THE SCIENTIFIC IMAGE

BAS C. VAN FRAASSEN

(1980)

CLARENDON PRESS · OXFORD

unobservable distinction, this distinction has no importance. The point at issue for the realist is, after all, the reality of the entities postulated in science. Suppose that these entities could be classified into observables and others; what relevance should that have to the question of their existence?

Logically, none. For the term 'observable' classifies putative entities, and has logically nothing to do with existence. But Maxwell must have more in mind when he says: 'I conclude that the drawing of the observational–theoretical line at any given point is an accident and a function of our physiological make-up, ... and, therefore, that it has no ontological significance whatever.'[6] No ontological significance if the question is only whether 'observable' and 'exists' imply each other—for they do not; but significance for the question of scientific realism?

Recall that I defined scientific realism in terms of the aim of science, and epistemic attitudes. The question is what aim scientific activity has, and how much we shall believe when we accept a scientific theory. What is the proper form of acceptance: belief that the theory, as a whole, is true; or something else? To this question, what is observable by us seems eminently relevant. Indeed, we may attempt an answer at this point: to accept a theory is (for us) to believe that it is empirically adequate—that what the theory says *about what is observable* (by us) is true.

It will be objected at once that, on this proposal, what the anti-realist decides to believe about the world will depend in part on what he believes to be his, or rather the epistemic community's, accessible range of evidence. At present, we count the human race as the epistemic community to which we belong; but this race may mutate, or that community may be increased by adding other animals (terrestrial or extra-terrestrial) through relevant ideological or moral decisions ('to count them as persons'). Hence the anti-realist would, on my proposal, have to accept conditions of the form

> If the epistemic community changes in fashion Y, then my beliefs about the world will change in manner Z.

To see this as an objection to anti-realism is to voice the requirement that our epistemic policies should give the same results independent of our beliefs about the range of evidence accessible to us. That requirement seems to me in no way rationally compelling; it could be honoured, I should think, only through a thoroughgoing scepticism or through a commitment to wholesale leaps of faith. But we cannot settle the major questions of epistemology *en passant* in philosophy of science; so I shall just conclude that it is, on the face of it, not irrational to commit oneself only to a search for theories that are empirically adequate, ones whose models fit the observable phenomena, while recognizing that what counts as an observable phenomenon is a function of what the epistemic community is (that *observable* is *observable-to-us*).

The notion of empirical adequacy in this answer will have to be spelt out very carefully if it is not to bite the dust among hackneyed objections. I shall try to do so in the next chapter. But the point stands: even if observability has nothing to do with existence (is, indeed, too anthropocentric for that), it may still have much to do with the proper epistemic attitude to science.

§3. *Inference to the Best Explanation*

A view advanced in different ways by Wilfrid Sellars, J. J. C. Smart, and Gilbert Harman is that the canons of rational inference require scientific realism. If we are to follow the same patterns of inference with respect to this issue as we do in science itself, we shall find ourselves irrational unless we assert the truth of the scientific theories we accept. Thus Sellars says: 'As I see it, to have good reason for holding a theory is *ipso facto* to have good reason for holding that the entities postulated by the theory exist.'[7]

The main rule of inference invoked in arguments of this sort is the rule of *inference to the best explanation*. The idea is perhaps to be credited to C. S. Peirce,[8] but the main recent attempts to explain this rule and its uses have been made by Gilbert Harman.[9] I shall only present a simplified version. Let us suppose that we have evidence E, and are considering several hypotheses, say H and H'. The rule then says that we should infer H rather than H' exactly if H is a better explanation of E than H' is. (Various qualifications are necessary to avoid inconsistency: we should always try to move to the best over-all explanation of all available evidence.)

It is argued that we follow this rule in all 'ordinary' cases; and that if we follow it consistently everywhere, we shall be led to scientific realism, in the way Sellars's dictum suggests. And surely there are many telling 'ordinary' cases: I hear scratching in the wall, the patter of little feet at midnight, my cheese disappears—and I infer that a mouse has come to live with me. Not merely that these apparent signs of mousely presence will continue, not merely that all the observable phenomena will be as if there is a mouse; but that there really is a mouse.

Will this pattern of inference also lead us to belief in unobservable entities? Is the scientific realist simply someone who consistently follows the rules of inference that we all follow in more mundane contexts? I have two objections to the idea that this is so.

First of all, what is meant by saying that we all *follow* a certain rule of inference? One meaning might be that we deliberately and consciously 'apply' the rule, like a student doing a logic exercise. That meaning is much too literalistic and restrictive; surely all of mankind follows the rules of logic much of the time, while only a fraction can even formulate them. A second meaning is that we act in accordance with the rules in a sense that does not require conscious deliberation. That is not so easy to make precise, since each logical rule is a rule of permission (*modus ponens* allows you to infer *B* from *A* and (if *A then B*), but does not forbid you to infer (*B or A*) instead). However, we might say that a person behaved in accordance with a set of rules in that sense if every conclusion he drew could be reached from his premisses via those rules. But this meaning is much too loose; in this sense we always behave in accordance with the rule that any conclusion may be inferred from any premiss. So it seems that to be following a rule, I must be willing to believe all conclusions it allows, while definitely unwilling to believe conclusions at variance with the ones it allows—or else, change my willingness to believe the premisses in question.

