4 In defence of scientific realism

Thus far, I have offered arguments against reductive empiricism, several versions of instrumentalism, either of the eliminative variety or of the Duhemian (non-eliminative) variety. We have seen that the so-called 'Ramsey way' does not offer a stable and satisfactory compromise between realism and instrumentalism. So, the only alternative is to adopt a realist attitude towards the unobservable entities posited by our best theories. If *semantic realism* is adopted, then we have a straightforward answer to the question: what is the world like, according to a given scientific theory? (Or, similarly, what is the world like, if a certain scientific theory is true?) The answer is none other than that the world is the way the scientific theory – literally understood – describes it to be.

This answer seems to have certain implications for epistemological questions. Bluntly put, once semantic realism is adopted, the issue of warranted belief in the existence of unobservable entities seems to take care of itself: insofar as scientific theories are well confirmed, it is rational to believe in the existence of the entities they posit. For, what other than our best theories should we look to in order to decide what it is reasonable to believe about the world? If our best science is not our best guide to our ontological commitments, then nothing is.

The *realist turn* in the philosophy of science since the early 1960s has aimed to remove the last scruples one might have against the confirmability and the actual confirmation of scientific theories. What realists have offered is a battery of arguments which aim to defend a scientific realist attitude towards our best scientific theories, while blocking their opponents' counter-arguments purporting to show that scientific theories cannot be accepted as approximately true. So, the realist turn has aimed to secure the epistemic optimism associated with scientific realism – a view which was explained in the Introduction to this book. In this chapter, I try to show that this attitude of epistemic optimism is well-motivated and warranted.

A central argument in defence of scientific realism is the famous 'no miracle argument' (henceforth NMA) which aims to show that our best scientific theories can be reasonably believed to be approximately true. NMA has found its 'textbook' formulation in these words of Hilary Putnam:

> The positive argument for realism is that it is the only philosophy that does not make the success of science a miracle. That terms in mature scientific theories typically refer (this formulation is due to Richard Boyd), that the theories accepted in a mature science are typically approximately true, that the same terms can refer to the same even when they occurs in different theories – these statements are viewed not as necessary truths but as part of the only scientific explanation of the success of science, and hence as part of any adequate description of science and its relations to its objects.
>
> (1975: 73)

So, NMA aims to defend the realist claim that successful scientific theories should be accepted as true (or, better, near true) descriptions of the world, in both its observable and its unobservable aspects. In particular, the realist claim is that accepting that successful scientific theories describe truly (or, near truly) the unobservable world *best explains* why these theories are empirically successful. That is, it best explains why the observable phenomena are as they are predicted to be by those theories.

As stated by Putnam, NMA is intended to be an instance of inference to the best explanation (henceforth IBE, or abduction). What needs to be explained, the explanandum, is the overall empirical success of science. NMA intends to conclude that the main theses associated with scientific realism, especially the thesis that successful theories are approximately true, offer the best explanation of the explanandum. Hence, they must be accepted precisely on this ground. This IBE-based reading of NMA underwrites the current defence of realism as developed by Richard Boyd and elaborated by me in the present chapter. Hence, I shall call this argument the Putnam–Boyd argument. It has, however, been repeatedly claimed that the Putnam–Boyd argument is viciously circular and begs the question against the critics of realism. For, it is noted, although the critics of realism deny (or simply doubt) that IBE is a reliable inferential method, NMA presupposes its reliability. As Fine (1991: 82) has put it, an IBE-based defence of realism lacks any argumentative force since it employs 'the very type of argument whose cogency is the question under discussion'. Dispersing the charge of vicious circularity and question-beggingness should be a central task in my own defence of realism. But before that, some detailed discussion is required in respect of the structure of the main realist argument. In particular, in the subsequent sections I try to disentangle several versions of NMA. The next two sections motivate and articulate carefully what I take to be the most forceful version of NMA, showing that it can offer a good defence of realism, provided that it is seen as part-and-parcel of a thorough externalist and naturalistic realist epistemological package.

Cosmic coincidences and the success of science

What appear to be variants of NMA had been put forward long before Putnam's slogan appeared by J. J. C. Smart and Grover Maxwell. Smart argued against instrumentalists that they 'must believe in *cosmic coincidence*' (1963: 39). To be sure, he referred to 'phenomenalism about theoretical entities', but took this to be eliminative instrumentalism, i.e. the view that 'statements about electrons, etc., are only of instrumental value: they simply enable us to predict phenomena on the level of galvanometers and cloud chambers' (ibid.).

We have already seen (Chapter 2) that eliminative instrumentalism takes scientific theories to be merely syntactic/mathematical constructs for the organisation of experimental and empirical facts, and for grouping together empirical laws and observations which would otherwise be taken to be irrelevant to one another. On this view, theoretical claims are not even truth-conditioned, i.e. capable of being true or false; nor do theories imply existential commitments to unobservables. The emergence of Craig's theorem coincided with the culmination of this view. For, as we have seen (pp. 22–23), it offers the instrumentalist a systematic way to eliminate theoretical terms.

On the eliminative instrumentalist account, a vast number of ontologically disconnected observable phenomena are 'connected' only by virtue of a purely instrumental theory: they just *happen* to be, and just *happen* to be related to one another in such a way that a Craig-style theory is true. If so, what other than a gigantic coincidence makes a Craigian theory true? Accepting the vast number of purely instrumental connections implied by the Craig-style theory exceeds the limits of tolerance, especially when there is a handy account that does away with all this happenstance. But look at scientific realism, says Smart. It leaves no space for coincidence on a cosmic scale: it is because theories are true and because the unobservable entities they posit exist that the phenomena are, and are related to one another, the way they are. Here is the contrast in Smart's own words:

> Is it not odd that the phenomena of the world should be such as to make a purely instrumental theory true? On the other hand, if we interpret a theory in the realist way, then we have no need for such a cosmic coincidence: it is not surprising that galvanometers and cloud chambers behave in the sort of way they do, for if there are really electrons, etc., this is just what we should expect.
>
> (1963: 39)

One may take Smart's argument to be a version of the 'no miracle' argument put forward by Putnam. At first glance, it seems that we are indeed dealing with one and the same argument. The only difference seems to be lexical: Smart bars cosmic coincidences, while Putnam bars miracles. Smart himself, after all, has also talked about a 'cosmic miracle' (1979: 364). Both arguments, it seems, rely on what they take to be the best explanation of why the observable phenomena are as they are predicted by scientific theories. As a rough approximation, this might be alright. However, if we look carefully at the details of the two arguments, it is pertinent to distinguish Smart's version from the Putnam–Boyd version of the NMA.

Smart's argument is not meant to be an inference to the best explanation. It is more of a general philosophical argument, what is sometimes called a plausibility argument (cf. Smart 1963: 8–12). For Smart, the argument for realism is largely a priori. He takes it that part, at least, of the distinctively philosophical method is to clarify conceptual disputes, i.e. disputes which are not amenable to empirical tests. On this view, the philosopher's job is to offer arguments in favour of each side of the dispute. Consistency is not at stake here, because every position can be made into a consistent one, given enough ingenuity. Rather, the philosopher should aim to examine the plausibility or arbitrariness of each position, especially in those grand disputes that 'affect our overall world view' (Smart 1963: 8). The realist–instrumentalist controversy is conceived by Smart to be such a grand conceptual dispute about the interpretation of scientific theories. Accordingly, Smart's 'no cosmic coincidence' argument relies on primarily intuitive judgements as to what is plausible and what requires explanation. It claims that it is intuitively more plausible to accept realism over instrumentalism because realism leaves less things unexplained and coincidental than does instrumentalism. Its argumentative force, if any, is that anyone with an open mind and good sense could and would find the conclusion of the argument intuitively plausible, persuasive and rational to accept – though not logically compelling: not because one would recognise the argument as an instance of a trusted inferential scheme, but because of intuitive considerations about what is more and what is less plausible.

