

# Empirical Equivalence, Explanatory Force, and the Inference to the Best Theory

Igor Douven\*

Department of Philosophy, Erasmus University Rotterdam  
douven@fwb.eur.nl

## Abstract

In this paper I discuss the rule of inference proposed by Kuipers under the name of Inference to the Best Theory. In particular, I argue that the rule needs to be strengthened if it is to serve realist purposes. I further describe a method for testing, and perhaps eventually justifying, a suitably strengthened version of it.

In his impressive work *From Instrumentalism to Constructive Realism* (Kuipers 2000; hereafter referred to as ICR) Theo Kuipers proposes a rule of inference under the name of Inference to the Best Theory (IBT) that is meant to be an ameliorated version of the Inference to the Best Explanation (IBE), a rule generally believed to be of crucial importance to the case for scientific realism. The present paper argues that, even though it does indeed greatly improve on IBE and eludes what many regard to be a fatal or near-fatal objection to the latter rule, IBT is, as it stands, too weak to serve realist purposes. However, it will also be seen that the rule can be strengthened so as to make it adequate to its purported task. The paper further considers the question whether there is any reason to trust the conclusions reached by means of IBT. It is argued that such reasons may well have to come from a test originally proposed in Douven (2002a) and summarized and subsequently further elaborated in the present paper. I start, however, by briefly discussing the argument that is at present the main source of antirealist sentiments, paying special attention to the thesis of empirical equivalence, which serves as one of the argument's premises and which, in my view, Kuipers dismisses too quickly. This will help to elucidate the role such rules as IBE and IBT play in the realism debate, and, later on, why from a realist perspective somewhat stricter criteria for the goodness of theories are required than the ones that accompany IBT in Kuipers' presentation.

**1. Empirical Equivalence and Underdetermination.** According to scientific antirealists, theory choice is radically underdetermined by the data. They conclude from this that we can never be in a position to warrantably attribute truth to theories beyond their observational consequences. The common antirealist argument for the underdetermination thesis starts from the premise that for each scientific theory there are empirically equivalent rivals, i.e., contraries that have the same observational consequences that it has (call this claim EE). If EE is correct, then no matter how many empirical tests a theory has already passed, this success cannot be taken as an indication that the theory is true, for each of its empirically equivalent rivals will or would pass the same

---

\*I am greatly indebted to Roberto Festa, Theo Kuipers, and Jan-Willem Romeyn for valuable comments on an earlier version of this paper.

tests just as successfully. Thus, unless the data refute a theory, no amount of them suffices to determine its truth-value. If we then further assume that if the data alone do not suffice to determine a theory's truth-value, then nothing does (call this Knowledge Empiricism, or KE for short), as antirealists typically do, it follows that the truth-value of any theory having non-observational consequences must forever remain beyond our ken.

Since the argument is deductively valid, scientific realists will have to rebut at least one of its premises. Traditionally, realists have believed the fault is to be sought in the second premise, KE. The present paper also is mainly concerned with the latter premise; for, as will become clear, if KE is correct, then rules such as IBE and IBT cannot be correct. In this section, however, I concentrate on the first premise, EE, and in particular on what Kuipers has to say about it.

For a long time, both realists and antirealists have taken the truth of EE for granted. However, in the past decade or so, philosophers have become more skeptical about it. It is no exaggeration to say that this change in attitude is mainly due to work done by Laudan and Leplin.<sup>1</sup> Kuipers seems to share the new skepticism regarding EE. In the context of a discussion of some intuitive arguments for the referentiality of theoretical terms, Kuipers notes that "it is difficult to make such intuitive arguments formally convincing" (ICR, 227). He then goes on as follows:

However, there is no more reason to be pessimistic in this respect than about the presupposition of the referential skeptic, according to whom it will always be possible to invent 'empirically equivalent theories' ... which ... can explain the same variety of successes and success differences. It should be conceded that inventing such theories can not be excluded. In the same way, the skeptic can always tell a story that explains the seemingly veridical nature of our experimental procedures, without them really being veridical. We have to admit, however, that (natural) scientists, after a certain period of hesitation, make the inductive jump to the conclusion that a certain term refers ... (*ibid.*)

As I understand this passage, Kuipers' point is that, although it may always be possible to come up with empirically equivalent rivals for any scientific theory, this possibility is typically not taken very seriously by scientists, and thus, I assume, we are to conclude that we (philosophers) should not take it very seriously either. Another way to put the same point may be to say that, just as Cartesian skepticism at most succeeds in raising philosophical, but definitely not real, doubts about the possibility of our gaining knowledge in general, so antirealism at most succeeds in raising philosophical doubts about the possibility of our gaining scientific knowledge regarding the unobservable.

Let me first remark that at least the only well-developed antirealism to date, to wit van Fraassen's constructive empiricism, is not presented with the intention of simply rerunning the debate on skepticism within the philosophy of science. Skeptics are positively rare. For most (and perhaps even all) of us, skepticism is not a live option—we could not even *come* to hold it. We accept as a fact that we know quite a bit, and we regard any theory of knowledge that does not imply such as fundamentally defective. Nevertheless, we are sometimes willing to entertain skepticism; playing the skeptic's role can be a useful strategy for finding out whether a theory of knowledge is indeed defective for the reason just mentioned. But it will be clear to anyone familiar with van Fraassen's writings that his antirealism is not offered for such purely methodological reasons. Though van Fraassen agrees that we know quite a bit (cf. in particular 1989:178), according to him this bit is, and cannot but be, restricted to the observable

---

<sup>1</sup>See, e.g., Laudan (1990), Laudan and Leplin (1991), and Leplin (1997).

part of the world. Antirealism thus definitely is a live option for him. More than that, he actually urges us to be (or become) antirealists.<sup>2</sup>

Leaving van Fraassen's intentions to one side, I also think Kuipers' rather dismissive remarks on EE, as well as the current skepticism about this thesis among many other authors, are not altogether well-founded. Even if the arguments the antirealist can advance in support of EE are perhaps not quite as convincing as she might wish (and as they were once believed to be), it seems to me that, especially if taken in conjunction, they are convincing enough to sustain a case for antirealism as a real contender (as opposed to a mere skeptical or methodological alternative) for scientific realism.

First, antirealists can point to some actual examples of empirically equivalent rivals. Special Relativity and the æther theory in the Lorentz/Fitzgerald/Poincaré version are demonstrably empirically equivalent, as are standard quantum mechanics and Bohm's mechanics. Admittedly, there are not many more such examples. But the antirealist seems perfectly able to explain why there are so few. As Fine (1984:89), for instance, notes, in scientific practice it is typically quite hard to come up with even one theory that fits the data, let alone a number of such theories. By way of further explanation, we might add that it will in general not be a scientist's first ambition to find empirically equivalent rivals for extant theories (if only because success in this respect is very unlikely to be rewarded with a Nobel prize).

Secondly, there exist several proofs of EE; see Earman (1993), Douven (1995), and Douven and Horsten (1998). One might worry that the empirically equivalent rivals that these authors prove to exist for any scientific theory postulating unobservables are not genuine theories, but merely formal equivalents of the skeptic's Evil Demon or Brain in a Vat scenarios (cf. Laudan and Leplin 1991). However, although none of the proofs is constructive, they do give sufficient insight into the nature of the empirically equivalent rivals to justify the claim that they are not of that variety, or at least not all of them. And while it must be acknowledged that each of the currently available proofs of EE rests on non-trivial assumptions, these assumptions seem plausible enough for the proofs to lend considerable support to the thesis (even if the assumptions are not so obviously correct that the proofs can count as incontrovertible demonstrations of EE). And that seems to be all the antirealist needs. After all, we do not require the realist to *demonstrate* the correctness of her position, and so should not require this from her opponent. At any rate, in the face of these proofs it is at best misleading to assert, as Kuipers does, that it cannot be excluded that empirically equivalent rivals can be 'invented'.

In brief, I think it is fair to say that, although EE cannot be considered as being established once and for all, antirealists have, pending realist arguments to the contrary, bona fide grounds for holding that the existence of empirically equivalent rivals is to be seriously reckoned with (even if Kuipers is right that scientists typically do not do so).<sup>3</sup>

---

<sup>2</sup>At least this is what he does in his (1980). Later publications (in particular his 1985 and 1989) are more guarded on this point; some passages in these works suggest that van Fraassen has come to conceive of both scientific realism and scientific antirealism as (equally?) rational positions vis-à-vis science. See Kukla (1995, 1998) on the development of van Fraassen's thoughts on the (ir)rationality of scientific realism.