Therefore the statement that we all follow a certain rule in certain cases, is a *psychological hypothesis* about what we are willing and unwilling to do. It is an empirical hypothesis, to be confronted with data, and with rival hypotheses. Here is a rival hypothesis: we are always willing to believe that the theory which best explains the evidence, is empirically adequate (that all the observable phenomena are as the theory says they are).

In this way I can certainly account for the many instances in which a scientist appears to argue for the acceptance of a theory or hypothesis, on the basis of its explanatory success. (A number of such instances are related by Thagard.[8]) For, remember: I equate the acceptance of a scientific theory with the belief that it is empirically adequate. We have therefore two rival hypotheses concerning these instances of scientific inference, and the one is apt in a realist account, the other in an anti-realist account.

Cases like the mouse in the wainscoting cannot provide telling evidence between those rival hypotheses. For the mouse *is* an observable thing; therefore 'there is a mouse in the wainscoting' and 'All observable phenomena are as if there is a mouse in the wainscoting' are totally equivalent; each implies the other (given what we know about mice).

It will be countered that it is less interesting to know whether people do follow a rule of inference than whether they ought to follow it. Granted; but the premiss that we all follow the rule of inference to the best explanation when it comes to mice and other mundane matters—that premiss is shown wanting. It is not warranted by the evidence, because that evidence is not telling *for* the premiss *as against* the alternative hypothesis I proposed, which is a relevant one in this context.

My second objection is that even if we were to grant the correctness (or worthiness) of the rule of inference to the best explanation, the realist needs some further premiss for his argument. For this rule is only one that dictates a choice when given a set of rival hypotheses. In other words, we need to be committed to belief in one of a range of hypotheses before the rule can be applied. Then, under favourable circumstances, it will tell us which of the hypotheses in that range to choose. The realist asks us to choose between different hypotheses that explain the regularities in certain ways; but his opponent always wishes to choose among hypotheses of the form 'theory T_i is empirically adequate'. So the realist will need his special extra premiss that every universal regularity in nature needs an explanation, before the rule will make realists of us all. And that is just the premiss that distinguishes the realist from his opponents (and which I shall examine in more detail in Sections 4 and 5 below).

The logically minded may think that the extra premiss can be bypassed by logical *léger-de-main*. For suppose the data are that all facts observed so far accord with theory *T*; then *T* is one possible explanation of those data. A rival is *not-T* (that *T* is false). This rival is a very poor explanation of the data. So we *always* have a set of rival hypotheses, and the rule of inference to the best explanation leads us unerringly to the conclusion that *T* is true. Surely I am committed to the view that *T* is true or *T* is false?

This sort of epistemological rope-trick does not work of course. To begin, I am committed to the view that T is true or T is false, but not thereby committed to an inferential move to one of the two! The rule operates only if I have decided not to remain neutral between these two possibilities.

Secondly, it is not at all likely that the rule will be applicable to such logically concocted rivals. Harman lists various criteria to apply to the evaluation of hypotheses *qua* explanations.[10] Some are rather vague, like simplicity (but is simplicity not a reason to use a theory whether you believe it or not?). The precise ones come from statistical theory which has lately proved of wonderful use to epistemology:

> H is a better explanation than H' (*ceteris paribus*) of E, provided:
> (a) $P(H) > P(H')$—H has higher probability than H'
> (b) $P(E/H) > P(E/H')$—H bestows higher probability on E than H' does.

The use of 'initial' or *a priori* probabilities in (a)—the initial plausibility of the hypotheses themselves—is typical of the so-called *Bayesians*. More traditional statistical practice suggests only the use of (b). But even that supposes that H and H' bestow definite probabilities on E. If H' is simply the denial of H, that is not generally the case. (Imagine that H says that the probability of E equals $\frac{3}{4}$. The very most that *not-H* will entail is that the probability of E is some number other than $\frac{3}{4}$; and usually it will not even entail that much, since H will have other implications as well.)

Bayesians tend to cut through this 'unavailability of probabilities' problem by hypothesizing that everyone has a specific subjective probability (degree of belief) for every proposition he can formulate. In that case, no matter what E, H, H' are, all these probabilities really are (in principle) available. But they obtain this availability by making the probabilities thoroughly subjective. I do not think that scientific realists wish their conclusions to hinge on the subjectively established initial plausibility of there being unobservable entities, so I doubt that this sort of Bayesian move would help here. (This point will come up again in a more concrete form in connection with an argument by Hilary Putnam.)

I have kept this discussion quite abstract; but more concrete arguments by Sellars, Smart, and Putnam will be examined below. It

should at least be clear that there is no open-and-shut argument from common sense to the unobservable. Merely following the ordinary patterns of inference in science does not obviously and automatically make realists of us all.

§4. *Limits of the Demand for Explanation*

In this section and the next two, I shall examine arguments for realism that point to explanatory power as a criterion for theory choice. That this is indeed a criterion I do not deny. But these arguments for realism succeed only if the demand for explanation is supreme—if the task of science is unfinished, *ipso facto*, as long as any pervasive regularity is left unexplained. I shall object to this line of argument, as found in the writings of Smart, Reichenbach, Salmon, and Sellars, by arguing that such an unlimited demand for explanation leads to a demand for hidden variables, which runs contrary to at least one major school of thought in twentieth-century physics. I do not think that even these philosophers themselves wish to saddle realism with logical links to such consequences: but realist yearnings were born among the mistaken ideals of traditional metaphysics.

