between our example, as we understood it, and Kuhn's. Kuhn, despite the change in mode of discourse, fails to rise to the challenge of fictionalization. His imagined case is too much like the real one, in which, as I put it, we *present* our theories with objects which we (authoritatively) identify as ravens. Based on our discussion of (ii), we can see that Kuhn has failed really to imagine (i) as a piece of scientific theory. And, indeed, nobody learns what a post-main-sequence star is by being presented with ten paradigmatic examples and told authoritatively that they form a natural kind. Nor does anyone learn what an electron is by being shown ten paradigmatic electrons,[32] nor did any Newtonian learn what force was by being shown ten paradigmatic forces and told, authoritatively, that they all had something natural in common. Kuhn is correct that learning about such things will often, if not always, involve being exposed to some kind of "concrete examples" (813), but the educational function of such examples, whatever it may be, is clearly nothing like that of his ten authoritatively identified birds. Those ten birds, in effect, just *are* what one is learning: one is learning that these (and things like them) are swans. In the case of genuine theoretical concepts, on the other hand, what one is learning—with or without the aid of concrete examples—is not a (kind of) thing, but a theory.

Hence if *raven* were a true theoretical concept, learning what a raven is would involve, and be a part of, learning a theory like (T) or (T'). And in that case the question Kuhn goes on to ask—namely, what would make us "embrace" or "commit to" a generalization like (i)—falls away. (More precisely: it gains a transcendental answer.) The theory must, directly or indirectly, commit us to generalizations like (i) because it must claim its own concepts, i.e. must give them experiential import. We cannot, therefore, take from Kuhn's example what he wants us to. But it does show something important about what a theory must claim. It is not enough, namely, in describing a natural property, to give or deduce a list of characteristics peculiar to it: one must also claim that it is surrounded by "perceptual space." We saw how, on the negative side, this allowed the data of fig. 3 to undermine a theory like (T) or (T'). But it also explains, on the positive side, why such a theory might be confirmed by data like those of fig. 4.

Now, however, we are in a position to understand Lewis's "excellent reason" for believing that certain concepts are (relatively) natural. For what is it, after all, that makes a law or system of laws satisfactory? A law would be unsatisfactory if it led to false empirical predictions (committed us to false judgments). But even if we side with Popper (against Putnam, Lakatos and others) on the possibility of falsification, simply avoiding it is not enough to make a universal generalization into a satisfactory law. As Goodman (1983, 17–27) long ago pointed out, a claim like

(iii) All birds in this box, Γ, are black.

(where Γ is the proper name of some box) cannot, in our current situation, be accepted as a law, no matter how many birds are in Γ and no matter how many of them are black. Why not? The temptation is to claim that (iii) is somehow insufficiently universal, perhaps because it mentions an individual object or spatiotemporal region, but this will not do, for reasons discussed by Hempel (1965, 342) and Nagel (1961, 57–9), among others. The real problem, as Goodman points out and as Hempel agrees, is that the predicate "bird in the box Γ" is not "projectible"—or, in Lewis's language: that it does not name a particularly natural property. What Lakatos and Kuhn notice, in different ways, about (i) is that, in our current situation, it suffers from exactly the same problem: it is not much of a law (relatively unlawlike), and therefore without scientific interest, because our concept *raven* is not very projectible (relatively unnatural). Or, to put the same point in a way which I hope will sound more plausible: if someone asks, "Why is this thing black?", and the answer is, "Because it's a raven, and all ravens are black," then that is not much of a scientific explanation. How could it be? After all, identifying something as a raven consists, as Kuhn correctly points out, of something like comparing it to authoritatively identified paradigms. In fact, there is officially exactly one such paradigm: the so-called lectotype of the species *Corvus corax*. No one actually learns what a raven is by being exposed to exactly that one specimen, or to any one fixed set of specimens, but Kuhn is right to say that there might as well be, say, ten ravens which we use for that purpose. And surely "resembling these ten things" is not a projectible predicate. This is why, as we noted to begin with, finding a raven which is not black, though it might be surprising (as could be an unexpected refutation of (iii)), could violate no law and thus could not falsify or disconfirm any of our scientific theories. But the situation would be different had we some theory such as (T) or (T'): such a theory would itself claim that *raven* is a relatively natural concept, and thus make our assertion of (i) far more like an assertion of natural law.[33]

3. Science and Superstition

It emerges, therefore, that (i), regarded as a judgment of scientific theory, would actually be tested in the same way as (ii). But it may still seem that there are other kinds of scientific research, involving more straightforward testing. A feature of my examples which might raise this suspicion is my repeated assumption that the data shown were the only kind available for making the conceptual distinction we want to make. Things like this do happen: in fact, we have little if any empirical basis for classifying stars other than color (spectral type) and luminosity. Still it can also happen that some distinction we need is already established and we now want to test something further. Say, for example, that we have already distinguished between ravens and swans based on beak length and nesting habits. Can't we now entertain the further judgment that all ravens are black? And wouldn't the testing of this judgment, finally, proceed as inductivists and/or Popperians imagine? That is: wouldn't we proceed to go out and find ravens (which we would identify by their beak lengths and nesting habits) and then check to see if they are black?