<sup>3</sup>One might suggest that there is a quick and easy way for the antirealist to end the recent skirmishes over EE, viz., by supplanting in the argument from underdetermination EE—according to which every scientific theory *has* empirically equivalent rivals—by the weaker premise that, *for all we know*, every scientific theory has empirically equivalent rivals. Logically speaking, this would suffice to make the argument from underdetermination go through. However, it should be noted that not any claim that will make the antirealist argument go through will also suffice to make the argument a real threat to scientific realism. And indeed, the mere possibility that a scientific theory has empirically equivalent rivals is hardly sufficient to *seriously* challenge the realist claim that the theory constitutes (or may constitute) knowledge.

For the realist much hangs on whether she has an adequate response to the second premise of the argument from underdetermination, KE, according to which only the data can determine a theory's truth-value, if that can be determined at all. If this thesis is right, then it does indeed follow, given EE, that knowledge of the unobservable is unachievable. However, realists have objected that by endorsing KE, antirealists totally neglect the role played by explanatory considerations in theory validation. Scientists assess a theory not just on the basis of its conformity with the data; they also take into consideration the theory's explanatory force. And such considerations, realists claim, are truth conducive and not of merely pragmatic significance, as antirealists typically hold. If this realist claim is right, as also most scientists seem to believe, then of course it does not hold that we cannot possibly come to know the truth-value of a theory which makes claims that go beyond the observable. For although empirically equivalent theories necessarily conform (or fail to conform) to the same data, it may well be that one of them provides a better explanation of those data than the other(s). Under the current supposition, this would give reason to believe it is true.

The problems connected with this realist response to the argument from underdetermination are manifold. Chief among them is the fact that realists have so far been unable to answer the antirealist challenge to make plausible that explanation is a mark of truth (the mere fact that most scientists take it as such is, the current enthusiasm for naturalism notwithstanding, philosophically not a sufficiently good reason to believe it is). In section 4 I describe a strategy that may well provide the means to meet that challenge. However, I first want to consider a quite ingenious argument for scientific realism devised by Dorling (1992),<sup>4</sup> and argue that it fails. This intermezzo has a double motivation. First, if it were correct, Dorling's argument would show that any appeal to explanatoriness is dispensable in a defense of scientific realism, and thus that the whole project of justifying the confirmation-theoretical role realists assign to explanation (and thereby section 4 of this paper) is otiose. Second, the discussion of the argument allows me to introduce in a natural way van Fraassen's ideas about how antirealism is to be modelled in confirmation-theoretic terms. In section 4.4 I try to show how we can be maximally concessive to van Fraassen by presupposing his preferred confirmation theory and yet be in a position to find empirical support for (a version of) IBT.

**2. A Bayesian Defense of Scientific Realism.** Dorling's (1992) argument focuses on local realism/antirealism disputes, i.e., disputes concerning the proper epistemic attitude towards a particular scientific theory (though as we shall see, it can, if successful at all, also be regarded as offering a defense of scientific realism *tout court*). His suggestion is that such disputes can be settled by simple Bayesian conditionalization on the available evidence. He tries to demonstrate this with the aid of an example. In his example, ' $T_R$ ' denotes some particular scientific theory not solely about observables, and ' $T_P$ ' denotes the set of observable consequences of  $T_R$ . Dorling then considers two philosophers (scientists?), one of whom is supposed to be a realist, the other an antirealist.<sup>5</sup> Their distinct attitudes are supposed to be reflected in the differences in the initial probabilities they assign to  $T_R$ : The realist assigns it a probability of .6, the antirealist a probability of .2. Since  $T_P$  is implied by  $T_R$ , both assign a conditional probability to the former given the latter of 1. The conditional probability of  $T_P$  given the negation of  $T_R$  is less straightforward, but Dorling assumes that the realist and antirealist agree

<sup>4</sup>The argument is also discussed in (ICR, 223f).

<sup>5</sup>An assumption left implicit in Dorling's example is that both are Bayesian learners, i.e., both calculate new probabilities by means of Bayes' theorem. This assumption is far from innocent, of course, but we will go along with it here.

that it is .2. We are now to suppose that we obtain enough evidence for  $T_P$ 's correctness to make us virtually certain of it. What would that mean for the realist's and antirealist's confidence in  $T_R$ , respectively? Two easy calculations show that the realist now believes  $T_R$  to a degree of (approximately) .9 and that the antirealist believes it to a degree of (approximately) .6.<sup>6</sup> A first, rather unsurprising, observation Dorling makes about this result is that both the realist and the antirealist are now more confident in  $T_R$  than they were before they received the evidence for  $T_P$ . More surprising may be the observation that the antirealist's new confidence in  $T_R$  has increased to such an extent that she now has more confidence in  $T_R$  than in  $\neg T_R$ . She might thus be said to have been converted to realism with regard to  $T_R$  (Dorling 1992:368f).

It seems that in this case the realism issue has been settled in favor of the realist, and—most relevant to our concerns—that this has been accomplished without any appeal to the explanatory power of the theory under consideration. The example is rather abstract, but according to Dorling many local realism debates in the history of science fit the example in all relevant respects. Something that Dorling does not discuss but that is certainly noteworthy is that, if his argument is sound, it may even offer a defense of scientific realism in general. For it suffices that *some* local realist disputes have been, or can be, settled in favor of the realist in order to establish modern, relatively modest versions of realism such as Leplin's minimal epistemic realism, according to which "there are possible empirical conditions that would warrant attributing some measure of truth to theories" (Leplin 1997:102). And although the same would not suffice as a defense of more ambitious versions of scientific realism, such as for instance Boyd's (1981, 1984, 1985), Devitt's (1991), or Niiniluoto's (1999), according to which scientific theories are typically approximately true,<sup>7</sup> a defense of these stronger versions along the lines indicated by Dorling is obviously possible as well.

Unfortunately I do not think Dorling's defense succeeds in the first place. It may be that in his argument he has managed to correctly represent *some* antirealists, but it would be a flagrant misrepresentation of the modern, sophisticated antirealist (such as van Fraassen's constructive empiricist), who bases her position on the argument from underdetermination, if we were to identify her with the antirealist Dorling puts on the scene. The latter's ontological claim "is simply the negation of [the realist's ontological claim]" (Dorling 1992:363); for example, the antirealist "reject[s] the existence of atoms" (Dorling 1992:367). That, however, is not at all what a sophisticated antirealist does. Remember that her point merely was that, since there are empirically equivalent rivals for every scientific theory, there is no way of *knowing* the truth-value of a theory which postulates unobservables, and thus also no way of knowing that the theory is *false*. According to this antirealist, what can at most be claimed is that a given theory is empirically adequate, meaning that that theory is a member of a class of empirically equivalent theories one of which is true; to claim that the theory is false would be no less justified than to claim that it is true or approximately true, as realists under certain circumstances consider justified. She therefore counsels *agnosticism* as the proper attitude with respect to what our theories tell us about the unobservable part of the world.

---

<sup>6</sup>A further, not entirely uncontroversial, assumption in Dorling's argument is that being *virtually* certain of a proposition allows one to conditionalize on that proposition. However, like Kuipers (ICR, 65ff), I approve of this practice and believe that the problems associated with it can be solved; see Douven (2002b). Also, as Howson (2000:198–201) shows, the argument can be modified so that conditionalizing on  $T_P$  is avoided.

<sup>7</sup>The epistemological claim embodied in Kuipers' constructive realism, as presented in ICR, is certainly stronger than that made by Leplin but seems to be somewhat weaker than that of the versions of scientific realism just referred to. However, Kuipers is not very explicit on this point.

How does this affect Dorling's argument? This is made perfectly clear in the following passage from van Fraassen's (1989:193f), which in fact anticipated Dorling's argument:

Consider ... the hypothesis that there are quarks .... The scientist has initially some degree of belief that this is true. As evidence comes in, that degree can be raised, to any higher degree. That is a logical point: if some proposition  $X$  has positive probability, conditionalizing on other propositions can enhance the probability of  $X$ .

The mistake in this argument is to assume that agnosticism is represented by a low probability. That confuses lack or suspension of opinion with opinion of a certain sort. To represent agnosticism, we must take seriously the vagueness of opinion ....

Van Fraassen then goes on to argue that a person's agnosticism concerning a proposition  $H$  is to be identified with her assigning a vague or interval-valued probability  $[0, p]$  to  $H$ , where  $p$  is the probability of  $H$ 's least probable consequence(s), and that conditionalizing on other propositions can at most have the effect of removing the upper limit on that interval, or, as one might also put it, it can only increase the vagueness of  $H$ 's probability (I shall be more explicit on all this in §4.3). Thus, in Dorling's argument,  $T_P$ 's becoming certain or nearly certain would for a sophisticated antirealist *à la* van Fraassen at most effect an increase of the upper bound on her degree of belief in  $T_R$ . But that would leave her as agnostic about  $T_R$  as she was before. In particular, she cannot be said to have converted to realism with regard to  $T_R$ .