In his book *Between Science and Philosophy*, Smart gives two main arguments for realism. One is that only realism can respect the important distinction between *correct* and *merely useful* theories. He calls 'instrumentalist' any view that locates the importance of theories in their use, which requires only empirical adequacy, and not truth. But how can the instrumentalist explain the usefulness of his theories?

> Consider a man (in the sixteenth century) who is a realist about the Copernican hypothesis but instrumentalist about the Ptolemaic one. He can explain the instrumental usefulness of the Ptolemaic system of epicycles because he can prove that the Ptolemaic system can produce almost the same predictions about the apparent motions of the planets as does the Copernican hypothesis. Hence the assumption of the realist truth of the Copernican hypothesis explains the instrumental usefulness of the Ptolemaic one. Such an explanation of the instrumental usefulness of certain theories would not be possible if *all* theories were regarded as merely instrumental.[11]

What exactly is meant by 'such an explanation' in the last sentence? If no theory is assumed to be true, then no theory has its usefulness explained as following from the truth of another one—granted. But would we have less of an explanation of the usefulness of the

This sort of epistemological rope-trick does not work of course. To begin, I am committed to the view that T is true or T is false, but not thereby committed to an inferential move to one of the two! The rule operates only if I have decided not to remain neutral between these two possibilities.

Secondly, it is not at all likely that the rule will be applicable to such logically concocted rivals. Harman lists various criteria to apply to the evaluation of hypotheses *qua* explanations.[10] Some are rather vague, like simplicity (but is simplicity not a reason to use a theory whether you believe it or not?). The precise ones come from statistical theory which has lately proved of wonderful use to epistemology:

H is a better explanation than H' (*ceteris paribus*) of E, provided:
(a) $P(H) > P(H')$—H has higher probability than H'
(b) $P(E/H) > P(E/H')$—H bestows higher probability on E than H' does.

The use of 'initial' or *a priori* probabilities in (a)—the initial plausibility of the hypotheses themselves—is typical of the so-called *Bayesians*. More traditional statistical practice suggests only the use of (b). But even that supposes that H and H' bestow definite probabilities on E. If H' is simply the denial of H, that is not generally the case. (Imagine that H says that the probability of E equals $\frac{3}{4}$. The very most that *not-H* will entail is that the probability of E is some number other than $\frac{3}{4}$; and usually it will not even entail that much, since H will have other implications as well.)

Bayesians tend to cut through this 'unavailability of probabilities' problem by hypothesizing that everyone has a specific subjective probability (degree of belief) for every proposition he can formulate. In that case, no matter what E, H, H' are, all these probabilities really are (in principle) available. But they obtain this availability by making the probabilities thoroughly subjective. I do not think that scientific realists wish their conclusions to hinge on the subjectively established initial plausibility of there being unobservable entities, so I doubt that this sort of Bayesian move would help here. (This point will come up again in a more concrete form in connection with an argument by Hilary Putnam.)

I have kept this discussion quite abstract; but more concrete arguments by Sellars, Smart, and Putnam will be examined below. It should at least be clear that there is no open-and-shut argument from common sense to the unobservable. Merely following the ordinary patterns of inference in science does not obviously and automatically make realists of us all.

§4. *Limits of the Demand for Explanation*

In this section and the next two, I shall examine arguments for realism that point to explanatory power as a criterion for theory choice. That this is indeed a criterion I do not deny. But these arguments for realism succeed only if the demand for explanation is supreme—if the task of science is unfinished, *ipso facto*, as long as any pervasive regularity is left unexplained. I shall object to this line of argument, as found in the writings of Smart, Reichenbach, Salmon, and Sellars, by arguing that such an unlimited demand for explanation leads to a demand for hidden variables, which runs contrary to at least one major school of thought in twentieth-century physics. I do not think that even these philosophers themselves wish to saddle realism with logical links to such consequences: but realist yearnings were born among the mistaken ideals of traditional metaphysics.

In his book *Between Science and Philosophy*, Smart gives two main arguments for realism. One is that only realism can respect the important distinction between *correct* and *merely useful* theories. He calls 'instrumentalist' any view that locates the importance of theories in their use, which requires only empirical adequacy, and not truth. But how can the instrumentalist explain the usefulness of his theories?

> Consider a man (in the sixteenth century) who is a realist about the Copernican hypothesis but instrumentalist about the Ptolemaic one. He can explain the instrumental usefulness of the Ptolemaic system of epicycles because he can prove that the Ptolemaic system can produce almost the same predictions about the apparent motions of the planets as does the Copernican hypothesis. Hence the assumption of the realist truth of the Copernican hypothesis explains the instrumental usefulness of the Ptolemaic one. Such an explanation of the instrumental usefulness of certain theories would not be possible if *all* theories were regarded as merely instrumental.[11]

What exactly is meant by 'such an explanation' in the last sentence? If no theory is assumed to be true, then no theory has its usefulness explained as following from the truth of another one—granted. But would we have less of an explanation of the usefulness of the

Laws and Symmetry

Bas C. van Fraassen

(1990)

CLARENDON PRESS · OXFORD

INTRODUCTION

We have now seen that any philosophical account of laws needs a good deal in the way of metaphysics to do justice to the concept at all. We have also seen that, as a result, any such account founders on the two fundamental problems of identification and of inference. The extant accounts come to grief additionally in their attempts even to meet the most basic criteria relating to science and explanation. Their promises have all proved empty.