An analogous argument for realism was offered Maxwell (1962a). To the best of my knowledge, he was the first to appeal explicitly to the success of scientific theories in order to to defend realism. The overall empirical success of science, says Maxwell, is a fact that calls for an explanation. The instrumentalist claim that theories are 'black boxes', which when fed with true observational premisses yield true observational conclusions, would offer no explanation whatsoever of the fact that these 'black boxes' are so successful. In light of this, he claims: 'The only reasonable explanation for the success of theories of which I am aware is that well-confirmed theories are conjunctions of well-confirmed, genuine statements and that the entities to which they refer in all probability exist' (1962a: 18). As he has pointed out elsewhere, the difference between realist and instrumentalist accounts of science is that

> as our theoretical knowledge increases in scope and power, the competitors of realism become more and more convoluted and ad hoc and

> explain less than realism. For one thing, they do not explain why the theories which they maintain are mere cognitively meaningless instruments are so successful, how it is that they can make such powerful, successful predictions. Realism explains this very simply by pointing out that the predictions are consequences of the true (or close true) propositions that comprise the theories.
>
> (Maxwell 1970: 12)

Maxwell's argument differs from Smart's in an interesting way. It includes an attempt to ground the plausibility judgements that are required for the defence of realism and to show that such judgements are not, after all, distinctively philosophical. In a certain sense, Maxwell's argument is the 'bridge' between Smart's a priori argument and the subsequent Putnam–Boyd *naturalistic* version. Maxwell suggests that considerations of simplicity, comprehensiveness and lack of ad hocness are virtues that make judgements displaying them more plausible than judgements lacking them. What is more, Maxwell (1970) gives a Bayesian twist to his argument for realism. He emphasises that on standard probabilistic accounts of confirmation, if two or more mutually inconsistent hypotheses entail the same piece of evidence, the only way in which the evidence can be made to support one hypothesis more than the other(s) is via some kind of initial plausibility ranking of the competing hypotheses. This ranking is then reflected in the prior probabilities ascribed to the competing hypotheses. His argument for realism capitalises precisely on this well-worn fact. Suppose, he says, that both realism (*R*) and instrumentalism (*I*) *entail* that scientific theories are successful (*S*). Then, the likelihoods of realism and instrumentalism are both equal to unity; i.e.:

$$\mathit{prob}(S/R) = \mathit{prob}(S/I) = 1.$$

By Bayes' theorem, the posterior probability of realism is

$$\mathit{prob}(R/S) = \mathit{prob}(R)/\mathit{prob}(S)$$

and the posterior of instrumentalism is

$$\mathit{prob}(I/S) = \mathit{prob}(I)/\mathit{prob}(S),$$

where *prob*(*R*) is the prior probability of realism,
prob(*I*) is the prior of instrumentalism and
prob(*S*) is the probability of the 'evidence', i.e. of the success of science.

Given that *prob*(*S*) is the same for both realism and instrumentalism, any difference in the degree of confirmation of *R* and *I* should reflect a difference in their respective priors. Based on the thought that the realist explanation of the success of science is simpler, more comprehensive and less ad hoc than any instrumentalist attempt at such an explanation, Maxwell (1970: 17–18) argues that the *prior probability* of realism should be much greater than the prior of instrumentalism: i.e. *prob*(*R*)>>*prob*(*I*). Hence, the incremental confirmation of Realism is much greater than that of Instrumentalism.

I think Maxwell's point is two-fold. On the one hand, relying on prior probabilities is a routine aspect of all human judgement. It is also evident in scientific practice itself: not all theoretical hypotheses which entail the same evidence are ranked as equally plausible by scientists. In fact, the very virtues of simplicity, comprehensiveness and lack of ad hocness are those used by scientists to rank competing scientific hypotheses. On the other hand, philosophical problems, such as the realist–instrumentalist dispute are not much more difficult than – nor qualitatively different from – ordinary scientific problems, where no evidence can distinguish between two competing hypotheses. So, they call for the same treatment as ordinary scientific problems. As Maxwell puts it: 'My reasons for accepting realism are of the same kind as those for accepting any scientific theory over others which also explain current evidence' (ibid.).[1]

All this is a bit quick, the reader might think. Prior probabilities might indeed be indispensable in ampliative reasoning. But on what basis, the reader may ask, do we say that realism's prior probability is greater than that of (eliminative) instrumentalism? Since the conclusion of the argument depends crucially on assigning different priors to realism and instrumentalism, this conclusion would have been otherwise had we adopted an initial ordering which favoured instrumentalism over realism. How, then, can this ordering be decided? In particular, is this ordering supposed to be objective or subjective? If the former, then we need some further argument as to why realism is *objectively* more probable than instrumentalism. If the latter, what do subjective degrees of belief, or subjective estimates of prior probability, have to do with the alleged superiority of realism?

What it is correct to stress, I think, is that when it comes to the realism–instrumentalism debate an assignment of higher prior probability to realism can be rational – and hence objective – in these two senses. First, judgements of initial plausibility can be the subject and outcome of rational deliberation. One way, for instance, to argue for the greater initially plausibility of realism is to point out that realism derives much of its plausibility from a judgement which all parties in the realism debate would find rational – the very judgement which underlies the positing of middle-sized material objects. Against eliminative instrumentalism, realists rightly stress a certain analogy – and continuity – between positing middle-sized material objects to account for the orderly and coherent streams of sensory experience and positing scientific unobservables to account for the observable phenomena. If common sense has been the only thing required for the former, then so much the better for the latter. In denying the existence of unobservable entities, eliminative instrumentalists have to adopt 'double existential standards'. But as we have seen (pp. 18–22), there are no good arguments to support such double standards.

Second, judgements of initial plausibility can be rational and objective because they rely on sound expectations. Why is it initially more plausible to interpret scientific theories realistically? Because on an instrumentalist construal – such as mere 'black boxes', syntactic calculi and the like – *there is no reason to expect that theories are capable of being empirically successful*. To be sure, 'black boxes' and the like are constructed so that they systematise known observable regularities. But it does not follow from this that black boxes have the capacity to predict either hitherto unknown regularities or hitherto unforeseen connections between known regularities. Nor can such a thing be expected on any rational ground. However, if the theory is understood realistically, then novel predictions about the phenomena occasion *no surprise*. Realistically understood, theories entail too many novel claims, most of them about unobservables (e.g. that there are electrons, that light bends near massive bodies). It comes as no surprise that some of the novel theoretical facts a theory predicts may be such that they give rise to novel observable phenomena, or that they may reveal hitherto unforeseen connections between known phenomena. For instance, James Clerk Maxwell's theoretical identification of light with an electromagnetic wave predicted a hitherto unknown connection between the laws of light-propagation and the propagation of electric waves. At any rate, it would be very surprising if the causal powers of the entities posited by scientific theories were exhausted in the generation of the already-known empirical phenomena that led to the introduction of the theory. So, on a realist understanding of theories, novel predictions and genuine empirical success are to be expected (given, of course, that the world co-operates).

The fact of the matter is that such judgements as those above have been strong enough to mitigate the force of standard instrumentalist accounts. As we saw in Chapter 2, similar plausibility judgements have been put forward by 'textbook instrumentalists' like Pierre Duhem and Henri Poincaré. Both have argued that novel predictive success – a feature that has not been stressed sufficiently well by Maxwell – is at odds with an eliminative instrumentalist construal of scientific theories as 'racks filled with tools' (Duhem 1906: 334) or as 'simple practical recipes' (Poincaré 1902: 174). This is not surprising: on an instrumentalist account, novel predictive success is, if anything, an accidental feature of theories. Maxwell's argument makes good on precisely this state of affairs. It suggests that scientific realism is the only alternative that overcomes the problem which makes instrumentalism implausible – how novel successful predictions are possible. What, I think, it adds to this suggestion is the following. Once theories are treated as semantic realism suggests, then their novel empirical success can accrue only to the theory's confirmation: the more unlikely the prediction, the greater the incremental confirmation of the theory which makes it available.