Van Fraassen's way of modelling agnosticism is not the only one nor necessarily the best; see Hájek (1998) and Monton (1998) for recent criticisms. However, the point against Dorling's defense of scientific realism arguably stands on any reasonable construal of agnosticism (like, e.g., the one suggested in Earman's 1993:35, which models agnosticism by assigning *no* probabilities to theoretical hypotheses). It thus appears that Dorling's defense is without any force against a sophisticated antirealist. Against such an antirealist, the realist has to make clear that reasons can be supplied for the claim that some particular theory is true which are not actually only reasons for the much weaker claim that it is empirically adequate, i.e., that the observable part of the world in every respect is and behaves *as if* the theory were true. In other words, the realist must provide reasons to believe that we can justifiably assign *sharp* high probabilities to particular scientific theories. Such reasons may well have to do with the explanatory force of those theories. That this is so, is the root intuition behind the rule of IBE. I shall now turn to this rule, and in particular to Kuipers' version of it.

**3. The Inference to the Best Theory and Explanatory Force.** The idea that explanatoriness is of confirmation-theoretical significance can be, and indeed has been, fleshed out in quite a variety of ways. Presumably the simplest of these, and certainly the one most frequently encountered in the literature, is the following:

IBE Given evidence  $E$  and rival (potential) explanations  $H_1, \dots, H_n$  of  $E$ , infer to the (probable/approximate) truth of the  $H_i$  that explains  $E$  best.

Even this is a rule schema rather than a rule, at least as long as it has not been supplemented by a precise account of explanation and by a set of criteria for the goodness of explanations. Here let me just note that realists agree that which theory of a collection of theories is to count as the best explanation, is to be determined on the basis of what are commonly called theoretical or non-empirical virtues, such as simplicity, elegance,

inner beauty, fertility, coherence with background theories and/or metaphysical suppositions.<sup>8</sup> Exactly how this is to be determined (how, for instance, these virtues are to be weighed against each other), is a matter of considerable controversy among realists, but we shall leave that discussion aside here.

It is also not important for van Fraassen's (1989:142ff) critique of IBE: That applies regardless of the precise understanding of the notion of explanation. The crucial point of that critique is that to make IBE a rationally compelling rule of inference it must be assumed that the truth generally is among the available potential explanations of the data to the truth of the best of which IBE allows us to infer. For, clearly, unless that is the case, IBE cannot be reliable. And since we will only rarely manage to consider all possible explanations of the evidence, that assumption seems to require some sort of privilege, viz., that we are somehow predisposed to come up with the truth when we contemplate what might explain the data. As van Fraassen (1989:143–149) convincingly argues, there is *a priori* scant reason to believe we are thus predisposed.

Numerous objections have been levelled against this so-called argument of the bad lot; see for instance Devitt (1991), Lipton (1991, 1993), Kitcher (1993), and Psillos (1996, 1999). However, for reasons given elsewhere, I believe that these objections fail (cf. Ladyman *et al.* 1997, Douven 2002a). In my view the argument of the bad lot is successful, at least to the extent that it shows IBE to rest on an unfounded assumption. But of course this is not to say that it succeeds in showing that there can be no rationally compelling rule of inference based on explanatory considerations. After all, it may well be that versions of such a rule other than IBE can do without the indeed not very plausible assumption of privilege that IBE requires. Kuipers has proposed just such a rule.

As a matter of fact, the feature of IBE the argument of the bad lot capitalizes on is one Kuipers had discovered as being undesirable independently of van Fraassen's critique (cf. Kuipers 1984, 1992, ICR). As Kuipers (ICR, 171) notes, the rule licenses a non-comparative conclusion—that a given theory is true—on the basis of a comparative premise, viz., that the particular theory is the best explanation of the evidence relative to the other theories available. That is to say, the rule displays a rather awkward asymmetry. Once the defect has thus been diagnosed, it is obvious how it can be repaired: One can either require a non-comparative premise for the rule to apply (for instance, that a given hypothesis is the absolutely best explanation, whatever other hypotheses have gone unnoticed) or one can have the rule license only a comparative conclusion when given a comparative premise as input. Kuipers opts for the second strategy, and proposes the following amended version of IBE, which he calls Inference to the Best Theory (*ibid.*):

IBT If a theory has so far proven to be the best one among the available theories, then conclude, for the time being, that it is, of the available theories, the one closest to the truth.

For later purposes I should immediately note an interesting feature of this rule, namely, that it licenses an inference to the unqualified truth of the absolutely best theory, i.e., the theory that is better than any other theory, whether considered or not. For if a theory for a given domain is better than any other theory for that domain, then it must also be closer to the truth about the domain than any other theory. But no theory can be closer to the truth than the truth itself. Hence the absolutely best theory must be identical to the true theory of the domain. By consequence, if one is certain

---

<sup>8</sup>Some authors count explanatory power itself among the theoretical virtues, but as McMullin (1996) points out, this is wrong.

that a particular theory is the absolutely best, then applying IBT yields the same result as an application of IBE would have yielded.

IBT clearly is invulnerable to the argument of the bad lot; this rule could well be reliable without our being privileged in the way we must be privileged if IBE is to be a reliable rule. And if it can be shown to be compelling, then it is—in principle (see below)—sufficient for a defense of at least a convergent scientific realism such as Kuipers’ constructive realism (and perhaps even for a defense of stronger versions of scientific realism—see Douven 2002a). After all, although the rule does not in general license belief in the unqualified truth of a theory, it does license believing that a theory is closer to the truth than any of its predecessors, provided it qualifies as better than those predecessors according to the criteria IBT assumes—and this is true even if the theory is about unobservables. This epistemic attitude contrasts sharply with the agnosticism we saw van Fraassen counsel in such cases.

In the previous paragraph the qualification ‘in principle’ was added because IBT requires a slight modification (or rather the criteria of goodness that accompany it need such modification) if it is to serve the (convergent) scientific realist’s goal. It will be apparent from our presentation of the argument from underdetermination that for the realist it is crucial that the criteria for goodness IBT assumes are such that distinct empirically equivalent theories can satisfy them to *differing* degrees. And as Kuipers understands the notion of best theory, this is not the case. Another way to put the problem is that, given the criteria for goodness that Kuipers assumes, it follows from EE that there is never a unique best theory in cases in which the theory goes beyond the observable.

To see why, consider that Kuipers equates the best theory with the most successful theory (ICR, 170), where the latter notion is spelled out in purely empirical terms. Theoretical virtues, as described at the beginning of this section, have no part whatsoever in this, and thus the notion of best theory as supposed by IBT has little to do with the notion of best explanation that is typically invoked by realists in order to (or at least in the hope that it will enable us to) discriminate between empirically equivalent theories.<sup>9</sup>

To be more specific, according to Kuipers (ICR, 112) a theory  $T_1$  is *at least* as successful as a second theory  $T_2$  exactly if both of the following conditions hold:

- (1) all individual problems of  $T_1$  are also individual problems of  $T_2$ ;
- (2) all general successes of  $T_2$  are also general successes of  $T_1$ .

$T_1$  is *more* successful than  $T_2$  exactly if at least one of the following conditions holds:

- (a)  $T_2$  has some extra individual problems in addition to those it shares with  $T_1$ ;
- (b)  $T_1$  has some extra general successes in addition to those it shares with  $T_2$ .

And, finally,  $T_1$  and  $T_2$  are *equally* successful exactly if  $T_1$  is at least as successful as  $T_2$  but not more successful than  $T_2$ .

Now an individual problem of a theory is a counterexample to what Kuipers calls a General Test Implication (GTI) of the theory; a general success of a theory is an established GTI of the theory. Without going into all the details, a GTI of a theory can be characterized as an empirical or observational law-like consequence of the theory (see ICR, 96 for a detailed account). Most significant here is, of course, the word

---

<sup>9</sup>It will have been noted that IBT does not speak of best *explanation* but only of best *theory*. However, Kuipers (ICR, 170) seems to suggest that the only reason for this is that IBT is also meant to apply to theories that have already been falsified, and of course it would be odd to call a theory the best explanation of the data if the data refute it. My point is that as Kuipers understands the notion of best theory, it would be misleading to call it the best explanation even if the theory were unfalsified.



‘observational’. For being an *observational* consequence of a theory, a GTI will be a consequence of any empirically equivalent rival theory as well; by the definition of empirical equivalence, empirically equivalent rivals have exactly the same observational consequences and thus also exactly the same GTIs. Consequently, empirically equivalent rivals are bound to have both the same individual problems (if any) and the same general successes (if any). They thus are bound to be equally successful in the sense just defined. Of course Kuipers is well aware of this, as witness, e.g., the remark in the quotation given in §1 that empirically equivalent theories “can explain the same variety of successes and success differences”. The reason he does not seem to be too bothered by this is that, as we saw, he more or less refuses to take EE seriously, but not, as we also saw, for any good reason.