But there are still those traditional arguments, which conclude first that there *must be* laws of nature, and secondly, that we *must believe* that there are such laws. In modern terms, the threats are these: without laws of nature we can make no sense of science and its achievement, nor of rational expectation, and must succumb inevitably to the despair of scepticism.

In this Part I shall answer these epistemological arguments. More constructively, I shall propose a programme for epistemology and for philosophy of science which will allow them to flourish in the absence of laws of nature or belief therein.

6.
Inference to the Best Explanation: Salvation by Laws?

> As a man of science you're bound to accept the working hypothesis that explains the facts most plausibly.
>
> The Arch-Vicar of Belial, to Dr Poole, in Aldous Huxley, *Ape and Essence*.[1]

The inference from the phenomena that puzzle us, to their best explanation, appears to have our instinctive assent. We see putative examples of it, in science and philosophy no less than in ordinary life and in literature.

It is exactly this pattern of inference, to the best explanation offered, that philosophers have drawn upon to claim confirmation of laws. They support this appeal in two ways: by pointing to the failures of traditional ideas of induction and by arguing that this inference pattern is the true rock on which epistemology must build. After examining their reasons, I shall argue instead that they would build on shifting sands. As long as the pattern of Inference to the Best Explanation—henceforth, IBE—is left vague, it seems to fit much rational activity. But when we scrutinize its credentials, we find it seriously wanting.[2] (For those more interested in IBE itself than in its connection with laws, sections 2 and 3 may be skipped.)

If both induction and IBE fail as rational basis for opinion and expectation of the future, traditional epistemology is indeed in serious difficulty. But rather than proclaim the death of epistemology, and submit either to an irenic relativism or to sceptical despair, I shall try to show in the next chapter that a new epistemology has been quietly growing within the ruins of the old (as well as show that it issues in a still more drastic critique of IBE).

I. ON THE FAILURES OF INDUCTION[3]

One contention, common to many writers, is that without some such concept as laws of nature, we can make no sense of rational expectation of the future. I have earlier presented this point in what I take to be its primordial form: if anyone says that there is no reason for the observed regularities, then he can have no reason to expect them to continue.

This assertion clearly denies the cogency of induction in a narrow sense (belief based on straight extrapolation from the data) while it holds out the hope of induction in a very broad sense (rationally formed expectation of the future). Let us here use the term 'induction' everywhere in its narrow sense: the procedure whose independent rationality friends of laws tend to deny. I may as well add at once that I agree with them on the critical point. My discussion will aim to underline their legitimate objections, but to show simultaneously how their critique is misdirected, and where it rests on dubitable premisses of their own.

Here is the ideal of induction: of a rule of calculation, that extrapolates from particular data to general (or at least ampliative) conclusions. Parts of the ideal are (*a*) that it is a *rule*, (*b*) that it is *rationally compelling*, and (*c*) that it is *objective* in the sense of being independent of the historical or psychological context in which the data appear, and finally, (*d*) that it is *ampliative*. If this ideal is correct, then support of general conclusions by the data is able to guide our opinion, without recourse to anything outside the data—such as laws, necessities, universals, or what have you.

Critique of this ideal is made no easier by the fact that this rule of induction does not exist. The rule was indeed baptized—presumably after conception, but before it was ever born. Sketches of rules of this sort have been presented, with a good deal of hand-waving, but none has ever been seriously advocated for long.[4] Every generation of philosophers, beginning with Aristotle, has seen that mere numerical extrapolation in any specific form, cannot be the rule described in our ideal. Criticisms brought forward in this century however, exhibit difficulties to plague every possible realization of the ideal. If the reader is already convinced of the inadequacy of induction in the narrow sense, there is no need to read the rest of this section.

What is extrapolation? Example of the alien die

Consider a die, which is to be tossed ten times, and the hypothesis that all ten tosses will come up *ace*. Let the evidence so far be that it has come up *ace* for the first seven tosses. Now, how could the rule of induction relate these data to the hypothesis? Should it tell us that all ten will be like the first seven? Rules of extrapolation can't be expected to do well if some relevant evidence is left unstated, and we do have other information about human dice. So suppose we found this die on an alien planet, and 'ace' is the name we give to one of the six sides of this geometric cube of unknown composition.

Of course, reader, you are still unwilling to suggest that the rule should tell us to infer that all tosses (or equivalently, the last three) will come up *ace*. After all, in this situation, *you* would not infer that. But perhaps the rule should be sophisticated beyond anything Bacon and Newton, the great advocates of this ideal, could imagine. Let it tell us the probability of the hypothesis, bestowed on it by the data. Then it could be asserted that a rational person must follow the rule of induction, in the sense that it provides him with the probability that the hypothesis is true, given the evidence.

This suggestion marks quite a shift, because it takes us from induction as extrapolation from mere numbers to something much more general. But every discussion of induction is forced to this. Suppose, for example, that instead we try to maintain the rule in as simple a form as possible, with as one corollary: if all instances have been favourable, and you have no other evidence, then believe that the next instance will be favourable as well. You will immediately insist, surely, that 'believe' must be qualified here, if you are ever to follow it. Believe with what confidence? Believe to what degree? Are a hundred instances not better evidence than ten, even if all have been favourable? Any such reaction replaces the simple rule with a more sophisticated one, of the order of probability assignments. Of course we should hasten to accept all worthy suggestions for improvement of the rule, rather than insist on beating a dead horse.