We need to be careful about exactly what case we're imagining. In general we can think of a theory as introduced into a situation in which certain properties are taken to be relatively natural: say, color and luminosity, or albedo, neck length, beak length, and nesting habits (as opposed, for example, to some gruesome counterparts of these). In introducing its own concepts, the theory proposes some other properties as yet more natural. The theory of stellar evolution, for example, proposes that we classify stars according to mass and age, and (T') proposes a classification of birds by ecological niche. As for (T): given the way we were forced to understand its existence-and-uniqueness claim, we should see it as proposing that ravenness per se is a natural property. But a (scientifically testable) theory cannot simply require us to adopt its new concepts. Rather, it must give those concepts empirical import by implying that, under the conditions we take mostly to hold at our world, there will some recognizable pattern somewhere in our data. In the simplest cases the required pattern will be a simple clumping, as in fig. 4, but, as fig. 1 shows, it may be more complicated. In any case, the pattern will require some "perceptual space" in some diagram like fig. 4 or 1, in which the axes are defined by our old concepts (the properties which we already took to be relatively natural). There need not, however, be perceptual space on *every* such diagram. So, for example, (T') could safely claim that ravens and swans overlap in albedo (e.g. that ravens are black, and so are some swans), so long as it also implied they could be clearly distinguished in other ways—for example, by neck length and/or beak length.

The relevance of these considerations to the question at hand is as follows. If *raven* is a theoretical concept of (T'), and (T') implies (under certain approximations, etc.) that (for the most part, given our background assumptions) ravens are black, then, no matter how many other other ways it may supply for picking them out, there is no chance of falsifying (i) by finding a white raven, nor of confirming it by finding black ones. It would always remain the case, in this fictional scenario, that, unless we are somehow induced to change our mind about what we are (at least temporarily) inclined to think is the case at our world, we would have to take the fact that a given bird is not black as evidence that it is not a raven. That is: it would still be impossible for us to check ravens and see if they are black. Data refuting (undermining) (i) would still have to resemble those in fig. 3, showing that we do not, taking *all* relevant properties into account, have any objective way of distinguishing between ravens and other birds. If we want a different kind of case, therefore, we have to imagine one in which (T'), although it does imply various things about ravens—enough for us to test the naturalness of the concept *raven*—fails to imply anything about raven albedo. Assuming that *raven* survives such tests, could we not then go on to hypothesize (i)? And couldn't we then check its truth by a simple process of induction and/or attempted falsification?

There is nothing to stop us, logically speaking, from forming any hypothesis about anything. I can hypothesize that there are only black birds in the box Γ, that there is a penny in my pocket, that, when I leave my office, I will see a raven flying on my left. If I were now (in my current situation) to entertain these hypotheses, then it would indeed be possible simply to confirm or disconfirm them. Nor would they be foolish or implausible. But, I will claim, they would not, in the way I might now entertain them, belong to any modern scientific theory (though they might, if correct, be among the data to which such theories are responsible[34]). If (i) turns out to be such a hypothesis as this, then we will be back in the position Lakatos describes: with a judgment which is easily testable only because it is trivial and of no scientific interest.

To understand why such hypotheses are not scientifically interesting, it will be easier, once again, to start with a more realistic example. Consider the following, taken from an early paper by Charles Sanders Peirce:

(iv) Bleeding tends to cure cholera.