There is no reason to doubt that Smart's and Maxwell's arguments undermine drastically the rationale of eliminative instrumentalism. But they are ineffective against sophisticated empiricist positions *à la* van Fraassen (1980; 1989). For a long time eliminative instrumentalism was the dominant alternative to a realist understanding of scientific theories. Smart and Maxwell (and, for that matter, Feigl too) aimed to kill two birds with one stone. Their central point was that the success of scientific theories lent credence to the two theses:

- that scientific theories should be interpreted realistically; and
- that, so interpreted, these theories are well confirmed because they entail well-confirmed predictions.

So, their arguments operate on the assumption that an argument for the realist interpretation of scientific theories can be, *ipso facto*, an argument for *believing* in the existence of the entities they posit. Given what has been said in Chapters 1 and 2 about the fate of reductive empiricism and eliminative instrumentalism, this is a reasonable assumption. Once it is accepted that theories should be interpreted realistically, the only remaining issue is whether these theories are well confirmed. If one and the same argument can establish both, then so much the better for realism.

However, the empiricist position advocated by van Fraassen accepts a realist interpretation of the semantics of scientific theories but challenges the *rationality of belief* in unobservable entities the existence of which these theories, if true, imply. Hence, in a certain sense, van Fraassen's position starts precisely where Smart's and Maxwell's arguments stop: that eliminative and reductive accounts of theoretical commitment in science are wrong-headed and discredited. As we shall see in detail in Chapter 9, one of van Fraassen's central points against scientific realism is that abductive–explanatory reasoning, by means of which theoretical beliefs are formed, cannot be shown to be truth-conducive, and therefore that belief in the approximate truth of particular theories is not rationally compelling. In other words, he questions the reliability of the methods scientists employ to arrive at their theoretical beliefs. On van Fraassen's view, the collapse of eliminative instrumentalism does not make realism the only rational option. An agnostic variety of empiricism is not, *ipso facto*, ruled out: one can always remain *agnostic* as to the truth-value of the particular theoretical descriptions of the world offered by a theory.

Boyd's important contribution to the debates over scientific realism, to which I now turn, is precisely that he has employed and strengthened the 'no miracle' argument in an attempt to defend the reliability and rationality of ampliative–abductive reasoning in science.

The explanationist defence of realism

Boyd's 'explanationist defence of realism' (henceforth EDR) is a programme for the development and defence of a *realist epistemology of science*. Boyd suggests that this epistemology should be thoroughly naturalistic. On the one hand, it should rest on the claim that it is a radically contingent fact about the world that scientific theories can and do deliver theoretical truth. On the other hand, in its attempt to investigate the epistemic credentials of science, and in particular to answer the question why scientific methodology is instrumentally reliable, a realist epistemology of science should employ no methods other than those used by scientists themselves. Boyd's defence of realism is *explanationist* because it is based on the claim that the realist thesis that scientific theories are approximately true is the best explanation of their empirical success. Boyd's naturalism makes his use of the NMA distinctively different from Smart's and (to a lesser extent) from Maxwell's: there is no distinctive philosophical method which is either prior to scientific method or can be used to resolve first-order scientific disputes. In this section I focus on the place of the 'no miracle' argument in EDR.

Boyd[2] has set out to show that the best explanation of the instrumental and predictive success of mature scientific theories is that these theories are approximately true, at least in those respects relevant to their instrumental success. I shall reconstruct the main argument as follows:

> That the methods by which scientists derive and test theoretical predictions are theory-laden is undisputed. Scientists use accepted background theories in order to form their expectations, to choose the relevant methods for theory-testing, to devise experimental set-ups, to calibrate instruments, to assess the experimental evidence, to choose among competing theories, to assess newly suggested hypotheses, etc. All aspects of scientific methodology are deeply theory-informed and theory-laden. In essence, scientific methodology is almost linearly dependent on accepted background theories: it is these theories that make scientists adopt, advance or modify their methods of interaction with the world and the procedures they use in order to make measurements and test theories.
>
> These theory-laden methods lead to correct predictions and experimental success.
>
> How are we to explain this?
>
> The best explanation of the instrumental reliability of scientific methodology is that: the theoretical statements which assert the specific causal connections or mechanisms by virtue of which scientific methods yield successful predictions are approximately true.

NMA is a philosophical argument which aims to defend the reliability of scientific methodology in producing approximately true theories and hypotheses. Its strength, however, rests on a more concrete type of explanatory reasoning which occurs all the time in science. It can be stated as follows. Suppose that a background theory T asserts that method M is reliable for the generation of effect X in virtue of the fact that M employs causal processes $C_1, \ldots, C_n$ which, according to T, bring about X. Suppose, also, that we follow T and other established auxiliary theories to shield the experimental set-up from factors which, if present, would interfere with some or all of the causal processes $C_1, \ldots, C_n$, thereby preventing the occurrence of effect X. Suppose, finally, that one follows M and X obtains. What else can better explain the fact that the expected (or predicted) effect X was brought about than that the theory T – which asserted the causal connections between $C_1, \ldots, C_n$ and X – has got these causal connections right, or nearly right? If this reasoning to the best explanation is cogent, then it is reasonable to accept T as approximately true, at least in those respects relevant to the theory-led prediction of X. To be more precise, more is needed for the acceptance of T as relevantly approximately true. For instance, T is to be contrasted with available alternative hypotheses, and should emerge as *the* best explanation. T should also offer a 'good enough' explanation in its own right, e.g. an explanation which can adequately account for all salient features of the experimental facts.[3] But such considerations are part and parcel of these more concrete applications of explanatory reasoning in science. And although we may not always be in position to choose a hypothesis as clearly the best explanation, that does not entail that we never are.

The relation between this more concrete type of explanatory reasoning in science and the NMA should be clear: successful instances of such reasoning provide the basis (and the initial *rationale*) for this more general abductive argument. However, NMA is not just a generalisation over scientists' abductive inferences. Although itself an *instance* of the method that scientists employ, NMA aims at a broader target: to defend the thesis that Inference to the Best Explanation, or abduction (that is, a *type* of inferential method), is reliable. The (first-order) instances of explanatory reasoning involve the claim that it is reasonable to accept that *particular* theories are relevantly approximately true. NMA is, then, based on these instances to defend the more general claim that science *can* deliver *theoretical truth*. NMA is a kind of *meta*-abduction. The explanandum of NMA is a general feature of scientific methodology – its reliability for yielding correct predictions. NMA asserts that the best explanation of why scientific methodology has the contingent feature of yielding correct predictions is that the theories which are implicated in this methodology are relevantly approximately true.

So, what makes NMA distinctive as an argument for realism is that it defends the achievability of theoretical truth. But how exactly does this argument defend IBE and hence how exactly does NMA become the pivot for a realist epistemology of science? As I have noted, it suggests that the best explanation of the instrumental reliability of scientific methodology is that background theories are relevantly approximately true. These background scientific theories have themselves been typically arrived at by abductive reasoning. Hence, it is reasonable to believe that abductive reasoning is reliable: it tends to generate approximately true theories. This conclusion is not meant to state an a priori truth. The reliability of abductive reasoning is an empirical claim, and if true is contingently so.

Having said this, let me stress that NMA should be suitably qualified. There is enough historical evidence to persuade any *bona fide* realist, first, that scientific theories have encountered many failures as well as successes and, second, that some past theories which once were empirically successful and were accepted as 'best explanations' of the evidence were subsequently abandoned as inadequate and false. In light of this, the realist argument should be qualified in two respects:

1 The realist argument should acknowledge the existence of failures. Their actuality does not impair scientific methodology. Nor does it sever the explanatory link between approximate truth and empirical success – especially novel empirical success. Clearly, the fact that I have occasionally failed to find my lost keys does not entail that a thorough search of the places where they could have been left is not a reliable method for finding lost keys. In any case, realists should concentrate on *particular* theory-led successes – and there are very many of those – and argue that it is *these* successes that require explanation. It is, after all, a salient feature of scientific methodology that it does lead to empirical successes. Things could be otherwise, and scientific theories might have been *total* failures. So, to ask how it is possible at all that scientific theories yield correct predictions, especially novel ones, and to offer explanations of this contingent feature of scientific methodology are essential for understanding science. (The notion of novelty in prediction, to which realists should appeal, is analysed in Chapter 5.)
2 The realist argument should become more local in scope. Accordingly, the main realist point should be the following: although most realists would acknowledge that there is an explanatory connection between a theory's empirical success and its being, in some respects, right about the unobservable world, it is far too optimistic – if at all defensible – to claim that *everything* the theory asserts about the world is thereby vindicated.