In both forms, the same problem about induction appears very clearly. It is that, being a *rule* it must have certain structural features—and as a result, its extrapolation from any data will be heavily influenced by what it does with small increases in data. But how it does that, must be either uninformative or arbitrary. That is the dilemma this ideal always foundered on.

To illustrate this, begin with the naïve 'straight rule': believe that the ratio of *A* to *B* overall equals the ratio observed so far. That tells us to believe that the sun will always rise. Unfortunately, it also tells us to expect all *aces* as soon as we have seen a single toss of the alien die, if it came up *ace*. We can't very well suggest that another ratio is any more plausible. So we must fiddle with the confidence: always believe what this rule tells you, but believe it weakly after one toss and strongly after many tosses.[5] Now what is needed is an exact prescription of what beliefs, degrees of belief, or confidence I should have at the outset; and an auxiliary rule about how this should change with the outcome of each new toss. The former could be perfect neutrality of opinion. (I do not assume it must be a precise subjective probability or anything like it.) But then the auxiliary rule must still say exactly how much non-neutrality there should be after one toss; and indeed, after $n + 1$ tosses (as a function of n, the previous outcomes, and the new outcome). Now you can look at the numbers and ratios as much as you like, but they will give you no clue at all to this auxiliary rule for massaging your confidence in the observed ratio. Myriads, continua, of such functions exist, and however little they diverge in the small, they lead to widely different consequences down the line. You can try to remain a little neutral among them: the more you do so, the *less arbitrary* will your rule of induction be, but also the *less informative*. Now, next problem: try and formulate a measure of balance between arbitrariness and informativeness. There you will again find a continuum of functions to choose from, and you will again confront the dilemma presented by the spectrum from capricious choice to trivializing neutrality.[6]

There are many other problems with the ideal of induction, even if this is (as I think) its fundamental flaw.[7] As to the empiricists who followed this banner *sans* device, their hope was placed in an empty promise. But does there indeed lie a better hope in the mobilization of laws to found rational expectation, as Dretske, Armstrong, and Tooley contend?

2. DRETSKE ON THE REMARKABLE CONFIRMATION OF LAWS

The scheme we inspected above might be called simple or bare induction; many have been the proposals to replace it by more complex schemes. One such is Dretske's proposal that we take note of the (supposed) remarkable tendency of laws to become well confirmed on the basis of very little evidence. He uses exactly the reflections and types of examples exhibited above, to sketch a rival picture of how we can rationally go beyond the evidence.

It appears, at first sight, that laws are beyond our epistemic grasp. The sun has risen every day; this appears to confirm the universal generalization that it always has and always will. But if we speculated that this was a matter of law, we would be asserting more: something beyond and in addition to the universal statement. The conforming instances support the generalization, but surely they do not support anything beyond that? Dretske tells us that this apparent problem is a pseudo-problem, resting on a mistaken empiricist epistemology. In fact, he claims, it is quite the other way round. This sort of evidence, of positive instances, does not at all support the universal statement, if taken in isolation. But it does support the hypothesis that it is a law that the phenomenon always occurs in that same way.

Dretske calls this conclusion 'mildly paradoxical' ('Laws of Nature', p. 267), but it seems more than that. Surely if it is a law that *A* then it is also true that *A*; hence I can become no less confident that it is so than that it is a law. The air of paradox is perhaps removed, if we take Dretske to be attacking the conception of evidential support that was implicit in the proffered argument. This conception appears to be the old ideal of purely numerical induction.

To tackle also more sophisticated epistemic schemes, Dretske makes the preliminary point that raising our probabilities may not amount to real confirmation.[8] Let us use the alien die, introduced in the preceding section, to illustrate his point. Suppose I begin with the initial assumption that the die is fair. Then my initial probability that it will come up *ace* all ten tosses, is very low—$(\frac{1}{6})^{10}$ which equals about six in a thousand million. After I have seen seven *aces* come up in in a row, while maintaining this assumption, the probability that the last three will come up, is still the same as it was: $(\frac{1}{6})^3$. But this is now also the new probability of the proposition that all ten come up *ace*. The probability of that proposition has therefore become 6^7 times—about 300 000 times—higher than it was. Our probability for the universal statement has increased dramatically—but we are in no better position to predict it that inference to a good, with luck the best, explanation has force *even* in the sphere of metaphysical analysis'—(my italics)). To support (P_6) he does not see the need for more than rhetorical questions: 'If making such an inference is not rational, what is?' (p. 53); 'To infer to the best explanation is part of what it is to be rational. If that is not rational, what is?' (p. 59).

In sum, therefore, Armstrong has reached powerful conclusions on the basis of an assumption which is supported solely by a challenge to those who would doubt it.

In the next two sections I hope to meet Armstrong's challenge. I shall argue that inference to the best explanation cannot be a recipe for rational change of opinion. And then I shall try to answer the question, 'If that is not rational, what is?'

4. WHY I DO NOT BELIEVE IN INFERENCE TO THE BEST EXPLANATION[10]

There are many charges to be laid against the epistemological scheme of Inference to the Best Explanation. One is that it pretends to be something other than it is. Another is that it is supported by bad arguments. A third is that it conflicts with other forms of change of opinion, that we accept as rational.