Peirce discusses the following argument supporting (iv):

> A certain man had the Asiatic cholera. He was in a state of collapse, livid, quite cold, and without perceptible pulse. He was bled copiously. During the process he came out of collapse, and the next morning was well enough to be about. Therefore, bleeding tends to cure cholera. (Peirce 1868, 43–4)

According to Peirce, this is an example of a *good* inference, rather than of a bad one. A possibly good inference, that is—although not, of course, possibly apodictic. Its validity, if it had any, would depend on our lack of knowledge about other cases. "If we knew," he explains, "that recoveries from cholera were apt to be sudden, and that the physician who had reported this case had known of a hundred other trials of the remedy without communicating the result, then the inference would lose all its validity" (44). Now, I do not at all intend to pick out Peirce's analysis as particularly mistaken. In fact, I think it unusually sophisticated. The idea that what makes inductive reasoning possible is our *lack* of knowledge is appealing. Still, Peirce's version of this, according to which the reasoning stated would under some circumstances be scientific, cannot be correct. We do often make inferences like the one he outlines; in all likelihood, they even lead, fairly often, to correct conclusions. But the rationality of a method is not to be established by truth of its conclusions. And the rationality of modern science rests precisely on avoiding such inferences wherever possible. Regarded as a clinical trial, the procedure Peirce describes is deficient in at least two key respects: (1) the sample size is too small; (2) there is no control.[35] Of these, it is (2) that is the more fundamental. If there were a control, we would need a large sample size in order to carry out appropriate statistical tests; since there is none, there are no such tests to be made, and hence no possibility whatsoever of scientifically confirming or falsifying (iv).[36]

To see this more clearly, and to see why I use the strong term "superstition," consider the following, similar in form to (iv):

(v) Having a black cat cross one's path is bad luck.

Suppose my evidence for this is that a black cat once crossed my path, and that I subsequently had bad luck. Or suppose even more: suppose that a black cat crosses my path every day, and that every day, without exception, I have bad luck. On Peirce's analysis, or indeed on any of the usual analyses of induction, I would then have very good reason to believe (v)—even Popper would agree that (v) would then have passed a severe test and hence be very well corroborated.[37] But although, once again, we often do reason in that way, the conclusions which we reach are not scientific theories, but superstitions. Scientists are not now engaged in testing (iv) or (v). But if they were, then the scientific testing would begin only with the examination of cases in which the alleged remedy was *not* applied and of days on which a black cat did *not* cross my path—that is, with the establishment of a control.

Why is this? If we go back to one of the hypotheses I mentioned above—e.g. (iii), or that there is a penny in my pocket—there seems to be no parallel issue: I can test my hypothesis without checking the contents of other boxes or of other people's pockets. What is the difference?

One might respond that (iv) and (v) assert causal relationships, while those other hypotheses do not. Actually, in the case of (v), that is not at all clear: black-cat-crossing might well be a sign or omen or harbinger of bad luck, rather than a cause of it. But, even if we take both (iv) and (v) as causal claims, we need to understand what kind of claim that is. "Cause," like "essence" and "definition," is an old Aristotelian term, not so easy to apply correctly in our current philosophical context. Still, we can make some headway by asking why we don't make causal claims in the case of those other hypotheses. If I hypothesize that there is a penny in my pocket, why not also hypothesize that being in my pocket causes it to be a penny? Perhaps I remember having put a penny in there; since (one might think) a cause must precede its effect in time, being in the pocket as a cause of penniness might then seem superfluous. But, first of all, even if I had no recollection at all of what, if anything, I had previously put in my pocket, I still would not hypothesize that anything was caused to be a penny by being in there. And, second of all, there are other, similar cases, where I *would* claim causation—for example, if I have put a bag of ice cubes in the freezer. They were ice cubes when I put them in, but, nevertheless, being in the freezer is what has caused them now to be ice cubes.

The real reason I am unwilling even to hypothesize a causal relationship in this case, or in the case of (iii), is that I am unwilling to assert the mutual relevance of the concepts involved. Whatever a causal claim is, it at least involves a claim of relevance: I must think that whether something is in my pocket has *something* to do with whether it is a penny or not, if I am so much as to suggest that the former is the cause of the latter. To test such a relevance claim rationally, however, I would indeed have to examine birds not in Γ or things not in my pocket (perhaps pennies out there melt like ice cubes left out on the counter). (iv) and (v) require testing in a controlled experiment because they both do make such claims of relevance—and that is so even if (v) asserts signification, ominousness, or harbingery, rather than causation. Those other hypotheses, on the other hand, can be tested without a control precisely because they do not claim such relevance. But that very fact is what makes them mere guesses, rather than belonging to any theory about the world.