So, realists should *refine* the explanatory connection between empirical and predictive success, on the one hand, and truth-likeness, on the other. They should assert that these successes are best explained by the fact that the theories which enjoyed them have *truth-like theoretical constituents* (i.e. truth-like descriptions of causal mechanisms, entities and laws). The theoretical constituents whose truth-likeness can best explain empirical successes are precisely those which are essentially and ineliminably involved in the generation of the predictions and the design of the methodology which brought these predictions about. From the fact that not every theoretical constituent of a successful theory does and should get credit from the successes of the theory it certainly does *not* follow that none do (or should) get some credit. If, on top of that, it is shown that, far from being abandoned, the theoretical constituents of past theories which did essentially contribute to their successes were retained in subsequent theories of the same domain, then the realist case is as strong as it can be. In Chapter 5 this point is explained in detail, since the argument just expressed captures in a nutshell the way in which I try to block the argument from the 'pessimistic induction'.

From this point onwards I assume that the above considerations constitute the intended reading of NMA. EDR has caused a heated discussion among philosophers of science (cf. Laudan 1984; McMullin 1987 and 1991; Musgrave 1988; Newton-Smith 1987; Lipton 1991). As already noted, the main line of criticism is that EDR is viciously circular. Since it employs IBE, critics suggest that it therefore presupposes what needs to be shown – that IBE is a reliable inferential method. Arthur Fine (1986; 1986a; 1991) has summarised and defended this line in the most forceful way. He points out that the realist is 'not free to assume the validity of a principle whose validity is itself under debate' (1986a: 161). As he has put it elsewhere, an IBE-based defence of realism lacks any argumentative force since it employs 'the very type of argument whose cogency is the question under discussion' (1991: 82). Fine concludes that 'there is, in general, no rational defence of realism' (1986a: 163). But Fine has also put forward two more objections. Let us suppose, he says, for the sake of the argument, that abduction *is* reliable. It would not be wise for realists to use an abductive argument in their defence of realism, since they must demand more stringent methods of proof of their philosophical doctrines (cf. Fine 1986: 114). At any rate, he notes, there are better instrumentalist explanations of the success of science (Fine 1986a: 154).

In what follows I explore some new and systematic ways in which realists can attempt to block the foregoing objections.

EDR and circularity

To call an argument viciously circular is to level an epistemic charge which indicates that the argument in question cannot, and perhaps should not, be *persuasive* since it in some way assumes, or postulates, that which needs to be independently shown. A typically circular argument is one in which the conclusion is either identical to or a mere paraphrase of one of its premisses. Note, however, that the mere fact that a premiss is identical to the conclusion is not sufficient ground for attributing *vicious* circularity. To show that an argument is *viciously* circular one should not just look at the sentences employed in the argument, but also take account of what the argument presumes to show by its use of the specific sentences. So, for instance, if we look only at the sentence-structure involved in it, the argument-type '*a* & *b*, therefore *b* & *a*' is circular. But it is *not* viciously circular since, I take it, it purports to show only the commutativity of logical conjunction. Similarly, the argument-type '*p*, therefore *p*' should not be deemed viciously circular if it is meant to show that every sentence is a logical consequence of itself. But it would be viciously circular were it meant to show that *p* is true. For then it would pretend to *prove* that *p* is true where it just *assumes* that *p* is true.

What is necessary in order for an argument to be correctly judged *viciously* circular is that the argument should purport to offer reasons for accepting a certain sentence (the conclusion), where (one of) the reasons cited is that sentence itself. Following Braithwaite (1953), one may call viciously circular arguments 'premiss-circular'. In the latter, one claims to offer an argument for the truth of α, but explicitly *presupposes* α in one's premisses. Such an argument has no probative force for anyone who does not already accept that α is true.[4]

In his attempt to defend an inductive vindication of inductive learning from experience, Braithwaite (1953: 274–278) also noted that there is a type of circular argument which is not premiss-circular. On the surface level, the argument is as non-circular as anything can be. It begins with the premisses $P_1, \ldots, P_n$, and then, by employing an inference rule R, it draws a certain conclusion Q. However, Q has a certain logical property: it asserts or implies something about the rule of inference R *used* in the argument, in particular that R is reliable. Braithwaite called this argument-type 'rule-circular'. In general, rule-circular arguments are such that the argument itself is an instance of, or involves essentially an application of, the rule of inference vindicated by the conclusion.

Braithwaite took it that rule-circularity was not vicious. I think this is correct. There are a few relevant differences between premiss-circularity and rule-circularity. The conclusion of a rule-circular argument is *not* one of the premisses. Nor is the argument such that one of the *reasons* offered for the truth of the conclusion is the conclusion itself. Hence, to say the least, rule-circular arguments are not *obviously* viciously circular. The case of rule-circular arguments has been defended, in connection with induction, by Braithwaite (1953), van Cleve (1984) and Papineau (1993). But, first appearances aside, there is a *residual* suspicion that rule-circular arguments are vicious. Before I try to disperse this doubt, I want to show that NMA is, if anything, a rule-circular argument.

As we saw in the last section, the premisses of NMA assert the theory-ladenness of scientific methodology and its widely accepted instrumental and predictive success. Then, by means of a meta-IBE, the argument concludes that the background theories are approximately true. Since these approximately true theories have been typically arrived at by *first-order* IBEs, this information together with the conclusion of the meta-IBE entail that IBE is reliable. So, the truth of the conclusion of NMA is (part of) a sufficient condition for accepting that IBE is reliable. NMA is clearly *not* premiss-circular. The conclusion of the meta-IBE (that theories are approximately true) is not among the premisses of the argument. In fact, no assumption about the approximate truth of theories is made within the premisses, either explicitly or implicitly. Besides, there is no a priori guarantee, as clearly there would have been if this argument were premiss-circular, that the conclusion of NMA will necessarily be that theories are (approximately) true. The conclusion is true, if at all, on the basis that it is the best explanation of the premisses, but it might not have been the best explanation. As we shall see, this point is implicitly conceded by the critics of NMA, since they take pains to argue that there are better explanations of the success of science. By arguing that the conclusion of NMA need not be the intended realist conclusion, they acknowledge implicitly that NMA is not premiss-circular.

Let us now examine in some detail whether rule-circularity is, nonetheless, vicious. How could it be? The thought here might be that in a rule-circular argument one has to assume the reliability of the rule invoked in the argument. But if this assumption is based on the prior acceptance of the conclusion of the rule-circular argument, then the proponents of a rule-circular argument apparently traffic in a vicious circle. For they would have to prove the conclusion *before* they accepted the rule used to derive it. But they could not prove the conclusion unless they *first* accepted the reliability of the rule.

I want to reply to this objection by denying that any assumptions about the reliability of a rule are present, either explicitly or implicitly, when an instance of this rule is used. Nor should the reliability of the rule be established *before* one is able to use it in a justifiable way. This is controversial. But here I am in good company. Externalists in epistemology have argued for this extensively (see Goldman 1986). The point is the following. When an instance of a rule is offered as the link between a set of (true) premisses and a conclusion, what matters for the correctness of the conclusion is whether or not the rule *is* reliable that is, whether or not the contingent assumptions which are required to be in place in order for the rule to be reliable *are* in fact in place. If the rule of inference *is* reliable (this being an objective property of the rule) then, given true premisses, the conclusion will also be true (or, better, likely to be true – if the rule is ampliative).[5] Any assumptions that need to be made *about* the reliability of the rule of inference, be they implicit or explicit, do not matter for the correctness of the conclusion. Hence, their defence is not necessary for the *correctness* of the conclusion.