It should thus be clear that the theoretical virtues will have to be taken into account in determining the betterness and bestness of theories if we want to base our defense of (convergent) scientific realism on IBT. One obvious way to modify the definition of best theory is to let the best theory among the available theories be the one that is the most successful (in Kuipers’ sense) of these theories *if there is a unique one*; else, let it be the one of the ‘equally most successful’ theories that does, on average, best with respect to the theoretical virtues (that Kuipers wants the notion of best theory also to apply to refuted theories is no impediment to this definition; refuted theories may be no less simple or beautiful than unrefuted theories).<sup>10</sup> Whether the foregoing is the optimal way to give theoretical virtues a role in determining the best of the available theories and, even more importantly, how these theoretical virtues are to be incorporated into the formal framework developed by Kuipers, are further and not readily answerable questions. Here let me just note that it is encouraging to know that in one of his most recent papers, Kuipers has made a start on the latter project (see Kuipers 2002; incidentally, in this paper he does seem to take EE seriously).

In whatever precise way theoretical virtues are going to play a role in comparing the goodness of theories, I shall henceforth assume that IBT operates on the basis of a definition of ‘best theory’ that takes these virtues into account in some formally and intuitively acceptable way. It is worth noting that this assumption does not in the least jeopardize the superiority of Kuipers’ rule over IBE, as this solely depends on the distinction between the two rules with regards to input/output symmetry.

**4. What Justification is there for the Inference to the Best Theory?** In the previous section we saw that, in contrast to IBE, IBT does not rest on an unfounded and implausible assumption of privilege. Now it is one thing to show that a rule is invulnerable to certain objections, but it is quite another to justify the rule. And if IBT is to help us in blocking the argument from underdetermination we must, of course, make plausible that IBT is indeed justified.<sup>11</sup> It is evident that the rule has no analytical justification (at least not if we assume a traditional, correspondence notion of truth).<sup>12</sup>

<sup>10</sup>Even given these more demanding criteria there is no guarantee that there will always a unique best theory; several ‘most successful’ theories in Kuipers’ sense may do equally well with respect to our additional criteria. But note that this does not bring the argument from underdetermination back. It follows from EE that *no* theory that qualifies as best, given Kuipers’ criteria, can be unique. There is no plausible thesis, however, saying that there is never a unique best theory given the criteria just proposed.

<sup>11</sup>Psillos (1999), in the course of defending IBE, argues that what is needed to justifiably apply some rule of inference is that there be no reason to doubt the rule; in the absence of such a reason, there is no requirement to actively seek to justify it. This is certainly an interesting proposal. However, for reasons given in Douven (2001) I do not endorse it.

<sup>12</sup>Kuipers (ICR, 171) seems to suggest that it does have such a justification, but that is definitely false for a version of the rule amended along the lines suggested at the end of the previous section.

Thus it seems that, if it can be justified at all, its justification must rest on empirical grounds. Several authors in the realist camp have hinted at what at first sight is the evident way to proceed in order to obtain such an empirical justification of IBT.

I will start by describing this seemingly straightforward procedure and show that it is destined to beg important antirealist issues. It is also shown that, initial appearances to the contrary notwithstanding, Hacking's famous arguments from microscopy cannot *by themselves* save the procedure, for as they stand these arguments, too, beg the issue against the antirealist (§4.1). I will then argue that we can combine Hacking's arguments with the idea of the simple procedure for testing IBT in a way that yields a—slightly more complicated—testing procedure that is *not* question-begging (§4.2). In §4.4 it is shown that this new procedure can even be made to work within the confines of the confirmation theory that van Fraassen advocates (cf. §2). To that end this confirmation theory must first be presented in greater detail than has so far been done by van Fraassen or indeed anyone else; this I do in §4.3.

It should be emphasized right away that in this section I am only concerned to establish that there is no principled difficulty in empirically justifying IBT in a way that is acceptable to both the realist and the antirealist; I do not try to argue that IBT is in fact justified. One reason for this is that carrying out the latter task would require careful historical research, which is beyond my competence to undertake. Another is that the relation between empirical support and justification (*How much empirical support for IBT is needed in order for it to be justified?*) is a tangled issue that I want to sidestep here.

*4.1. A Simple Tacking Argument for IBT (?)* What was just referred to as the *prima facie* evident way to obtain an a posteriori justification of IBT is to check the success ratio of IBT, i.e., to investigate in what percentage of its applications IBT yields a correct result, where a correct result of course does not mean that the application of IBT led to the acceptance of a true theory, but only to the acceptance of a theory that is closer to the truth than the theories it once competed with for acceptance. To give a homely example of how such a check may be carried out: When in the morning we find a used plate on our kitchen table we conclude, by means of IBT, that one of our housemates made him- or herself a midnight snack (this is undoubtedly the very best explanation for the phenomena and thus IBT allows us to conclude to the unqualified truth of it—as was explained in the previous section). That might be wrong—surely there are other explanations for the phenomena. However, we can check whether our conclusion is correct simply by asking our housemates about it. If one of them did make a midnight snack, that would constitute a success for our rule of inference. Examples like this can be multiplied at will. And it may seem that, given enough of them, we obtain the required empirical justification of IBT. For would not the evidence in that case show the rule to be reliable?

The problem is that the example just given is an example of what, slightly adapting terminology used by Psillos (1996, 1999), we could call a horizontal inference to the best theory, i.e., an inference from phenomena to a conclusion strictly about observable entities or states of affairs (in contrast to vertical inferences to the best theory, in which we infer to a conclusion solely or partly about unobservables). So the antirealist might respond to a purported justification along the above lines, that all it shows is that IBT is reliable as long as what we come to accept by means of it is strictly about the observable realm. This does not in any way help the realist who is combating the argument from underdetermination. To block that argument by means of an appeal to IBT, the rule must also be shown to be reliable when it leads to the acceptance of hypotheses concerning the unobservable.

So now the all-important question is whether there exist examples of applications of IBT that could show that IBT also leads to correct conclusions when it is applied to hypotheses about unobservables. One is immediately inclined to answer this question positively. Were viruses not postulated for explicitly explanatory reasons at the end of the nineteenth century, and then discovered about fifty years later after the electron microscope had become available? This indeed appears to be an example of a successful application of IBE or IBT<sup>13</sup> of the kind required to establish the reliability of vertical inferences to the best theory. And there seem to be many more of such examples (see Harré 1986, 1988, and Bird 1998).

This line of thought will not do to convince the antirealist, however. Recall that according to the antirealist our knowledge is restricted to the observable, where by ‘observable’ she means ‘observable by the unaided senses’. She will thus certainly object to the assertion that the existence of viruses has ever been established. Perhaps they have been discovered, perhaps not. The tiny spots biologists identified (and identify) as viruses may in reality have been (and be) due to certain aberrations of the microscopes used. As with any theoretical hypothesis, the antirealist (van Fraassen type) will want to remain agnostic about whether or not viruses were discovered last century, so that to speak of the discovery of viruses would seem to beg a central antirealist question. Hence, a tacking argument for the justification of IBE or IBT such as the one proposed above seems bound to fail.

Is the antirealist not overly cautious in refusing to take data from microscopy as evidence for the existence of viruses and the like? No doubt the antirealist will respond, and rightly so to my mind, that the history of microscopy gives reason for caution at this point; more than once, entities were ‘discovered’ and then later shown (to use a realist terminology) to be artifacts, due to some aberration(s) of the instrument or the devices used to prepare the specimens for study (cf. Hacking 1981:138ff, Atlas 1986:52f).

But even though there is reason for caution when it comes to existence claims made on the basis of data obtained from microscopes, and even though realists may in general have been too quick to assume that, thanks to technological advances, erstwhile unobservable entities have become observable, it may be possible to argue for at least certain observation devices that they are veridical. Hacking (1981) has given two intuitively very appealing triangulation arguments for the conclusion that modern types of microscope are indeed veridical.<sup>14</sup> In one of these, Hacking notes that very different types of microscope give essentially the same output when given the same input. He argues that it is just too much to believe that this is due to some coincidence, and thus that it is reasonable to believe that the different types of microscope are all veridical. If this or the other triangulation argument is correct, then it seems we may after all be able to hold in a non-question-begging way that entities once postulated on purely explanatory grounds were at a later stage actually seen by means of (a) certain type(s) of microscope. In other words, Hacking’s arguments seem to provide exactly what is needed in order to make the tacking argument for the reliability of IBT go through.

But here another problem appears. Several authors have argued, quite convincingly to my mind, that Hacking’s triangulation arguments are, implicitly, inference to the

---

<sup>13</sup>It is immaterial exactly what rule the scientists who postulated viruses were using. Even if it was not IBT that they were using, we may be sure that had that rule been applied to the theories available at the time, it would have led to the acceptance of viruses all the same—and that is what matters for present purposes.