This demand for a control, however, obviously resembles our previous demand for perceptual space. In fact, to call this "resemblance" is to understate the point. Imagine for a moment that both bleeding and being cured of cholera were a matter of continuously variable degree, so that (iv) could be interpreted to mean: sufficient bleeding tends to produce a high degree of curedness. In that case we could draw the results of a successful controlled experiment on a figure something like fig. 4—in fact, *exactly* like fig. 4, only with the horizontal and vertical axes relabeled "amount of bleeding" and "curedness," respectively.[38] The figures are the same because the cases are. (i) as scientifically interesting says that there are (at least) two (mostly) distinct kinds of birds: in the simplest case, as represented in (fig. 4), those that are black and short-necked and those that are neither. It claims, that is, that albedo and neck-length are in a certain way relevant to each other among birds. Just so, (iv) is scientifically interesting if it says (in the simplest case) that there are two kinds of patients: those who have been bled and are cured, and those who have not and are not. In both cases the issue is the relative naturalness of a concept. In the case of (i) we introduce a term, "raven," for that concept. In the case of (iv) we do not, but the decision to subject (iv) to scientific testing implies that such a term could be chosen, and that, if we did introduce it, we would intend to designate by it a property of patients which is more natural than either *bled* or *cured of cholera*. That is: the decision to test (iv) scientifically implies that we think there is a *mechanism* which links the treatment to the cure, and the successful testing of (iv) involves implicitly something like a reduction. If (iv) fails, on the other hand, we are led to say that there is no such mechanism, i.e. that there is no one (relatively) natural property common to all and only cured cholera patients who have been bled.

The suspicion raised at the beginning of this section is therefore unfounded: there are not really two different kinds of research here. If the concept *raven* is already legitimated by way of a theory which gives it empirical import—say (T')—then we could test further hypotheses about ravens which are not implied by the theory: say, for example, that they are short-beaked. Such a hypothesis could fail under test without thereby undermining the concept *raven*. But if it is a scientifically interesting hypothesis, rather than a mere guess, it must mean, not just that all ravens happen in fact to have short beaks, but that their ravenness is relevant to their beak length. That is: it must propose that there is some (relatively) natural property, the presence of which implies (under normal circumstances, etc.) that its bearer is both a raven and short-beaked. And we can test for the existence of such a property only if the hypothesis is further interpreted to claim that there is perceptual space on the relevant diagram—in the simplest case, by implying that (for the most part, under normal circumstances) whatever lacks the property in question will be neither short-beaked nor a raven. We might introduce a new term for this property, or we might, as in the case of (iv), leave it nameless, or we might attach it to one of our existing terms—say, the term “raven” (in which case we might not give notice, and might not even say, if asked, that the meaning of “raven” had changed). This last possibility could actually be quite convenient, despite the confusion it would cause to us philosophers, who are not free to juggle our own technical terminology in this way. In any case, whatever our terminological decisions, we would have to test our hypothesis in the same way we test (ii)—that is, by exposing our new concept to the possibility of being illegitimated by the data.

4. Conclusion

Could these few mostly imaginary cases be enough to settle the question of what modern science is, or of what makes it distinctively rational? Actually I doubt we know how even to begin answering that question, which concerns, not the method of testing in science, but the much more difficult case of testing in philosophy. Despite the many attempts to declare the contrary by fiat, I take it as obvious that philosophy doesn’t, at least for the time being, much resemble a science—not even in the grossest respects. It seems unlikely, therefore, that we should hope to legitimate philosophical concepts such as *science* and *philosophy* in the same way as scientific concepts such as *main-sequence* and *post-main-sequence*, and thus further unlikely that statistical issues of sample size will have the same bearing on philosophical judgments as they do on scientific ones. So, in advance of further discussion, I will refrain from either claiming or disowning conclusiveness.

Still, I do hope to have prepared the ground for the reopening of some important questions. One, which I explicitly announced above, would be the interpretation and evaluation of Kuhn, and in particular of Kuhn’s normal science (which he characterizes as an attempt to force nature into preformed, inflexible “conceptual boxes” [1996, 5, 24]). Others which I have mentioned in passing include the nature of mathematics (pure and applied) in general, the role of statistics in particular, and the political prerequisites and effects of science. That last issue, especially, might have very far reaching implications, if (as I would suggest) the conjunction of such prerequisites and effects ought simply to be called “modernity.” And is not modernity—in art or politics, as much as in science—precisely a condition in which our conceptual boxes are seen to stand in need of justification, and at least occasionally discarded for lack of it? But this, obviously, goes beyond what I can take up here.

NOTES

1. This is not to say that those two lines of thought exhaust my intellectual debts, here or elsewhere. Nor is it intended as a rejection of more recent developments in philosophy of science, many of which I will have reason to mention below. I do think, however, that the issues at stake here are both important and (for many historical and philosophical reasons) not squarely faced in the more recent literature.