In order to highlight the point just made, let us envisage the following situation. Suppose that, in a fashion analogous to a Turing test, we come across a certain 'inference machine' and we start playing a game with it. We feed it with several sets of true premisses and ask it to draw conclusions from them. Suppose also that in all (or most) cases the 'inference machine' draws true conclusions. To say the least, we would conclude that the 'inference machine' is (or is likely to be) reliable. We would also think that the 'inference machine' must operate according to some rules of inference in such a way that when the premisses are fed in it activates a rule and draws a conclusion. But *qua* machine, the 'inference machine' makes no assumptions about the rules it activates. It just activates them. And, given the success of the 'inference machine' in drawing true conclusions, can we protest that we should first identify the rules it activates, prove that they are reliable, and only then accept that the 'inference machine' is reliable? I think this would be unreasonable and, in any case, counter-productive. If the 'inference machine' started producing consistently false conclusions, we would have reason to start worrying. But in their absence, worrying is unnecessary.[6]

Pursuing the previous example, one might object that the issue is more complicated if we think, as we should, of reasoners as 'conscious inference machines'. For, the objector might note, the defence of the reliability of the rule of inference *does* matter for the *justification* that the reasoner might have for taking the conclusion to be correct (or, likely to be correct). This is really the point on which the allegedly vicious nature of rule-circularity turns. For whether or not the proof of reliability is required for justification will most likely depend on the epistemological perspective which one adopts. As is well known, *externalist* accounts sever the alleged link between being justified in using a reliable rule of inference and knowing, or having reasons to believe, that this rule is reliable. On such accounts, if the rule is reliable, then it thereby confers justification on a conclusion drawn using this rule, insofar as the premisses are true. Hence, given externalism, all we should require of a rule-circular argument is that the rule of inference employed *be* reliable; no more and no less than in any ordinary (first-order) argument. A rule-circular argument would be no more vicious than any other first-order application of the rule involved in it. Since first-order applications are not vicious, nor is the second-order application involved in the rule-circular argument. What is special with rule-circular arguments is what the conclusion says. It asserts that the rule of inference is reliable. But the correctness of this conclusion depends on the rule being reliable, and not on having any reasons to think that the rule is reliable. No less than the conclusion of any first-order ampliative argument, the conclusion of a rule-circular argument will produce a belief, this time about the rule of inference itself. This belief will be justified if the rule is reliable. But, if we keep with externalism, it is the truth of this belief and the (objective) reliability of the rule which generated it that matter. Justification requires no more than *reliability* and *truth*.

Rival *internalist* accounts of justification suggest that justification requires something over and above the fact, if it is a fact, that the rule *is* reliable, viz. knowing (or justifiably believing) that the rule of inference involved is reliable. So, if one took an internalist approach, then a separate justification of the reliability of the rule would be required for the overall warrant the reasoner might have for taking a belief issued by the rule to be true. On this understanding of justification, rule-circular arguments might appear to be vicious. For it seems that believing the conclusion of the rule-circular argument would be necessary in order to justifiably use the rule involved in it the first place. Hence, internalists would be likely to require an *independent* justification of the rule – that is, a justification of the kind that a rule-circular argument cannot possibly offer.

So, the issue of whether rule-circular arguments are vicious turns on the theory of justification one adopts. Realists should have to be externalists if they take NMA seriously. And their critics will have to argue for internalism, if the charge of vicious circularity is to go through. Given an externalist perspective, NMA does not have to assume *anything* about the reliability of IBE. Consequently, it does not have to assume anything about the reliability of IBE that anyone else (the critics of realism, in particular) *denies*. To be sure, the proponents of NMA have to assume an externalist theory of justification that some critics of realism might deny. But that is a different matter. That battle can be fought on general epistemological grounds which have nothing to do with the issue of circularity.

The point just made may give rise to further objections. One such might be that, even if we grant externalism, NMA does rely on the assumption that IBE is reliable. For, if the NMA does not presuppose or assume this, why should it employ an IBE in its defence of realism? Why not rely on some *other* type of inference? And if NMA does rely on this assumption, realists surely need to defend it in an independent way, would they not? Another objection might be that, if externalism is assumed, why should realists bother to offer NMA in the first place? By offering this argument, do they not implicitly assume that we need reasons to believe in the reliability of IBE? That is, do they not grant what the internalists have argued for all along? Let us take these objections in turn. Providing the answer to the first is a straightforward matter; but the second objection will not be met without some more work.

Why should NMA rely on an IBE in its defence of realism? Does that not imply that it assumes IBE to be reliable? I do not think it does. If one knew that a rule of inference was unreliable, one would be foolish to use it. This does not imply that one should first be able to prove that the rule is reliable before one uses it. All that is required is that one should have no reason to doubt the reliability of the rule – that there is nothing currently available which can make one distrust the rule. The defenders of NMA are 'guilty' of something: we would not use IBE if we had reasons to consider it unreliable. But we have no such reason. There is nothing vicious in admitting all this. If someone denied that abduction is reliable, they should have to give some reasons why this is so. This debate can go on independently of the issue of circularity. It will turn on arguments which aim to show that IBE should not be trusted. (Such arguments will be dealt with in Chapter 9.) But an analogy, due to Frank Ramsey (1926 [1978: 100]), will bring the present point home. It is only via memory that we can examine the reliability of memory. Even if we were to carry out experiments to examine this, we would still rely on memory: we would have to remember the outcomes of the experiments. But there is nothing vicious in using memory to determine and enhance the degree of accuracy of memory. For there is no reason to doubt its overall reliability.

Let us now focus on the second objection above: by offering the NMA, are realists not implicitly offering *reasons* to believe in the reliability of IBE? And, if so, should not these be independent reasons? I have two points against this objection.

1 The objection misunderstands what the NMA aims to do. NMA does not *make* IBE reliable. Nor does it add anything to its reliability, if it happens to be reliable. It merely generates a new belief *about* the reliability of IBE which is justified just in case IBE is reliable.
2 But, suppose we granted that NMA aimed to defend the reliability of IBE. This is certainly not excluded by externalism. It is just optional. Would the mere fact that the defence relies on a rule-circular argument make the attempted defence vicious – and hence lacking in rational force? I do not think so. If the rule-circularity of a defence is taken to be an outright vice, then we should simply have to forgo any attempt to explain or defend any of our *basic* inferential practices. What this implies is that even internalist defences, ultimately, will have to rely on rule-circular arguments. When it comes to the defence of our basic modes of reasoning, both ampliative and deductive, it seems that we either have no reasonable defence to offer or else the attempted defence will be rule-circular.