Still, the verdict I shall urge is a gentle one. Someone who comes to hold a belief because he found it explanatory, is not *thereby* irrational.[11] He becomes irrational, however, if he adopts it as a rule to do so, and even more if he regards us as rationally compelled by it. The argument for this conclusion will be begun here and concluded in the next chapter.

What IBE really is

Inference to the Best Explanation is not what it pretends to be, if it pretends to fulfil the ideal of induction. As such its purport is to be a rule to form warranted new beliefs on the basis of the evidence, the evidence alone, in a purely objective manner. It purports to do this on the basis of an evaluation of hypotheses with respect to how well they explain the evidence, where explanation again is an objective relation between hypothesis and evidence alone.

It cannot be *that* for it is a rule that only selects the best among

the historically given hypotheses. We can watch no contest of the theories we have so painfully struggled to formulate, with those no one has proposed. So our selection may well be the best of a bad lot. To believe is *at least* to consider more likely to be true, than not. So to believe the best explanation requires more than an evaluation of the given hypothesis. It requires a step beyond the comparative judgment that this hypothesis is better than its actual rivals. While the comparative judgment is indeed a 'weighing (in the light of) the evidence', the extra step—let us call it the ampliative step—is not. For me to take it that the best of set X will be more likely to be true than not, requires a prior belief that the truth is already more likely to be found in X, than not.

There are three possible reactions to this, each of which argues that IBE must be allowed *nevertheless* to play the role of leading to a new belief extrapolated from one's evidence. Clearly any such reaction must focus on the ampliative step, because the above objection is independent of the method of evaluation (of explanatoriness) that is used.

Reaction 1: Privilege

The first consists in a claim of privilege for our genius. Its idea is to glory in the belief that we are by nature predisposed to hit on the right range of hypotheses.[12]

We recognize here the medieval metaphysical principle of *adequatio mentis a rei.* Contemporary readers will not be happy to accept it as such, I think, and would hope for a justification. Such a justification could take two forms, allied respectively with naturalism and rationalism in epistemology.

The naturalistic response bases the conclusion on the fact of our adaptation to nature, our evolutionary success which must be due to a certain fitness. But in this particular case, the conclusion will not follow without a hypothesis of pre-adaptation, contrary to what is allowed by Darwinism.[13] The jungle red in tooth and claw does not select for internal virtues—not even ones that could increase the chance of adaptation or even survival beyond the short run. Our new theories cannot be more likely to be true, merely given that we were the ones to think of them and we have characteristics selected for in the past, because the success at issue is success in the future. The moths in industrial England became dark, not because they began to have more dark offspring but because the light ones were more vulnerable.

The rationalist response must be patterned after Descartes's argument for the correspondence of ideas to reality. Alvin Plantinga has suggested such a reason for privilege: given other beliefs about God, such as that we are made in his image, it is only reasonable to believe that we are specially adapted to hit on the truth when we come up with our (admittedly limited) guesses at explanation. Plantinga applies this even to belief in propositions and other abstract entities. But it takes more than a generally agreed concept of God to get this far. For even if he created us naturally able to perceive the truth about what is important for us in his eyes (perhaps to discern love from lust, or charity from hypocrisy, in ourselves), this may not extend to speculations about demons, quarks, or universals. *Privilege* is consistent, but seems incapable of either naturalistic or rationalist support.

Although it does not count as an argument, I should also point out that Privilege is entirely at odds with Empiricism. By this I mean the position that experience is our one and only source of information. Clearly this leaves open a great deal—experience may be very rich in its possible varieties; on the other hand, the information it brings us may not come as if in the voice of an angel, but in dubious and defeasible form. However that may be, the position sets one clear limit: if we do have innate or instinctive or inborn expectations, we'll be just lucky if they lead us aright, and not like lemmings into the sea. However basic or natural our inclination may be toward, for example, more satisfying explanations, that inclination itself cannot be relevant information about their content's truth.

Reaction 2: Force majeure

The *second* reaction pleads *force majeure*: it is to try and provide arguments to the effect that we *must* choose among the historically given significant hypotheses. To guide this choice is the task of any rule of right reason. In other words, it is not because we have special beliefs (such as that it will be a good thing to choose from a certain batch of hypotheses), but because we must choose from that batch, that we make the choice.

The *force majeure* reaction is, I think, doomed to fail. Circumstances may force us to act on the best alternative open to us.

They cannot force us to believe that it is, *ipso facto*, a good alternative.

Perhaps it will be objected that the action reveals the belief, because the two are logically connected. And in a general way that is certainly true. But it is exactly in situations of forced choice that action reveals very little about belief. Think of William James's example of a walker in the mountains who has the choice of jumping over a crevasse, or remaining for the night. Suppose that both a fall and exposure mean almost certain death. The prize is equal too: life itself. She jumps. Does this reveal that it was very likely to her that she would get across? Not at all, for even a very low chance would be reasonable to take in this case. Or if the Princess must open one of N doors, the Tigers skulking behind all but one, we can only conclude about the door she opened, that it seemed to her no less probable than $1/N$ that it would lead to freedom. And if there might be a Tiger behind each, and she is forced to choose nevertheless, we cannot even conclude that.

In the case of science we certainly observe theory choice. But just what belief is revealed there? Let us look carefully at the practice, and see what belief it entails, if any.