2. For the sake of brevity in what follows I will omit the translation into the formal mode (e.g., "use of terms in sentences" for "use of concepts in judgments," etc.).

3. Since the identification of these concepts and judgments depends (in principle) on scientific results about human beings (see, e.g., Carnap 1974, §67, pp. 91–3; 1932, 438; 1936, 454–5), it should be clear at the outset that the aim here is not to convert someone who doubts the authority of science, but to justify that authority to someone already confronted with it.

4. At least, this is one attempt at giving a good sense to that phrase. As Putnam (1962, 220–24) points out, the intended meaning is not always clear.

5. Whatever else may be said about Hacking's sustained defense of what *he* (no doubt with every right) calls "observation terms" against the charge of what *he* (again, very plausibly) calls "theory-ladenness" (1983, ch. 10), it does not speak to this issue at all. More relevant may be the treatment in Azzouni 2000, but that work is both too rich and too different from mine in basic approach to be usefully discussed in the space available here.

6. Other kinds of problems could be raised. Because Popper makes the primary issue methodological, i.e. practical, he is vulnerable to a claim that the social order resulting from and/or needed to maintain science in his sense is politically undesirable. Feyerabend's later works mount such a critique. I will not address this or other political issues directly here.

7. Popper himself considered that Kuhn's description of "normal science" was perfectly accurate as a description of an anomalous, degenerate activity, which had (unfortunately) become common in recent years: "a kind of modern blemish on my essentially routine-free picture of science" (1974, 1146).

8. See Lewis 1986, 24–5; 116–8; 121; 1983a, 35–7; cf. Popper 1959, 282; 430–31, and see also Lakatos 1974, 253; 256–7; 260–61 and n. 117.

9. The "once and for all" is unfair. As I noted above, the question of what observations are possible for beings like ourselves is itself empirical. Moreover, later Carnap entertains various proposals for the form and content of the observation language (see 1932, 438–40; 1956, 40–41), thus different proposals as to what concepts should count as meaningful—though, since such proposals are to be weighed practically, this is a philosophical, not a scientific,

issue. None of this, however, amounts to the testing of individual concepts (or systems of concepts) for legitimacy. (In particular, no reasonable proposal or plausible physiological/psychological discovery would allow "raven" and "green" but disallow "non-raven" or "grue.")

10. My hunch is that our name for this method is "applied mathematics," because mathematics itself is the pure method of concept expansion (cf. the very suggestive Buzaglo 2002, especially pp. 76–80; 88–9; 109–115), and that this explains how and why mathematics is applied in modern science. But I will not even begin to argue for those surprising claims here.

11. Obviously this is risky. I consider the risk in question essential to philosophy as such. See Nietzsche 1930, §381, pp. 300-301.

12. In fact, ancient: see Galen, *De temperamentis*, 589K; Porphyry, *Isagoge*, *CAG* 4.1:12,26–13,1.

13. In real life there might be other kinds of data available. I will come back to that issue below.

14. Note that the new, more natural characteristics can indeed be gruesome with respect to the old ones. For example (oversimpfying in several ways), we might define *having a mass of one solar mass* as: *yellow, dim and less than 10 billion years old, or red, bright and more than ten billion years old.*

15. An older and better term would be *Aufhebung*. (Think of using the concepts *eastern* and *western* to find that the earth is round.)

16. See points (4) and (5) in his 1974, 983.

17. Cf. Putnam 1962, 219: "A theoretical term, properly so-called, is one which comes from a scientific *theory* (and the almost untouched problem, in thirty years of writing about 'theoretical terms,' is what is *really* distinctive about such terms)." This section addresses a version of that problem.

18. What Putnam (1962, 216) actually suggests is that, in place of the distinction between observational and theoretical terms, one might consider one of many other distinctions, among them "'new' terms vs. 'old' terms" and "technical terms vs. non-technical ones." Lewis, like me, is interested in the first of these distinctions, but only in the case where the new terms are *theoretical*: "introduced by a given theory *T* at a given stage in the history of science" (79).

19. Implicitly in the notorious §127 (on assigning colors to space-time points); explicitly in §155 (on the elimination of the basic relation *Er*).

20. I.e., because such a definition would not be *formal einwandfrei*: see ibid., §96, p. 135.

21. How strong or long-lasting this inclination is will vary from case to case. We may be willing to give up some risky and ill-motivated approximations rather quickly, for example. But actually *making* such approximations at all means putting oneself, at least temporarily, in the epistemological position described in the text, and to "twist" one's way out of *all* of them would be to rob them of their methodological point.