This dilemma shows up already in the case of deductive inference. It goes back to Lewis Carroll and his 'What the tortoise said to Achilles' that one cannot prove the soundness of *modus ponens* unless one ultimately employs *modus ponens*. We need *modus ponens* (and other deductive rules) because we need truth-preserving rules of inference – rules such that, whenever the premisses of an argument are true, the conclusion is also true. But can we prove that *modus ponens* is truth-preserving? The best we can do is to prove a meta-theorem that *modus ponens* in the object-language is truth-preserving. This meta-proof, however, requires that the meta-language already has *modus ponens* (or other deductive rules) as a rule. Intuitively, the idea is that any kind of proof (even the proof that *modus ponens* is truth-preserving) requires some rule of inference in order for it to go through. In the case of *modus ponens*, the required rule must also be truth-preserving. But do we not need a proof that *this* rule is truth-preserving? And so on. A typical reply, expressed vividly by Salmon (1965: 54), is that we *should* trust *modus ponens* because we do not have any reason to doubt that it is truth-preserving: we can 'reflect' on instances of *modus ponens* and realise the inconceivability of the situation in which all of the premisses are true and the conclusion is false. Whether this is exactly right is still debatable. Van McGee (1985) and William Lycan (1994), for instance, have suggested that there are counter-examples to *modus ponens*. That is, there are instances of arguments which instantiate *modus ponens*, and yet have true premisses and a false conclusion.[7] I do not want to enter this interesting debate here, but the typical response to these counter-examples shows that the defence of the soundness of *modus ponens* is a far from trivial (and presuppositionless) exercise. The typical reply to these counter-examples, discussed by Kornblith (1994), is that if we just define *modus ponens* using the standard meaning of the logical connective for conditional statements of the form '*p q*', (where the conditional is true either when the antecedent is false or the consequent true), then there is *no* room for counter-examples: any purported counter-example is dismissed on the grounds that it should not be formalised as a purported instance of the schema $\{p;\ p \rightarrow q;$ therefore $q\}$. The issue here is not whether this dismissal is correct (Lycan 1994a, for instance, doubts that it is). Rather, the issue is that no justification of *modus ponens* is possible which does not rest on some presuppositions. All we can do is engage in a process of *explanation* and *defence*. By reflecting on *modus ponens* (and other deductive rules we use), we aim to systematise it, to explain to ourselves the ways in which we should use it, and to show that, *given* the meaning of the logical connectives and the truth-tables, it delivers its goods – it is truth-preserving.[8]

A similar, if more complicated, situation arises when it comes to inductive reasoning. Inductive rules are non-truth-preserving. However, it is wrong to apply deductive standards to inductive reasoning. While deduction is concerned with truth preservation, induction is concerned with learning from experience. The fact that induction is not deduction shows nothing other than that each should be treated as a distinct mode of reasoning. But how can the very possibility of rational learning from experience be defended, if not by a rule-circular argument? Carnap's work can help us address this issue in a systematic way. Carnap's major problem was to establish which kinds of inductive argument in his systems of inductive logic are valid, in the sense that they license conclusions with high inductive probability (or degree of confirmation). In particular, he wanted to find out which among a number of ampliative rules (straight rule, Laplace's rule, c^*, $c^\dagger$, etc.) can best represent inductive learning from experience. But, we all know that one cannot defend the validity of inductive arguments without using *some* form of inductive reasoning. Reflecting on this question, Carnap (1968: 265–267) suggested that the circularity involved in an attempt to vindicate inductive reasoning is both indispensable and harmless. Here is a reconstruction of his argument.

> Reasoners are either *inductively blind* – where 'inductively blind' refers to reasoners who make no inductive inferences and who are not disposed to make any – or they are not. If the reasoners are inductively blind, then we cannot possibly show them when an argument is inductively valid and when it is not. For learning to discriminate between these two cases, and therefore learning to recognise inductively valid arguments and to discard invalid ones, requires an 'inductive intuition'. This intuition should not be confused with the Cartesian idea of an infallible source of knowledge. Rather it should be seen as some sort of *disposition* to use inductive reasoning and to fallibly recognise that an argument is inductively valid. If there were such (unfortunate) inductively blind persons, they would be inductively blind precisely because they lack this disposition to learn from experience. When it comes to *our* attempts to persuade them why learning from experience is reasonable, we can rely only on some inductive argument – we have to rely on the past successes of inductive reasoning. What we are doing is indispensable, because no other argument could show them that learning from experience is reasonable. Yet being engaged in rule-circular reasoning also is harmless because, being inductively blind – nothing could persuade our interlocutors to reason inductively. If, on the other hand, the reasoners are not inductively blind – if they already operate within a network of dispositions to learn from experience – it is also both indispensable and harmless to engage in rule-circular reasoning in an attempt to explain to them the circumstances under which an inductive argument is or is not valid. It is indispensable because no non-inductive argument is available, and it is harmless because, in this case, it is an instance of a self-clarificatory procedure.
>
> So, in either case, in our attempt to vindicate learning from experience, being engaged in rule-circular reasoning is both indispensable and harmless. The situation is totally analogous to the defence of deductive reasoning. There is no way in which one can persuade a *deductively blind* person of the soundness or rationality of deductive arguments. However, all those who operate in a network of deductive intuitions – e.g. who have internalised the meaning of the logical connectives, etc. – can be made to discriminate between valid and invalid arguments.

Carnap's argument suggests a wholly new perspective on the issue of what exactly we do when we offer arguments in defence of our basic inferential practices. In one sense, no inferential rule carries an absolute rational compulsion, unless it rests on a framework of intuitions and dispositions which takes for granted the presuppositions of this rule (truth preservation in the case of deductive reasoning, learning from experience in the case of inductive reasoning, searching for explanations in the case of abductive reasoning). When we attempt to vindicate or defend certain rules of inference (e.g. certain deductive, inductive or abductive rules), this is not because we want either to justify them without any assumptions, or to prove that they are rationally compelling for any sentient being. It is because we want to evaluate our existing inferential practices: to reflect on the rules we use or are disposed to use uncritically, and to examine the extent to which and in virtue of what these rules are reliable. Such evaluations cannot be made from a neutral epistemological standpoint. They, too, have to employ some methods. In the final analysis, we just have to rely on some basic methods of inquiry. The fact that we make recourse to rule-circular arguments in order to defend them, if defence is necessary, is both inescapable and harmless.[9]

By parity of reasoning, if one is disposed to reason abductively one should have no special problem with using NMA in defence of the reliability of IBE. NMA is no worse than attempts to defend *modus ponens* and inductive rules. In fact, the class of reasoners who use abductive reasoning is much broader than the class of committed realist epistemologists who reflect on the reliability of IBE and defend it by offering the NMA. This class will most certainly include non-realists – those who do not take sides on the realism debate. But it will also include those critics of realism who employ abduction, but disagree with the conclusion of NMA, the thesis that scientific theories are approximately true. As I noted above, that this class is not empty follows from the fact that at least some critics of the realist NMA try to show that there are better potential explanations of the success of science than the realist one. If sound, NMA can have rational force for all of them.

So, NMA has not been shown to be viciously circular. That being so, I do not know what the problem with NMA is. In any case, Fine (1986: 115) is mistaken in maintaining that NMA is 'of no significance'.

Fine has, however, launched another criticism against EDR, what he calls 'a deep and . . . insurmountable problem with the entire strategy of defending realism' (1986: 114). He grants, for the sake of the argument, that EDR may be successful in convincing someone who already employs abductive reasoning about the truth of realism. Then he asks: 'should that not be of some solace, at least for the realist?' (ibid.: 117).

Fine thinks that EDR should give no comfort to realists. For one must demand that the proofs of one's meta-theories be more stringent than the proofs in one's theories. To this end Fine appeals to Hilbert's programme of showing the consistency of mathematical theories by using only the most stringent and secure means – in particular, means which fall outside the proof-theoretic tools of the theory under consideration. Fine argues:

> Hilbert's idea was, I think, correct even though it proved to be unworkable. Metatheoretic arguments must satisfy more stringent requirements

> than those placed on the arguments used by the theory in question, for otherwise the significance of reasoning about the theory is simply moot. I think this maxim applies with particular force to the discussion of realism.
>
> (1986: 114)

From a naturalist viewpoint, it is of great relevance to the debate if a requirement has proved to be utopian. It is plain from Goedel's second incompleteness theorem that there cannot be a stringent proof, in Hilbert's sense, of the consistency of Peano arithmetic. In particular, any consistency proof for such an axiomatic formal theory is – at least in some sense – less elementary than the formal methods which the axiomatic theory formalises. Hilbert's requirement might be in principle correct. Yet, it is unreasonable to demand that a philosophical theory must satisfy a requirement that mathematics, with an accurate notion of proof and a strict and rigorous deductive structure, fails to satisfy. Fine's demand (1986: 115) that a realist theory of science employ 'methods more stringent than those in ordinary scientific practice' is unnaturally strong and unnaturally non-naturalistic.

Are there better explanations of the success of science?