Scientists designing a research programme, bet their career and life's satisfactions on certain theoretical directions and experimental innovations. Here they are forced to choose between historically given theoretical bases. They are forced by their own decision to be scientists, to opt for the best *available* theory, by their own light. What beliefs are involved in this, can be gauged to some extent from their goals and values. Does this scientist feel that his life will have been wasted if he has spent it working on a false theory? Then he must feel that Descartes's and Newton's lives were wasted. Or does he feel that his life will have been worth while if he has contributed to progress of science, even if the contribution consisted in showing the limits and inadequacies of the theories he began with and the discovery of some new phenomenon that every future science must save? In the latter case his choice between theories, as basis for research, does not reveal any tendency to belief in their truth.

Reaction 3: Retrenchment

The *third* reaction is to *retrench*: 'Inference to the Best Explanation' was a misnomer, and the rule properly understood leads to a revision of judgement much more modest than inference to the truth of the favoured hypothesis. The charge should be that I have construed the rule of inference to the best explanation too naïvely. Despite its name, it is not the rule to infer the truth of the best available explanation. That is only a code for the real rule, which is to allocate our personal probabilities with due respect to explanation. Explanatory power is a mark of truth, not infallible, but a characteristic symptom.

This retrenchment can take two forms. The *first form* is that the special features which make for explanation among empirically unrefuted theories, make them (more) likely to be true.[14] The *second form* is that the notion of rationality itself requires these features to function as relevant factors in the rules for rational response to the evidence. I will take up both forms—the first in the remainder of this section, and the second in the next chapter. Let us note beforehand that the first must lean on *intrinsic explanatoriness*, which can be discerned prior to empirical observations, and the second specifically on *explanatory success* after the observational results come in. What the criteria are for either, we shall leave up to the retrencher.

Retrenchment, form 1

Is the best explanation we have, likely to be true? Here is my argument to the contrary.

I believe, and so do you, that there are many theories, perhaps never yet formulated but in accordance with all evidence so far, which explain at least as well as the best we have now. Since these theories can disagree in so many ways about statements that go beyond our evidence to date, it is clear that most of them by far must be false. I know nothing about our best explanation, relevant to its truth-value, except that it belongs to this class. So I must treat it as a random member of this class, most of which is false. Hence it must seem very improbable to me that it is true.

You may challenge this in two ways. You may say that we do have further knowledge of our own best explanation, relevant to its truth-value, beyond how well it explains. I'm afraid that this will bring you back to the reaction of Privilege, to glory in the assumption of our natural or historical superiority. Or you may say that I have construed the reference class too broadly. That is fair. The class of rivals to be considered should be on a par with

our own, in ways that could affect proportions of truth. So we cannot include in it two theories, each having the same disagreement with ours at point X, but then disagreeing with each other at point Y, on which ours has nothing to say. So for each statement ours makes, beyond the evidence we have already, we can include only one theory disagreeing with that statement, for every way to disagree with it. But of course, there is only one true way to agree or disagree at this point. So the conclusion, that most of this class is false, still stands.

David Armstrong, replying to a version of this argument, writes 'I take it that van Fraassen is having a bit of fun here.'[15] Yes, I had better own up immediately: I think I know what is wrong with the above argument. But my diagnosis is part and parcel of an approach to epistemology (to be explained in the next chapter), in which rules like IBE have no status (*qua* rules) at all. As a critique of IBE, on its own ground, the above argument stands.

One suspicious feature of my above argument is that it needed no premisss about what features exactly do make a hypothesis a good explanation. Let us consider a contrary argument offered by J. J. C. Smart, which does focus on one such feature, simplicity. I think that both argument and counterargument will be rather typical of how any such debate could go (for any choice of explanatory feature). Smart begins as follows:

> My argument depends on giving a non-negligible a priori probability to the proposition that the universe is simple. . . .
>
> Let p = the observational facts are as if there are electrons, etc.
> q = the universe is simple
> r = there really are electrons, etc.
>
> We can agree . . . that $P(p) > P(pr)$, as of course we have to! But I want to say $P(pr/q) > P(p\bar{r}/q)$.[16]

That is, the probability *on the supposition that the universe is simple*, is greater for our best explanation which entails that the phenomena will continue as before, than it is for a denial of that explanation which agrees on those phenomena.

The reason for this judgement, on Smart's part, must hinge on a connection between explanation and simplicity. So it does. It is exactly because the explaining hypothesis is simple (as it must be, to qualify as explanation) that the supposed simplicity of the universe makes the hypothesis more probable than its denial (under the further supposition that the explanation is right about all the phenomena, even those to come). And it seems at first sight plausible to say that a supposed simplicity in nature, will make simple theories more likely to be true.

But this plausibility derives from an equivocation. If the simplicity of the universe can be made into a concrete notion by specifying objective structural features that make for simplicity, then I can see how Smart may have arrived at the opinion that the universe is (probably) simple. For there can be evidence for any objective structural feature. But if the universe's simplicity means the relational property, that it lends itself to manageable description by us (given our limitations and capacities) I cannot see that. The successes we've had are all successes among the descriptions we could give of nearby parts of the universe, and of the sort which our descriptive abilities allow. Suppose it is true that the frog can distinguish only the grossest differences between objects at rest, but can notice even small moving objects. Then his success in catching insects flying by is no index of how many potential prey and potential enemies sit there quietly watching him.