22. As Lewis also points out, the problem in question is not so new and depends on no particularly technical results of model theory. In fact, Carnap's attempt to eliminate *Er* in his 1974, §§153–5, runs into precisely the

same problem. Lewis's solution is different from, but related to, Carnap's, which relies on a proposed new logical constant, *fund*. For a searching discussion of related arguments see Azzouni 2000, pt. 3.

23. See, however, van Fraassen 1997. In private correspondence on this issue van Fraassen and I have unfortunately been unable to come to complete agreement.

24. In the following discussion I implicitly adopt Lewis's counterpart theory, rather than Kripke's theory of trans-world identity (which may be more familiar to some readers). See Lewis 1986, ch. 4 for extensive discussion. I myself do not consider Lewis's arguments there compelling, but nevertheless find his approach clearer and more perspicuous in almost every way. Most importantly for present purposes, it makes for a very clear distinction between the modal behavior of individuals and of universals. Cf. LaPorte 2004, 38–9.

25. Otherwise, take two β things, b_1 and b_2. Then the property of being b_1, the property of being b_2, and the property of being b_1 or b_2 will all work (and perhaps we should also count the empty property as a fourth—in which case the conclusion should be, not that there must be only one β thing, but that there can't be any).

26. This is Lewis's own intention: see his discussion of the possibility that the theoretical role fulfilled by a certain property at our world is at other worlds fulfilled by other properties (1970, 86–7).

27. Lewis also (1984) uses natural properties, quite differently, to supplement global descriptivism, which I think has contributed to confusion about these issues.

28. However, (T') is still preferable: it allows room for us to be right both about our theory and about which world this is and nevertheless find that (i) is false because we have used inappropriate approximations, etc., to derive predictions—i.e., have failed at a task of the form Putnam (1974, 261) calls "Schema III"; see also Azzouni 2000, pt. 1, §2.

Note how crucial is the qualification "real" in the phrase "real essence." Any arbitrarily chosen class of possibilia defines (or rather, in Lewisian terms, *is*) a property essential to its members. But this essence is nominal. Cf. LaPorte 2004, 11–12 (LaPorte's own discussion of the distinction, pp. 49–50, relies too heavily, I think, on Locke's particular views about what kinds of *res* there can be, and about how a *nomen* can be assigned or meaningfully used).

29. See also Quine 1969 (which also fails to explain how the process works). It is this omission on Lewis's part which leaves him open to the objection that, since the concepts with which science begins are historically and biologically conditioned, the methods of science are inherently unsuited to a search for natural properties: see Elgin 1995, 293–5, also Azzouni 2000, 194–5 (though the issue there is ostensibly quite different).

30. This is also quite in line with the whole tradition of literature on concept formation, as can be seen, e.g., from Hempel 1952, 53; see also Popper, 1974, 983.

31. (i) would also be undermined, in a different way, if we failed to find any black, short-necked birds at all. From a (formal) logical point of view, that would cast no doubt on (i)—but it certainly *would* cast doubt on the legitimacy of the concepts involved. Nothing can be a natural property instantiated in our world unless it is instantiated in our world. (I.e., although it is possible and convenient, in formal logic, to understand universal judgments as not entailing existential ones, that is not possible in *transcendental* logic.)

32. I say this despite Hacking's example (1983, 179) of the technician who knows no physics but can recognize positrons better than a physicist can. The technician knows what a positron is, not by *having learned what it is*, but by deferring to an expert. (This is something like the opposite of Putnam's well-known example of "gold.")

33. Cf. on this point Nagel 1961, 48; 66 and Popper 1994, §13, pp. 30–31; 1959, 427–8.

It is no defect, incidentally, of our current biological theories that they fail to provide an account like (T'). Given evolutionary theories of speciation, a theory like (T') is unnecessary and, because unnecessary, impossible: results like those in fig. 4, according to such theories, are expected due to chance alone. But I cannot discuss this topic further here.

34. This is related to Popper's distinction between theories, which are "specifically" universal (lawlike), and "low level hypotheses," which may be only "numerically" universal (see 1994, §13, especially n. 1; 1959, §22, p. 67, n. *1). Unfortunately, however, this area of Popper's thought is not as clear as it might be.

35. I will ignore the special problems which arise in medicine due to placebo effects.

36. Statistical techniques (and their informal equivalents) obviously play a central role in scientific methodology as I am sketching it. This is not the place to examine that role more fully.

37. Strictly speaking: that it would be relatively very well corroborated—relative, that is, to its maximum possible degree of corroboration (which in the case of such a particular judgment is low).