What needs to be shown also is that NMA's conclusion is indeed the best explanation of the instrumental success of science. This is crucial because otherwise NMA cannot adequately defend the reliability of abduction; moreover Fine has argued that there is a better non-realist explanation of the success of science. In fact, Fine (1986a: 154) defends the rather bold thesis that anything which realists can do instrumentalists can do, and in a better way.

Fine's claim is that some notion of *instrumental reliability* of scientific theories best explains the success of science, where 'instrumental reliability' is a feature of scientific theories in virtue of which they are 'useful in getting things to work for the practical and theoretical purposes for which we might put them to use' (1991: 86). However, Fine's strategy faces a general problem. Suppose that he uses IBE in order to infer the *truth* of instrumentalism. Then he seems to admit that abduction *is* reliable, yet it just happens that, contrary to what realists expect, realism is not the best explanation of the success of science: rather, instrumentalism is. But then Fine would have to concede that abduction is reliable.

So Fine's use of IBE must be different. It should not, that is, be seen as an inference to the *truth* of the best explanation – the latter being, according to him, that science is instrumentally reliable. In fact, Fine has spoken of 'an instrumentalist version of the inference to the "best" explanation' (1991: 83). This version should still favour the best explanation, but it should assert that the best explanation is *empirically adequate* rather than true. Instrumentalism would get accepted as empirically adequate, *à la* van Fraassen. Yet there would still, I think, be a problem. For even if instrumentalism were shown to be the best explanation of the instrumental success of science, it could not be more empirically adequate than realism. Realism and instrumentalism are equally empirically adequate. They both entail the empirical success of science. And note that for most instrumentalists empirical adequacy is the only epistemic virtue of a potential explanation – the only feature that contributes to its belief-worthiness *qua* explanation. If Fine accepted this common instrumentalist tenet, then even if instrumentalism were a better explanation of the success of science, it would be no more belief-worthy than realism, since they would be equally empirically adequate. If, however, Fine thought that certain explanatory virtues could, alongside their empirical adequacy, make one explanation more belief-worthy than another, then he would move away from an instrumentalist version of IBE and would defend instrumentalism only at the cost of conceding a major point to realism, viz. that explanatory virtues are ultimately epistemic virtues.

Let me, however, leave aside these qualms and focus on the central question: is the instrumentalist explanation of the success of science better than the realist one? Fine (1986a: 153–154; 1991: 82–83) contrasts two forms of (simplified) abductive explanations of the success of science:

(a)	(b)
Science is empirically successful	Science is empirically successful
∴ (probably) theories are instrumentally reliable	∴ (probably) theories are approximately true

Fine suggests that pattern (a) is always preferable to (b) on the grounds that if the explanandum is the empirical success of scientific methodology, then we do not need to inflate the potential explanation with 'features beyond what is useful for explaining the output' (1991: 83). So Fine thinks 'the instrumentalist, feeling rather on home ground, may suggest that to explain the instrumental success we need only suppose that our hypotheses and theories are instrumentally reliable' (1991: 82–83).

I think Fine's argument rests on the hidden assumption that an appeal to the (approximate) truth of background scientific theories goes beyond the features that are useful for explaining the instrumental success of science. In his essay 'Unnatural Attitudes' (1986a: 153), he has in fact suggested that admitting anything more than instrumental reliability 'would be doing no explanatory work'. His argument goes like this. When realists attempt to explain the success of a particular theory, they appeal to the approximate truth of a theoretical story as the best explanation of the theory's success in performing certain empirical tasks. But if this explanation is any good at all, they must 'allow some intermediate connection between the truth of the theory and success in its practice. The intermediary here is precisely the pragmatist's reliability' (1986a: 154). So, Fine suggests, the job that truth allegedly does in the explanation of the success of a theory is actually done by this intermediate *pragmatic* reliability. Truth seems explanatorily redundant. Moreover, if pragmatic reliability is substituted for truth in the realist account of success, one gets an alternative account in terms of instrumental reliability (ibid.: 154). Fine concludes: 'since no further work is done by ascending from that intermediary to the realist's "truth", the instrumental explanation has to be counted as better than the realist one. In this way the realist argument leads to instrumentalism' (ibid.). On the basis of this argument, Fine proves a meta-theorem: 'If the phenomena to be explained are not realist-laden, then to every good realist explanation there corresponds a better instrumentalist one' (ibid.).

There are two strange aspects to Fine's argument.

1 It is not at all obvious that there is anything like a *pragmatic* notion of reliability which realists have to take into account in their explanation of the success of science. Between successful empirical predictions and theories there are methods, auxiliary assumptions, approximations, idealisations, models and probably other things. Let us suppose that this stuff is what Fine calls the 'pragmatic intermediary'. Let us also suppose that these things alone could be summoned to account for the empirical success of a theory. Would this make claims concerning the truth of the theory explanatorily superfluous? Surely not. For one also wants to know why some particular model represents successfully the target physical system whereas others do not, or why one model represents the target physical system better than others, or why the methods followed generate successful predictions, or why some idealisations are better than others, and the like. When realists argue for the approximate truth of background scientific theories, they, in effect, want to explain the success (or instrumental reliability) of this intermediary stuff. Approximate truth would be summoned in order to explain the successful constraints which theories place on model-construction as well as the features of scientific methods in virtue of which they produce successful results. So, if Fine meant this stuff when he spoke of a pragmatic intermediary between the (approximate) truth of theory and its success in practice, the existence of *this* pragmatic intermediary would not render approximate truth explanatorily superfluous.

2 Even if we assume that there is some other *pragmatic* notion of reliability to be interpolated between approximate truth and empirical success, and even if we equate this notion with Fine's instrumental reliability, that it has any real explanatory import would be open to doubt. Instrumental reliability is nothing but a summary statement of the fact that the theory performs successfully practical tasks. If we then try to explain the theory's empirical success by saying that background theories are instrumentally reliable, we simply paraphrase what needs to be explained. It is immaterial whether we phrase the explanandum as 'Theories are successful' or as 'Theories are instrumentally reliable'. No explanation is thereby offered, only a paraphrase of theories' success in terms of theories' instrumental reliability. The situation here is totally analogous with an attempt to 'explain' the fact that hammers are successful in driving nails into a wall by saying that hammers are instrumentally reliable for nail-driving. Recall that what is at stake is whether an instrumentalist explanation is better than the realist one. It turns out that, despite all the manoeuvring, it is not an explanation at all.

Fine has implicitly recognised that instrumental reliability is a rather poor explanation. For he has recently (1991) suggested a way to make claims of instrumental reliability potentially explanatory. He has outlined a dispositional understanding of the instrumental reliability of science. On this view, instrumental reliability involves a *disposition* to produce correct empirical results. Fine claims that this dispositional explanation of the success of science is 'an explanation of outcomes by reference to inputs that have the capacity (or "power") to produce such [i.e. instrumentally reliable] outcomes' (1991: 83).

This new understanding of instrumental reliability *is* potentially explanatory: it accounts for empirical success by an appeal to a *capacity*, or disposition, that theories have in virtue of which they are empirically successful. Although certainly in the right direction, this account is incomplete. Not because there are no dispositions, or powers, in nature, but rather because one would expect also an explanation of why and how theories have such a disposition to be instrumentally reliable; in particular an explanation that avoids the troubles of Moliére's 'explanation' of why opium sends somebody to sleep in terms of its 'dormitive power'. Is it a brute fact of nature that theories – being paradigmatic human constructions – have the disposition to be instrumentally reliable? This seems hardly credible. If dispositions of this sort need grounding, then there is an obvious candidate: the property of being approximately true would ground the power of scientific theories to be instrumentally reliable. Since Fine would certainly deny this account, he owes us an alternative story of how this disposition is grounded. Else, should this disposition need no grounding, he needs to show how can this be so.

I conclude, then, that Fine has failed to prove his meta-theorem in favour of instrumentalism. The realist account is the best overall explanation of the empirical success of science.

Could we not just deflate our quest for explanation?