Could simplicity in the first sense, which we might have reason to surmise, make more probable the simpler-for-us among the accounts we can give? That depends in part on how much simplicity of theories has to do with simplicity of the world as described by those theories. But suppose there is an intimate connection—unlike, say, in literature where the simplicity and economy of form in poetry does not limit it to simplicity of subject, in comparison with prose. Then still the allocation of probabilities is effectively prevented by a very modest consideration. Simplicity is global. A part of a structure, which is very simple overall, may be exceedingly complex considered in isolation. Here is a simply described set: that of all descendants of Geoffrey of Monmouth, alive today. Now, try to describe it purely in terms of features recognizable today (by geographical location, blood type, what have you) without reference to the past. Considered as part of a historical structure, the set is easily delineated; viewed short-sightedly in the twentieth century it is not. Similarly for the simplicity of the universe as a whole, and those aspects and parts of it on which our sciences focus. The simplicity a situation has in virtue of being part of a simple universe, does not make more likely any simple putative description of it by itself alone.

Retrenchment, form 2

I already raised the possibility of a sophisticated, probabilistic version of the rule briefly above. Combining the ideas of personal probability and living by rules, the new rule of IBE would be a recipe for adjusting our personal probabilities while respecting the *explanatory* (as well as predictive) success of hypotheses. This will be investigated in the next chapter, after we have taken a much closer look at the representation of opinion.

5. WHAT GOOD IS THERE IN THIS RULE?

There must clearly be a solid basis for the appeal and renown that IBE has enjoyed over the years. It is important to ferret this out. But as with any subject, we must carefully separate the inflated claims of philosophical exponents from the grains of common sense which gave those claims their initial appeal. Eventually we must also show that the common-sense part is equally respected in our own account.

If I already believe that the truth about something is likely to be found among certain alternatives, and if I want to choose one of them, then I shall certainly choose the one I consider the best. *That* is a core of common sense which no one will deny.

But how far is it from this common sense to IBE, conceived as cornerstone of epistemology? This rule cannot supply the initial context of belief or opinion within which alone it can become applicable. Therefore it cannot be what 'grounds' rational opinion.

That is only the first point. Next we can see that even if the rule is applicable, we might very well not wish to apply it. Suppose it seems likely to me that one of the first six horses will win the race, and that of these, horse No. 1 is the best. It does not follow at all that I shall then wish to predict that horse No. 1 will win, for this might mean no higher probability than $\frac{1}{5}$ for its winning. Similarly, if I turn away then, and just at the end of the race a great cheer goes up, the best explanation for me of this cheer will be that horse No. 1 has just won. And still I shall be no readier to say that this is what has just happened. If a *force majeure* makes me predict, then I shall indeed say 'No. 1'—but this will certainly not reveal then that I believe it.

A rule which we would often, when it is applicable, prefer not to apply, is not a rule we are following! Common sense will often prevent us from applying IBE; and it could not very well do that if it were the epistemic categorical imperative. Even more important, perhaps, is this: even if the rule is applicable, and we make our choice in accordance with it, we may not be following the rule. This sounds paradoxical, but think: must a choice among hypotheses, even if unconstrained, necessarily be a choice to believe?

In general, the common-sense choice will be in the context of an opinion to the effect that this batch of hypotheses have a balance of certain virtues, and are well fitted to serving our present aims. Likelihood of truth will presumably be among those virtues, or among those aims, but it need not be alone. Informativeness with respect to topics of interest may be another. *When we choose the best, therefore, the choice must be interpreted in terms of the basis for choice.* If likelihood of truth is not the sole basis, then the choice—choice to accept—must not be equated with choice to believe.

How little comfort common sense gives to philosophical fancy! What has happened to extrapolation of general truth from evidence alone? No more than the metaphysician do I think that common sense brings its own clarity with it. So I propose next to discuss the whole subject again in a higher—and more constructive—key. Specifically, I will consider versions of IBE that make full use of the conceptual resources of probability. And yet, its fortunes will not improve thereby.

7.
Towards a New Epistemology

> . . . it is by no means clear that students of the sciences . . . would have any *methodology* left *if* abduction is abandoned. If the fact that a theory provides the best available explanation for some important phenomenon is not a justification for believing that the theory is at least approximately true, then it is hard to see how intellectual inquiry could proceed.
>
> Richard Boyd[1]

So far I have only offered a critique of traditional epistemology with its ampliative rules of induction and inference to the best explanation (abduction). But this mainstream does not constitute the only tradition in epistemology. The seventeenth century gave us besides Descartes and Newton also Blaise Pascal, and from his less systematic writings there sprung a stream that in the succeeding three centuries has become a powerful river: the underground epistemology of *probabilism*.

After introducing its basic ideas, I shall show that it leads us into a much more radical and far-reaching critique than we have seen so far. The rule of IBE, and indeed the whole species of ampliative rules, is incoherent. I shall deliver this critique at the outset through a proxy, a foil, one particular sort of probabilist: the orthodox Bayesian. The rigours of *his* views are however also considerably more than his arguments can demonstrate. I will go on then to propose an epistemology that is certainly still probabilist, but offers a reconciliation with traditional epistemology. It does give room to practices of ampliation beyond the evidence. (In the next chapter, we shall see that it also allows us finally to illuminate how objective chance should guide our personal expectation—remember the fundamental question, concerning chance, which previous chapters had to leave unanswered?—and how it may enter the opinion that guides rational decisions.) To end I shall argue that, despite the ominous warnings of the past, our new epistemology is driven into neither sceptical despair, nor feeble relativism, nor metaphysical realism.