<sup>14</sup>Hacking’s argument only involves certain, and definitely not all, types of microscope. However, there is no reason to believe the argument cannot be extended to include other types of microscope. In fact, it seems possible to extend it to observation devices other than microscopes, such as X-ray crystallography, for instance.

best explanation arguments:<sup>15</sup> That all the different microscopes figuring in Hacking’s argument just summarized give a similar output, is reason to believe that they are veridical because that is the best explanation of the former fact—that, according to those authors, is what the argument must really be. Note that this does not mean the argument requires IBE. Given that its conclusion clearly seems to be the absolutely best explanation for the fact Hacking reports, and given the earlier noted fact that in case the absolutely best explanation is among the hypotheses considered, it makes no difference whether we apply IBT or IBE, Hacking’s argument can make do with IBT. However, for the antirealist this will make little difference, for she accepts neither IBE nor IBT. It seems, then, that we were wrong to think Hacking’s arguments can help us in empirically evaluating IBT.

So far we have considered two realist arguments—one for the conclusion that IBE/IBT is a reliable rule of inference, the other for the conclusion that our current microscopes are veridical—that at least individually are unsuccessful; each relies on an assumption that the antirealist cannot be expected to grant. What I have tried to show in Douven (2002a) is that, surprisingly, in combination they may well provide the means to test IBT in a way that does not beg any antirealist issues. The trick is to suitably link the two arguments, and the link needed is Glymour’s (1980) account of bootstrap confirmation, or so I argue in the paper mentioned above. I summarize the procedure in the next subsection.

*4.2. A Bootstrap Test for IBT.* Since Duhem (at least), we have known that, in general, evidence bears on a hypothesis only relative to one or more auxiliary hypotheses. Some have taken this to imply that there can only be a relativized notion of confirmation (and even as opening the door to epistemological relativism). Now, it is Glymour’s position that Duhem’s insight does not entail that confirmation cannot be absolute. Though it is true that the basic confirmation relation is three-place rather than two-place, under certain specific circumstances we can go from relative confirmation to absolute confirmation, i.e., to the claim that certain data (dis)confirm a certain theory, period. To make this more precise, let  $T$  be a finitely axiomatizable theory, consisting of the axioms  $H_1, \dots, H_n$ , and let  $D$  be our total evidence at this time. Now suppose that for each axiom  $H_i$  of  $T$  the following three conditions hold:

- (B1) there are other axioms  $H_{j_1}, \dots, H_{j_m}$  also of  $T$  such that  $D$  confirms  $H_i$  when these latter axioms are taken as auxiliaries;
- (B2) there are possible (but non-actual) data  $D'$  that disconfirm  $H_i$  when  $H_{j_1}, \dots, H_{j_m}$  are taken as auxiliaries, i.e., adopting these latter axioms as auxiliaries in testing  $H_i$  does not guard the latter hypothesis against disconfirmation whatever the data;<sup>16</sup>
- (B3) there are no axioms  $H_{k_1}, \dots, H_{k_l}$  of  $T$  such that  $D$  disconfirms  $H_i$  when  $H_{k_1}, \dots, H_{k_l}$  are taken as auxiliaries.

Then, Glymour claims, we are allowed to conclude that the data confirm  $T$ , period, and not just that they confirm  $T$  with respect to  $T$ . In Glymour’s (1980) presentation of it, this claim is backed up by a lot of sophisticated argumentation and is further buttressed by many examples from science, showing that the schema of bootstrap confirmation is an abstract but fair representation of the way theories are tested in actual scientific practice.

<sup>15</sup>Cf. van Fraassen (1985:298), Devitt (1991:112), and Reiner and Pearson (1995:64).

<sup>16</sup>Some formulations in Glymour (1980) suggest that he actually intends, instead of (B2), the slightly weaker condition that there exist possible data that *do not confirm* (rather than disconfirm)  $H_i$  relative to  $H_{j_1}, \dots, H_{j_m}$ . The weaker condition may also suffice; cf. Douven (2002a).

The following points out how the theory of bootstrap confirmation can link the two realist arguments considered in this section so as to yield a test that can help justify IBT. Since the theory I shall be concerned with consists of two axioms only, (B3) can further be ignored. After all, given any reasonable interpretation of the notion of (dis-)confirmation as used in (B1)–(B3) it will hold that, if evidence confirms a hypothesis  $H$  given another hypothesis  $H'$  as auxiliary, then it will not also disconfirm  $H$  given  $H'$  as auxiliary.

In the first, tacking argument we sought to show that IBT is a reliable rule; call the hypothesis that it is  $R$ . Hacking claimed to have provided empirical support for the hypothesis that modern types of microscope are veridical; call this hypothesis  $V$ . It was seen that the tacking argument for IBT has no force against the antirealist, because it assumes that thanks to technological advances, in particular the development of sophisticated types of microscope, shifts have occurred in the boundary between what can and cannot be observed (for only given that assumption can we claim, for instance, that we now have observational evidence for the existence of viruses). Hacking’s argument was without force, too, because it relies on either IBE or IBT, neither of which the antirealist finds a compelling (or even acceptable) rule of inference.

But now consider the theory  $T_{V\&R} = \{V, R\}$ , and say that  $D$ , our total current evidence, comprises (among much else) the data Hacking adduces in his triangulation arguments as well as all the data available about events—alleged discoveries of erstwhile unobservable but postulated types of entities—that in our discussion of the tacking argument were said to be required in order to support the claim that vertical IBT is reliable, and suppose that, at least from a naive realist perspective, the latter data are favorable to  $R$  (i.e., from that perspective it seems that most applications of vertical IBT have been successful).<sup>17</sup> Clearly, if  $R$  is assumed as an auxiliary, then  $D$  is confirming evidence for  $V$ . Likewise, if  $V$  is assumed, then  $D$  is confirming evidence for  $R$ . But note that we now have already gone some way toward a bootstrap confirmation of  $T_{V\&R}$ . What we have called condition (B1) is satisfied. So the only thing left to show is that (B2) is satisfied as well.

Condition (B2) requires that the assumption of  $R$  in testing  $V$  does not trivialize that test in the sense that it excludes a negative result, and that, similarly, the assumption of  $V$  in testing  $R$  does not trivialize this second test. However, it is obvious that (B2) is satisfied. First, it is certainly conceivable that Hacking had obtained very different outputs from the different types of microscopes, even though they were given the same or similar inputs, the assumption of the reliability of IBT notwithstanding. And if he had, that would have cast considerable doubt on  $V$ . Secondly, making the assumption that contemporary types of microscope are veridical, be it in testing  $R$  or in testing some other hypothesis, cannot possibly make it the case that we will not find an unfavorable track record for IBT. For example, it might turn out that only very few of the putative unobservable entities once accepted on purely explanatory grounds ‘survived’ the introduction of the electron microscope.

Thus our total current evidence (or rather, what we assume to be our evidence) confirms  $T_{V\&R}$ , i.e., it confirms in tandem the hypothesis stating that IBT is a reliable rule and the hypothesis stating that modern microscopes are veridical, and, using Glymour’s idea of bootstrapping, it apparently does so without begging any antirealist issues. Now there are many objections that can be raised against these claims, both general objections to Glymour’s account and more specific ones addressing the use made of that account in the test we have just outlined. The most pressing of these I have considered and, I believe, answered in Douven (2002a). Here let me only briefly repeat

---

<sup>17</sup>If Harré and Bird, mentioned earlier, are right, then the data are indeed as here assumed. However, it seems to me that the historical evidence they cite to support this claim is rather meager.

the answer to an objection that seems invariably to come first to people's minds when they are first confronted with the above construction. This is the objection that the proposed test of  $T_{V\&R}$  is circular. This reaction is quite understandable. After all, the test combines two tests that are separately clearly question-begging; it may thus easily seem that any test combining them will have to be 'doubly' question-begging. But there is in effect an easy way to see that this is not so: Just check whether it is possible to derive  $T_{V\&R}$  from our test even before you have gathered any data (or, rather, assuming the relevant data are still unknown)! If  $T_{V\&R}$  were really presupposed, that should be possible. As you will realize, however, condition (B2) ensures that it is not. But then it is hard to see what substance there could be to the charge of question-begging.<sup>18</sup>

In order to render its logical structure more clearly, let me present the bootstrap test of  $T_{V\&R}$  in a more formal fashion. To that end, we need some notation. Given a test of some hypothesis  $H_i$ , say that ' $A_i$ ' denotes the hypothesis or conjunction of hypotheses involved as auxiliaries in that test. (Note that, in the case of a bootstrap test, a theory  $T = \{H_1, \dots, H_n\}$  is tested by testing each of the  $H_i \in T$  relative to other hypotheses belonging to the same theory, i.e., in such a test it holds that, for each  $H_i$ , either  $A_i \in T$  or  $A_i = \bigwedge_{j < m} H_j$  for some  $m$  such that  $m < n$  and  $H_j \in T$  for each  $j \leq m$ .) Furthermore, let ' $C(D; H_i; A_i)$ ' mean that evidence  $D$  confirms hypothesis  $H_i$  relative to auxiliary  $A_i$ , and let ' $D(D; H_i; A_i)$ ' mean that  $D$  disconfirms  $H_i$  relative to  $A_i$ . ' $C(D; H_i; \emptyset)$ ' means, of course, that  $D$  confirms  $H_i$  absolutely.