38. Other possible outcomes would support more complicated conclusions. If there were many data points on the lower right (but still many on the lower left and none on the upper left) we would have evidence that the cure was effective, but also that *cholera patient* is not as natural a property as we had thought: only *some* patients are cured if sufficiently bled.

A different continuous-valued interpretation of (iv) would be: the more a patient is bled, the more cured she will be. To get an analogous case, we could introduce a term "raven-swanness" with a theory like

(T') … & the ecological niche Δt is filled by organism of raven-swanness t & ...

(where lower raven-swanness means being more of a raven and higher raven-swanness means being more of a swan). (i), interpreted as claiming a positive correlation between albedo and raven-swanness, might then follow from (T'') (given the proper background assumptions, etc.). (T'') is even less like our actual biological theories than is (T'). But it is far more like our actual theories of stellar structure and evolution.

BIBLIOGRAPHY

Azzouni, J. (2000), *Knowledge and Reference in Empirical Science*. New York: Routledge.

Buzaglo, M. (2002), *The Logic of Concept Expansion*. Cambridge: Cambridge University Press.

Carnap, R. (1932), "Die physikalische Sprache als Universalsprache der Wissenschaft", *Erkenntnis* 2: 432–65.

Carnap, R. (1936), "Testability and Meaning", *Philosophy of Science* 3, 419–71.

Carnap, R. (1956), "The Methodological Character of Theoretical Concepts", in Feigl, H. et al., (Eds.), *Minnesota Studies in the Philosophy of Science*, Vol. 1. Minneapolis: University of Minnesota Press, 1–74.

Carnap, R. (1974), *Der logische Aufbau der Welt*. Hamburg: Felix Meiner, fourth edition.

Elgin, C. Z. (1995), "Unnatural Science", *Journal of Philosophy* 92: 289–302.

Goodman, N. (1983), *Fact, Fiction and Forecast*. Cambridge, MA: Harvard University Press, fourth edition.

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

Hempel, C. G. (1952), *Fundamentals of Concept Formation in Empirical Science*. Vol. 2 (7) of *International Encylopedia of Unified Science*. Chicago: University of Chicago Press.

Hempel, C. G. (1965), "Aspects of Scientific Explanation", in *Aspects of Scientific Explanation, and other Essays in the Philosophy of Science*. New York: Free Press, 331–496.

Kuhn, T. S. (1974), "Logic of Discovery or Psychology of Research?". In Schilpp (1974), 798–819.

Kuhn, T. S. (1996), *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press, third edition.

Lakatos, I. (1970), "Falsification and the Methodology of Scientific Research Programs", in Lakatos, I. and Musgrave, A., (Eds.), *Criticism and the Growth of Knowledge*. New York: Cambridge University Press. Reprinted in Worrall, J. and Currie, G., (Eds.), *The Methodology of Scientific Research Programs*. Vol. 1 of *Philosophical Papers*. New York: Cambridge University Press, 1978, 8–101.

Lakatos, I., (1974), “Popper on Demarcation and Induction”, in Schilpp (1974), 241–273.

LaPorte, J. (2004), *Natural Kinds and Conceptual Change*. Cambridge: Cambridge University Press.

Lewis, D. K. (1970), “How to Define Theoretical Terms”, *Journal of Philosophy* 67. Reprinted in *Philosophical Papers*, Vol. 1. New York: Oxford University Press, 1983, 78–95.

Lewis, D. K. (1983a), “New Work for a Theory of Universals”, *Australasian Journal of Philosophy* 61. Reprinted in Lewis (1999), 8–55.

Lewis, D.K. (1983b), *Philosophical Papers*. Vol. 1. New York: Oxford University Press.

Lewis, D. K. (1984), “Putnam’s Paradox”, *Australasian Journal of Philosophy* 84. Reprinted in Lewis (1999), 56–77.

Lewis, D. K. (1986), *On the Plurality of Worlds*. Cambridge, MA: Blackwell.

Lewis, D. K. (1999), *Papers in Metaphysics and Epistemology*. New York: Cambridge University Press.

Nagel, E. (1961), *The Structure of Science: Problems in the Logic of Scientific Explanation*. New York: Harcourt, Brace & World.

Nietzsche, F. W. (1930), *Die Fröhliche Wissenschaft: “La Gaya Scienza”*. Leipzig: Alfred Kröner Verlag.

Peirce, C. S. (1868), “Some Consequences of Four Incapacities”, *Journal of Speculative Philosophy* 2, 39–72.

Popper, K. R. (1959), *The Logic of Scientific Discovery*. London: Hutchinson. Reprinted New York: Routledge, 1992.