There is an aspect of the *intuitive* epistemic thrust of Fine's critique of realism with which I have not yet dealt: that somehow 'going beyond the data' to posit 'theoretical entities' is more problematic than abandoning some forms of intuitively attractive abductive reasoning. A defender of Fine's critique of realism in particular might suggest that a deflationary account of explanation as licensing retrodiction and prediction might do just as well, without taking extra risks about theoretical commitments. Here is how. Suppose that someone accepts the foregoing distinctions between premiss-circularity and rule-circularity as well as the existence of abductive, or explanatory, intuitions. He might, therefore, acknowledge the *prima facie* force of the demand for an explanation of the reliability of scientific methodology. But instead of accepting the realist's explanation, he identifies explanation with retrodiction and prediction, and offers the following (Quinean) second-order induction about abduction as an epistemic justification of abductive practices in science: past abductive inferences have generated empirically successful theories; hence, based on a second-order *induction*, it is reasonable to expect that abductive inferences will keep providing empirically successful theories. So he concludes that one can be equipped with inductive generalisations about the instrumental reliability of abductive scientific methodology on the basis of which one can predict or retrodict the instrumental reliability of scientific methodology in particular cases. But, he stresses, these inductive generalisations do not commit one to the existence of unobservable entities, nor do they entail that abductive reasoning is a reliable guide to theoretical truth. All that they entail is that one can rely on abductive reasoning to get instrumentally reliable theories, but nothing more. I shall call this 'the induction-about-abduction' move.

I think this move is in the spirit of Fine's dispositional account of instrumental reliability discussed at the end of the previous section. In fact, the suggested inductive generalisations about the instrumental success of scientific methodology might be offered as a way to ground claims about the disposition of this methodology as instrumentally reliable. Two responses, which work in tandem, are available. First, that these generalisations do not really explain why scientific methodology is reliable; and, second that these generalisations are not free of theoretical commitments. Let us take them in turn.

Take the (second-order) generalisation that abductive reasoning generates instrumentally reliable theories. Let us call it A. A can be paraphrased as the conjunction of the following two claims:

A_1: abductive reasoning has generated instrumentally reliable theories in the past and present; and
A_2: abductive reasoning will generate instrumentally reliable theories in the future.

Now remember what needs to be explained: the instrumental reliability – past, present and future – of scientific theories. It is, then, not difficult to see that A_1 & A_2 is merely a paraphrase of what needs to be explained. More specifically, we can question whether this generalisation, as it stands, is suitable for prediction and retrodiction. If we use A (= A_1 & A_2) to *predict* a future instance of instrumental reliability, we need to assume that A (= A_1 & A_2) is already well-confirmed, which means that we need to assume what is really at issue: that A_1 on its own provides good inductive evidence for A_2. What exactly makes it the case that A_1 supports A_2? It may well be the case that hitherto instrumentally reliable theories fail when they are extended in new domains; unless, of course, we assume that they are truth-like. This appeal to truth-likeness would explain why theories are (or tend to be) instrumentally reliable, and would also warrant the projection to future instrumental reliability. On the other hand, if we use A (= A_1 & A_2) to *retrodict* the past instrumental reliability of scientific theories (A_1), we will have to appeal, implicitly, to their future reliability (A_2), a fact as much in need of explanation and grounding as is A_1. In any case, positing the approximate truth of scientific theories would offer a more satisfactory and highly non-trivial way to predict and retrodict their instrumental reliability: it is in virtue of theories being approximately true that we can

- retrodict their instrumental success in certain cases;
- predict future successes; and
- confirm the generalisation that abductive reasoning generates empirically successful theories.

This last claim would be in accord with the confirmation of empirical generalisations in scientific practice. Empirical generalisations are considered well confirmed mainly when they are embedded in larger theoretical structures which explain how the properties involved in the generalisation co-vary and how the generalisation gets connected with other well-supported ones. A framework which is (approximate) truth-linked plays precisely this role when it comes to the explanation of the instrumental reliability of scientific methodology and the instrumental successes of scientific theories.

At any rate, it is highly dubious that the 'induction-about-abduction' move can altogether avoid theoretical commitments. Boyd has in fact considered a similar objection to his attempt to defend the reliability of abductive reasoning (cf. 1984: 68–70; 1985: 236–241). The point is straightforward. Prior to performing the induction on past empirically successful scientific theories, we must naturally accept that instrumental success constitutes evidence for the truth of the inductive generalisations about observables made by these theories. But this judgement is not independent of all theoretical commitments. From myriad generalisations that involve observables, scientists pick only some as genuinely empirically supported and confirmed. Their choice is theory-dependent: theories suggest connections between hitherto unrelated observable phenomena; they determine which predicates are projectible, and which collections of individuals form natural kinds. But if ordinary judgements concerning inductive generalisations about observables involve theoretical commitments, any attempt to have an induction-about-abduction that is free of theoretical commitments will be seriously impaired.[10]

Can Darwin help?

Van Fraassen has offered a different explanation of the success of science. It is this:

> The success of science is not a miracle. It is not even surprising to the scientific (Darwinist) mind. For any scientific theory is born into a life of fierce competition, a jungle red in tooth and claw. Only the successful theories survive – the ones which *in fact* have latched on to actual regularities in nature.
>
> (1980: 40)

On this account, there is no surprise in the fact that current theories are empirically successful. For the Darwinian principle of the survival of the fittest has operated. Current theories have survived because they were the fittest among their competitors – fittest in the sense of latching on to universal regularities. Clearly, this is an elegant and simple explanation of the fact that current theories are successful. But does it undermine the realist explanation?

If we unpack van Fraassen's story, we find that it is *phenotypical*: it provides an implicit selection mechanism according to which entities with the same phenotype, i.e. empirical success, have been selected. But a phenotypic explanation does not exclude a *genotypic* account: an explanation in terms of some underlying feature which all successful theories share in common; a feature which has made them successful in the first place. The realist explanation in terms of truth provides this sort of genotypic account: every theory which possesses a specific phenotype, i.e. it is empirically successful, also possesses a specific genotype, i.e. approximate truth, which accounts for this phenotype. In order to see the point more clearly, compare van Fraassen's story with this (due to Peter Lipton): Each in a group of people has red hair. This is no surprise; but is explained by the fact that this group is comprised of members of the club of red-haired persons. (The club is, in a sense, a mechanism which selects only persons with red hair.) But this observation does not explain why George (or, for that matter, anyone of them taken individually) has red hair. A different, most likely genetic, story should be told about George's colour of hair.

Notice here that the realist explanation is *compatible* with van Fraassen's Darwinian account. Yet, the realist's is arguably preferable, because it is deeper. It does not stay on the surface – that is, it does not just posit a selection mechanism which lets through only empirically successful theories. It rather tells a story about the deeper common traits in virtue of which the selected theories are empirically successful.

As Lipton (1991: 170ff.) has suggested, there is another reason for preferring the genotypic to the Darwinian explanation: all that the phenotypic explanation warrants is that theories which have survived through the selection mechanism have not yet been *refuted*. There is no warrant that they will be successful in the future. Any such warrant must be external to the phenotypic story. For instance, this warrant can come from a combination of the phenotypic explanation with the principle of induction. On the other hand, the genotypic explanation has this warrant up its sleeve: if a theory is empirically successful because it is true, then it will keep on being empirically successful.

To sum up, then, there are no better explanations of the success of science than the realist one. Not that the discussion so far has exhausted all arguments levelled against IBE and its role in the realism debate. More is to come on this in Chapter 9, when I discuss van Fraassen's position. Additionally, there is a seemingly powerful argument against NMA which needs to be rebutted. It is the so-called 'pessimistic induction', enunciated by Laudan. Its thrust is that NMA cannot possibly be taken seriously because it flies in the face of the (alleged) fact that the history of science is the graveyard of supposed 'best explanations' of the evidence.

Part II will be devoted to defending realism against the pessimistic induction (Chapters 5 and 6), after which an attempt will be made to rebut the argument from the 'underdetermination of theories by evidence'.