To represent schematically our earlier test of  $T_{V\&R}$  we need two rules of inference.<sup>19</sup> The first rule, a conjunction-introduction rule (&I), is utterly straightforward and reads as follows: If  $D$  confirms  $H_i$  relative to  $A_i$  and  $D$  confirms  $H_j$  relative to  $A_j$ , then  $D$  confirms  $H_i\&H_j$  relative to  $A_i\&A_j$ , or, in natural deduction format:

$$\frac{C(D; H_i; A_i) \quad C(D; H_j; A_j)}{C(D; H_i\&H_j; A_i\&A_j)} \&I$$

The second rule captures what plainly is the most characteristic feature of bootstrap testing, namely, that it permits us to 'cancel' auxiliaries assumed in a test on the condition that the assumption did not trivialize the test. This cancellation mechanism is formalized by the following ('non-triviality') rule:

$$\frac{\diamond\exists D_1 D(D_1; H_1; A_1) \wedge \dots \wedge \diamond\exists D_n D(D_n; H_n; A_n) \quad C(D; H_1\&\dots\&H_n; A_1\&\dots\&A_n)}{C(D; H_1\&\dots\&H_n; \emptyset)}_{NT}$$

Recall that we assumed ourselves to be in the rather fortunate position of already having obtained data confirming  $R$  relative to  $V$ , that is to say,  $D$ , our total current evidence, is assumed to comprise both the data cited in Hacking's arguments as well as 'sufficiently many' positive reports concerning the discovery of particular, earlier only hypothesized, unobservable types of entity (and only relatively few negative such reports). In formal clothing, then, the argument that our data  $D$  confirm  $T_{V\&R}$  reads as follows:

<sup>18</sup>Except for the fact that the argument might be rule-circular (cf. Psillos 1999). In Douven (2002a) I show that it is not, however.

<sup>19</sup>Remember that for our test the third clause, (B3), of the definition of bootstrap confirmation above could be neglected. I am therefore not giving a rule of inference corresponding to it. However, it is obvious how the rule should read, so those who would like to have an inferential system entirely capturing the earlier definition can easily provide it themselves.

$$\frac{\diamond\exists D'D(D'; V; R) \wedge \diamond\exists D''D(D''; R; V)}{C(D; V \& R; \emptyset)} \frac{\frac{C(D; V; R) \quad C(D; R; V)}{C(D; V \& R; V \& R)}_{\&I}}{\text{NT}}$$

The application of NT in this derivation is justified by our respective observations that nothing in the way the test is constructed could have prevented Hacking from obtaining dissimilar outputs from the various microscopes he used and that there is no guarantee that historical research is not going to provide us with data disconfirming  $R$  relative to  $V$ .

Note that so far no specific assumptions have been made about the nature of the confirmation relation involved. Somewhat surprisingly, perhaps, it is not directly necessary to do so. As Glymour (1980) makes plain, the idea of bootstrapping is not tied to any particular notion of confirmation; it can be combined with a variety of confirmation theories. Still, some of the details of our test will depend on the particular theory to which we attach the bootstraps in order to test  $T_{V \& R}$ . For instance, the answer to the question whether bootstrap confirmation of  $T_{V \& R}$  automatically yields confirmation of  $V$  and  $R$  separately, is clearly *yes* if the underlying confirmation theory is Hempel's (due to the Special Consequence Condition of that account), but is clearly *no* if the underlying theory is Bayesian confirmation theory.<sup>20</sup> Since we wish to justify IBT in a way that will also be acceptable to the antirealist, it would strategically be optimal if we can test  $T_{V \& R}$  by means of the confirmation theory (plus bootstraps<sup>21</sup>) endorsed by the antirealist herself. It seems no exaggeration to say that van Fraassen is currently the realist's only serious opponent. Let us therefore attempt to formulate the test for  $T_{V \& R}$  in the terms of the confirmation theory he advocates. We now immediately encounter a problem, however, for much is unclear about this confirmation theory. So, before we can fill in further details of our test, we will first have to help ourselves to a number of assumptions concerning the confirmation theory van Fraassen endorses.

*4.3. A Bayesian Confirmation Theory Countenancing Vague Probabilities.* Van Fraassen is a Bayesian, but one who countenances vague probabilities. Vague probabilities have no place in standard Bayesian confirmation theory, and if we want to give them a place in it questions arise that are not readily answered (and that van Fraassen nowhere addresses). Yet to be able to cast our bootstrap test in terms of the confirmation theory van Fraassen seems to advocate, these questions need to be answered. In this subsection I make certain assumptions about what the answers should be. However, I should note that it is not my objective to develop a Bayesian confirmation theory countenancing vague probabilities; among other things, this means that I do not try to give full-fledged justifications for the assumptions I make.

We want to combine Bayesian confirmation theory with the idea that hypotheses concerning the unobservable should be assigned probabilities that are vague over an interval including 0, whereas hypotheses strictly about observables should have sharp probabilities.<sup>22</sup> One question that arises then, concerns the fact that under certain

<sup>20</sup>So in the former, but not in the latter case we could add to the rules of inference presented in the text a conjunction-elimination rule, i.e., a rule allowing us to infer both  $C(D; H_i; \emptyset)$  and  $C(D; H_j; \emptyset)$  from  $C(D; H_i \& H_j; \emptyset)$ .

<sup>21</sup>As van Fraassen (1980:222) rightly notes, bootstrapping itself is neutral with respect to the realism/antirealism controversy, so there is no reason to worry that the mere fact that we make use of the mechanism of bootstrapping will suffice to make the test unacceptable in the eyes of the antirealist.

<sup>22</sup>Some antirealists may want to remain agnostic even about propositions strictly about observables. However, it is difficult to see how such an epistemic attitude could be motivated by the argument from underdetermination as presented in §1.

circumstances even an antirealist may want to change a vague probability assigned to a proposition into a sharp (positive) probability. Suppose  $A$ , a convinced antirealist who assigns vague probabilities to any hypothesis not strictly about observables (but assigns sharp probabilities to hypotheses about observables), believes hypothesis  $H$  to be about unobservables and assigns it a probability interval  $[0, x]$  for some  $x : 0 \leq x \leq 1$ . Now  $A$  discovers that her belief concerning  $H$  was erroneous and that the entities  $H$  is about are not unobservable after all. She may have heard you describe those entities as unobservables, whereas in fact you said they were observable. Or consider this: According to van Fraassen (1980:57ff) it is up to science to delimit the observable from the unobservable; and of course scientists may find out—and report to  $A$ —that certain entities or processes once classified as being unobservable are visible to the unaided eye after all. Also,  $A$  may find out, or at any rate come to believe, that, although the entities  $H$  is about are unobservable, they are, for some reason, to be treated epistemologically on a par with observables. The question now is this: How should  $A$  change her degree of belief in  $H$  under these and similar circumstances?

It seems that upon discovering that  $H$  is not really about unobservables,  $A$  should no longer assign a vague probability to it, as this would conflict with her policy of assigning vague probabilities only to hypotheses about unobservables. However, it should be remarked that, if all belief change is to proceed via conditionalization, as it is according to standard Bayesian confirmation theory, then the only kind of case in which there can be a transition from a vague probability assigned to some hypothesis to a sharp probability assigned to it, is the one in which one comes to accept evidence that refutes the hypothesis so that its probability becomes 0 sharp.<sup>23</sup> Yet one would suppose that a confirmation theory countenancing both vague and sharp probabilities would contain a rule telling us what to do in the indicated kind of cases.

To have such a rule, I will make the following assumption:

**Assumption I** If a person  $A$  believes hypothesis  $H$  to be about unobservables and therefore assigns it a probability interval  $[0, x]$ , for some  $x : 0 \leq x \leq 1$ , then, if  $A$  comes to learn that  $H$  is strictly about observables (or strictly about entities that are epistemologically on a par with observables), the rational thing to do for her is to change her probability for  $H$  from  $[0, x]$  to  $x$ .