Popper, K. R. (1974), “Replies to my Critics”, in Schilpp (1974), 961–1197.

Popper, K. R. (1989), “Three Views Concerning Human Knowledge”, in *Conjectures and Refutations: The Growth of Scientific Knowledge*. New York: Routledge, fifth edition, 97–119.

Popper, K. R. (1994), *Logik der Forschung*. Tübingen: Mohr Siebeck, tenth edition.

Putnam, H. (1962), “What Theories Are Not”, in Nagel, E. et al., (Eds.), *Logic, Methodology and Philosophy of Science*. Stanford, CA: Stanford University Press. Reprinted in Putnam (1979), 215–27.

Putnam, H. (1974), “The ‘Corroboration’ of Theories”, in Schilpp (1974). Reprinted in Putnam (1979), 250–69.

Putnam, H. (1979), *Mathematics, Matter and Method*. Vol. 1 of *Philosophical Papers*. New York: Cambridge University Press.

Quine, W. V. (1969), “Natural Kinds”, in *Ontological Relativity and Other Essays*. New York: Columbia University Press, 114–38.

Schilpp, P. A., (Ed.), (1974), *The Philosophy of Karl Popper*. La Salle, IL: Open Court.

van Fraassen, B. (1997), “Putnam’s Paradox: Metaphysical Realism Revamped and Evaded”, *Philosophical Perspectives* 11, 17–42.

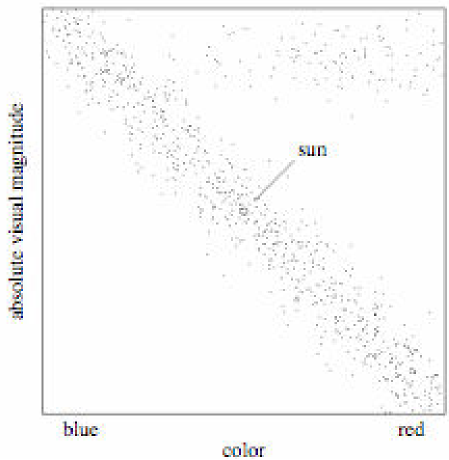

Figure 1: Hertzsprung-Russell diagram (confirming).

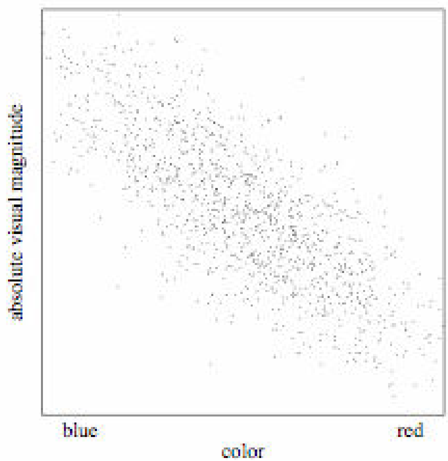

Figure 2: Hertzsprung-Russell diagram (undermining).

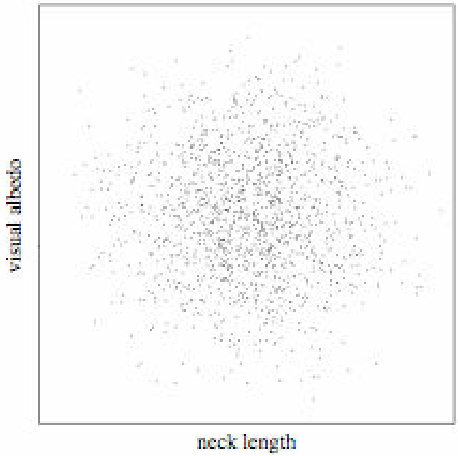

Figure 3: Ravens and swans (undermining).

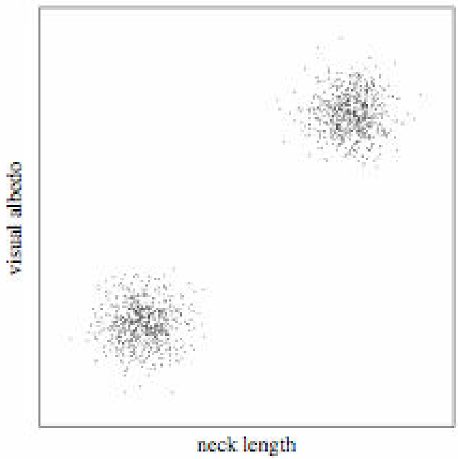

Figure 4: Ravens and swans (confirming).