As intimated above, I will not attempt to fully justify the assumptions made in this subsection. However, I do want to note that this assumption is at least *prima facie* plausible, for it seems that any reason one could have for assigning a probability to  $H$  that is either lower or higher than the upper bound of the old, vague probability would have been a reason to have a lower respectively higher upper bound on the probability interval assigned to  $H$  when one still believed the hypothesis was about unobservable entities.

Further questions about van Fraassen's confirmation theory that are relevant to our project concern the notions of confirmation and disconfirmation. There are well-known definitions of confirmation and disconfirmation, respectively, for standard Bayesian confirmation theory— $E$  confirms  $H$  exactly if  $p(H|E) > p(H)$ , and disconfirms  $H$  exactly

---

<sup>23</sup>For those familiar with van Fraassen's theory of vague probabilities, the reason for this is that if one is agnostic about  $H$ , then one's representor, i.e., roughly, the set of probability functions compatible with one's opinion, will include a probability function that assigns 0 to  $H$ . Since conditionalizing one's representor on new evidence comes down to conditionalizing each probability function in one's representor, and since conditionalizing a function that assigns 0 to  $H$  will result in another function that also assigns 0 to  $H$ , the only sort of occasion on which someone agnostic about  $H$  can come to have a sharp opinion about  $H$  purely by conditionalization is when all the probability functions in her representor come to assign  $H$  a probability of 0.



if  $p(H|E) < p(H)$ —but it is not straightforward to derive from these definitions, definitions of confirmation and disconfirmation for the theory we are assuming here. I will not try to supply such definitions for the theory, but will only make one assumption about a sufficient condition for confirmation and another about a sufficient condition for disconfirmation:

**Assumption II** If  $p(H) = [0, x]$  and  $p(H|E) = y$ , with  $y > x$ , then  $E$  confirms  $H$ .

**Assumption III** If  $p(H) = [0, x]$  and  $p(H|E) = [y, z]$ , with  $z < x$  and  $0 \leq y \leq z$ , then  $E$  disconfirms  $H$ .

Assumption II concerning confirmation should be unproblematic. As regards what counts as disconfirmation of a hypothesis that is assigned a vague probability, it seems to me that the only cases that are entirely unproblematic are those in which the evidence refutes the hypothesis (so that  $p(H|E) = 0$ ). However, although the matter is not quite so uncontroversial as in this latter kind of case, I do think it is sensible to say that disconfirmation occurs in the kind of case satisfying the condition of Assumption III.

*4.4. The Bootstrap Test Reconstructed.* I already said that our bootstrap test is sensitive to the underlying confirmation theory that is used. For reasons that will soon become clear, the effect of using the above confirmation theory is that we can obtain a positive test result only for a theory  $T_{V\&R'} = \{V, R'\}$  in which  $R'$  is a slightly restricted version of  $R$ ,<sup>24</sup> namely the thesis stating that IBT is reliable if it is used to derive conclusions about observables and/or unobservables of a specific type. For simplicity I assume that this type can be specified in terms of size. Let us say that, *if*  $V$  is correct, then entities of size  $S$ , but not of any smaller size, can be seen by means of our current microscopes (no doubt other features of an entity than just its size will be relevant to the question whether it can be seen by means of a microscope—provided the microscope is veridical—but as I said, I am simplifying here). Then  $R'$  is the hypothesis stating that IBT is reliable if it is used to derive conclusions about entities of size  $S$  or larger. Clearly, since  $R'$  is a weaker hypothesis than  $R$ , replacing  $T_{V\&R}$  by  $T_{V\&R'}$  diminishes the significance of the test.<sup>25</sup>

To begin our test of  $T_{V\&R'}$  within what we assume to be the confirmation theory accepted by the antirealist (or by van Fraassen, in any event), first note that both  $V$  and  $R'$  are about unobservables:  $V$  says that certain unobservables (things too small to be seen by the naked eye) *can* be seen by means of our microscopes;  $R'$  says that IBT is reliable when it licenses inferences about observables *and also when it licenses inferences about unobservables that can be seen by means of our microscopes in case  $V$  is true*. So, the antirealist will want to remain agnostic about these hypotheses and accordingly assign vague probabilities to both of them. Suppose that  $p(V) = p(R') = [0, .5]$  represents her prior degrees of belief in the two hypotheses (nothing much hinges on the exact values of the upper bounds, and nothing at all on the assumption that  $p(V) = p(R')$ ).

Let us now see whether the data support  $V$  relative to  $R'$  and  $R'$  relative to  $V$ , in that order.

---

<sup>24</sup>Or, more carefully, I can only see how to obtain a positive test result for  $T_{V\&R'}$ ; it is not excluded that, in a manner I am presently unaware of, a bootstrap test for  $T_{V\&R}$  can be constructed on the basis of the confirmation theory here assumed.

<sup>25</sup>The diminution may in the end not be very considerable, though, given that, as was said in note 14, the test Hacking proposes for microscopes may well be extensible to other observation devices, so that we could have a bootstrap test with  $V$  replaced by a hypothesis  $V'$  stating that these other devices plus our current microscopes are veridical.  $R'$  could then in turn be replaced by a stronger hypothesis (even though it may not be possible ever to replace it by  $R$ ).

Put briefly, the argument that the data support  $V$  relative to  $R'$  is as follows. Clearly  $V$  is about unobservables of size  $S$  or larger; it says that entities of at least size  $S$  can be seen by means of certain types of microscope. Thus  $R'$  is sufficient to draw from the data reported in Hacking's triangulation arguments, and from the fact that  $V$  is the very best explanation for those data, the conclusion that  $V$  is very likely correct. So, assuming  $R'$  as an auxiliary, we may conclude that the probability of  $V$  is 1 or close to 1, i.e., relative to  $R'$  the data bestow a sharp probability on  $V$  and a probability that is higher than what we assumed to be the upper bound of the antirealist's prior degree of belief in  $V$ . Hence, by Assumption II the data confirm  $V$  relative to  $R'$ .

The latter part of this argument is evident. However, I am sure it can do no harm to state in rather more detail the sub-argument for the conclusion that the data make  $V$  very likely given  $R'$  as an auxiliary. In this sub-argument we start from the premise that  $V$ , the hypothesis stating that entities of a size no smaller than  $S$  can be seen by means of modern microscopes, is the very best explanation of the data. By IBT, it follows from this that  $V$  is true (recall that in case a hypothesis constitutes the very best explanation for the data, IBT allows us to infer to the unqualified truth of that hypothesis). Now our auxiliary,  $R'$ , states that conclusions concerning hypotheses about observables and/or unobservables of size  $S$  and larger, reached on the basis of IBT, are very likely true. Thus the conclusion that  $V$  is true is very likely true, i.e.,  $V$  is very likely true, or, put differently again,  $V$  is very likely. That is, given  $R'$ , we have a deductively valid argument for the conclusion that our current microscopes are very likely veridical.

It may be noted that we do not reach the conclusion that  $V$  is very likely by updating the antirealist's initial degree of belief in  $V$  via conditionalization.<sup>26</sup> Now this may seem strange, for we clearly are assuming a non-standard Bayesian confirmation theory that, besides the rule described in Assumption I, normally only regards conditionalization as a legitimate way of updating one's degrees of belief. However, because in the argument  $R'$  is assumed as an auxiliary, a third rule becomes (at least temporarily) available, namely IBT, and it is this rule that allows us to conclude that  $V$  is very probable given the data.

As a next step in our test, we must see whether the data confirm  $R'$  relative to  $V$ . At this point Assumption I becomes relevant. For one problem we seem to encounter is that the realist and antirealist may disagree over what the data are. Assume that a realist would report the data relevant to this part of the test as follows: There exist data about discoveries by means of modern types of microscope of entities that once were postulated on explanatory grounds, and these data are in fact favorable to IBT. Evidently the antirealist will demur at this description. She will only be willing to grant that all observable phenomena are *as if* such entities were discovered. More specifically she will want to remain agnostic about the question whether the apparent discoveries were real discoveries, and thus will assign a probability to sentences such as

Viruses were discovered fifty years after they were postulated, (\*)

that is vague over the whole interval  $[0, 1]$  (here the assumption that the antirealist's upper bound is 1 is appropriate because the antirealist will certainly acknowledge that (\*) is empirically adequate: All observable phenomena are in any case *as if* viruses were discovered fifty years after they were postulated).

However, if  $V$  is assumed as an auxiliary, then (\*) and kindred sentences are about entities to which we have epistemic access no less than we have epistemic access to

---

<sup>26</sup>As explained earlier, that initial degree of belief *could* not be updated by conditionalization and result in a sharp, non-zero probability.

objects that can be seen by the unaided eye. Hence, given  $V$ , Assumption I applies to such sentences, and the antirealist's probability for them should 'collapse' to 1. This collapse should also occur in the case of her probability for  $R'$ , which we assumed to be  $[0, .5]$ . After all, given  $V$ ,  $R'$  is a hypothesis about observables and unobservables that are to be treated epistemologically on a par with observables. We may now assume that conditionalization on the data (provided they are as favorable to IBT as the realist hopes they are) will raise the antirealist's degree of belief in  $R'$  to some value  $x > .5$ . It then follows from Assumption II that the data confirm  $R'$  relative to  $V$ .

To complete the test, we must check whether bootstrap condition (B2) is satisfied, i.e., whether in both cases disconfirmation would have been possible or whether this was prevented by the specific choice of the respective auxiliaries. Here we can refer to what was said in our general exposition of the bootstrap test. Had the outputs from the various types of microscopes Hacking obtained been very dissimilar despite the fact that the inputs used were the same or at least similar to each other (clearly the occurrence of such data could not be prohibited by assuming  $R'$ ), the upper bound of the antirealist's degree of belief in  $V$  would certainly have been lowered. Thus in that case the data would, by Assumption III, have disconfirmed  $V$ . Furthermore, we have only assumed that the track record of IBT was favorable on the assumption of  $V$ ; in actuality we could certainly find an unfavorable track record for it on the same assumption. If we do, then, we may plausibly suppose, this will lead the antirealist to have a degree of belief in  $R'$  that is sharp (because the rule of Assumption I applies) but that also is lower than the upper bound on her initial degree of belief. Given Assumption III, the data would thereby disconfirm  $R'$ . Hence condition (B2) is satisfied.

**5. Conclusion.** We have seen that the realist seems to be in need of some rule like IBE if she is to successfully block the argument from underdetermination. It was argued that the realist is much better off with Kuipers' rule of inference, IBT, than with the more popular IBE, provided we adopt somewhat stricter criteria for the goodness of theories than the ones Kuipers proposes. As regards the question of the rule's legitimacy, in the main part of this paper I have argued that, initial appearances to the contrary notwithstanding, it is possible to test IBT, or in any event—depending on what confirmation theory we employ—it is possible to test it in combination with the hypothesis that the types of microscope currently in use are veridical, and that this can be done in a way that does not beg any of the antirealist's central tenets. So at least in principle it seems possible for the realist to justify IBT in a non-question-begging fashion. It has even been seen that we can bootstrap test (again in combination with the hypothesis concerning our microscopes) the reliability of IBT in a restricted, but for the realist still interesting, domain even if we employ as the underlying confirmation theory of the test the one preferred by van Fraassen.

## References

- Atlas, R. (1986) *Basic and Practical Microbiology*, New York: Macmillan.
- Bird, A. (1998) *Philosophy of Science*, London: UCL Press.
- Boyd, R. (1981) 'Scientific Realism and Naturalistic Epistemology', in P. Asquith and R. Giere (eds.) *PSA 1980*, vol. II, East Lansing MI: Philosophy of Science Association, pp. 613–662.
- Boyd, R. (1984) 'The Current Status of Scientific Realism', in Leplin (ed.) (1984), pp. 41–82.
- Boyd, R. (1985) 'Lex Orandi est Lex Credendi', in Churchland and Hooker (eds.) (1985), pp. 3–34.
- Churchland, P. and C. Hooker (eds.) (1985) *Images of Science*, Chicago: University of Chicago Press.
- Devitt, M. (1991) *Realism and Truth*, Oxford: Blackwell.
- Dorling, J. (1992) 'Bayesian Conditionalization Resolves Positivist/Realist Disputes', *Journal of Philosophy* 89:362–382.
- Douven, I. (1995) 'Boyd's Miraculous No Miracle Argument', in P. Cortois (ed.) *The Many Problems of Realism*, Tilburg: Tilburg University Press, pp. 89–116.
- Douven, I. (1999) 'Inference to the Best Explanation Made Coherent', *Philosophy of Science* (Proceedings) 66:S424–S435.
- Douven, I. (2000) 'The Antirealist Argument for Underdetermination', *Philosophical Quarterly* 50:371–375.
- Douven, I. (2001) 'Quests of a Realist', *Metascience* 10:354–359.
- Douven, I. (2002a) 'Testing Inference to the Best Explanation', *Synthese* 130:355–377.
- Douven, I. (2002b) 'A New Solution to the Paradoxes of Rational Acceptability', *British Journal for the Philosophy of Science* 53:391–410.
- Douven, I. and L. Horsten (1998) 'Earman on Underdetermination and Empirical Indistinguishability', *Erkenntnis* 49:303–320.
- Earman, J. (1993) 'Underdetermination, Realism, and Reason', in P. French, T. Uehling, Jr., and H. Wettstein (eds.) *Midwest Studies in Philosophy*, vol. XVIII, Notre Dame: University of Notre Dame Press, pp. 19–38.
- Fine, A. (1984) 'The Natural Ontological Attitude', in Leplin (ed.) (1984), pp. 83–107.
- Glymour, C. (1980) *Theory and Evidence*, Princeton: Princeton University Press.
- Hacking, I. (1981) 'Do We See Through a Microscope?', *Pacific Philosophical Quarterly* 62:305–322. (Reprinted in Churchland and Hooker (eds.) (1985), pp. 132–152; the page reference is to the reprint.)
- Hájek, A. (1998) 'Agnosticism Meets Bayesianism', *Analysis* 58:199–206.
- Harré, R. (1986) *Varieties of Realism*, Oxford: Blackwell.
- Harré, R. (1988) 'Realism and Ontology', *Philosophia Naturalis* 25:386–398.
- Howson, C. (2000) *Hume's Problem: Induction and the Justification of Belief*, Oxford: Clarendon Press.
- Kitcher, P. (1993) *The Advancement of Science*, Oxford: Oxford University Press.
- Kuipers, T. (1984) 'Approaching the Truth with the Rule of Success', *Philosophia Naturalis* 21:244–253.
- Kuipers, T. (1992) 'Naive and Refined Truth Approximation', *Synthese* 93:299–341.
- Kuipers, T. (2000) *From Instrumentalism to Constructive Realism*, Dordrecht: Kluwer.
- Kuipers, T. (2002) 'Beauty, a Road to the Truth?', *Synthese* 131:291–328.
- Kukla, A. (1995) 'The Two Antirealisms of Bas van Fraassen', *Studies of History and Philosophy of Science* 26:431–454.
- Kukla, A. (1998) *Studies in Scientific Realism*, Oxford: Oxford University Press.

- Ladyman, J., I. Douven, L. Horsten and B. van Fraassen (1997) ‘A Defence of van Fraassen’s Critique of Abductive Inference’, *Philosophical Quarterly* 47:305–321.
- Laudan, L. (1990) ‘Demystifying Underdetermination’, in C. Savage (ed.) *Scientific Theories*, Minneapolis: University of Minnesota Press, pp. 267–297.
- Laudan, L. and J. Leplin (1991) ‘Empirical Equivalence and Underdetermination’, *Journal of Philosophy* 88:449–472.
- Leplin, J. (1984) *Scientific Realism*, Berkeley: University of California Press.
- Leplin, J. (1997) *A Novel Defense of Scientific Realism*, Oxford: Oxford University Press.
- Leplin, J. (2000) ‘The Epistemic Status of Auxiliary Hypotheses’, *Philosophical Quarterly* 50:376–379.
- Lipton, P. (1991) *Inference to the Best Explanation*, London: Routledge.
- Lipton, P. (1993) ‘Is the Best Good Enough?’, *Proceedings of the Aristotelian Society* 93:89–104.
- McMullin, E. (1996) ‘Epistemic Virtue and Theory Appraisal’, in I. Douven and L. Horsten (eds.) *Realism in the Sciences*, Leuven: Leuven University Press, pp. 13–34.
- Monton, B. (1998) ‘Bayesian Agnosticism and Constructive Empiricism’, *Analysis* 58:207–212.
- Niiniluoto, I. (1999) *Critical Scientific Realism*, Oxford: Clarendon Press.
- Psillos, S. (1996) ‘On van Fraassen’s Critique of Abductive Reasoning’, *Philosophical Quarterly* 46:31–47.
- Psillos, S. (1999) *Scientific Realism: How Science Tracks Truth*, London: Routledge.
- Reiner, R. and R. Pierson (1995) ‘Hacking’s Experimental Realism: An Untenable Middle Ground’, *Philosophy of Science* 62:60–69.
- van Fraassen, B. (1980) *The Scientific Image*, Oxford: Clarendon Press.
- van Fraassen, B. (1985) ‘Empiricism in the Philosophy of Science’, in Churchland and Hooker (eds.) (1985), pp. 245–308.
- van Fraassen, B. (1989) *Laws and Symmetry*, Oxford: Clarendon Press.