Kuhnian theory-choice and virtue convergence: facing the base rate fallacy

Abstract

The arguably strongest argument for scientific realism, the no-miracles-argument, has been accused of the so-called base rate fallacy, namely the neglect of the dependence of posterior probabilities on prior probabilities. The apparent inaccessibility of the base rate of true theories has even been said to undermine the rationality of the entire realism debate. In this paper I confront this challenge by arguing, on the basis of the Kuhnian picture of theory choice, that a theory is likely to be true if it possesses multiple theoretical virtues, even when the base rate converges to zero.

Key words: scientific realism; no miracle argument; base rate fallacy; theoretical virtues; T.S. Kuhn; theory choice; convergence; witness testimony.

1 Introduction

The arguably strongest argument for scientific realism, the No-Miracles-Argument (NMA), has it that it would be a miracle if our theories were as successful as they are, and not be true. As Howson (2000) pointed out, however, as normally stated, the NMA commits the so-called base rate fallacy: it ignores the base rate of true theories. Expressed in Bayesian terms, it ignores the dependence of the posterior probability of a successful theory being true on the prior probability of a theory being true. But setting the base rates seems elusive. If probabilities are construed objectively, then it looks as though we have no way of finding out about them. If, on the other hand, probabilities are construed subjectively, then both the realist and antirealist can set the priors as they please. A rational debate about realism is therefore impossible (Magnus and Callender 2004).

In spite of the fact that the severity of Magnus and Callender’s challenge is widely appreciated, head-on confrontations of their claims have been far and between.1 Whilst the current paper does little to undermine Magnus and Callender’s fundamental point, it will nevertheless, in the face of it, try to tilt the balance to the realist’s favour on the basis of the Kuhnian picture of theory-choice.

1 See for example (Psillos 2009) and a reply by (Howson 2013). For another recent attempt see (Menke 2013).
Although the paper will assume large parts of the Kuhnian picture of theory-choice, the purpose of this paper is not exegetical. That is, the purpose of this paper is not to reconstruct Kuhn’s view of theory-choice in a way that makes best sense of his view in the context of his other works. Rather, the paper will seek to explore some interesting implications given (some parts of) the Kuhnian framework of theory choice. The view defended here may thus very well be detrimental to some of the views held by Kuhn.

The paper proceeds as follows. Section 2 specifies Magnus and Callender’s challenge. Section 3 outlines how the Kuhnian picture of theory choice provides the resources for generating an argument for realism via the convergence of scientists’ truth judgements about theories on the basis of those theories’ virtues. I refer to this argument as VIRTUE NO-COINCIDENCE-ARGUMENT. Section 4 develops a formal apparatus for VIRTUE NC with the help of Earman’s Bayesian rendering of the convergence of witness reports. Section 5 spells out this apparatus for VIRTUE NC in detail. Section 6 argues that there are rational constraints on scientists’ truth judgements. Section 7 concludes this paper.

2 Magnus and Callender’s challenge

Magnus and Callender distinguish between ‘wholesale’ and ‘retail’ arguments for realism, i.e., arguments about “all or most of the entities posited in our best scientific theories” and arguments about “specific kinds of things, such as neutrinos”, respectively (321). While they think that there may be good grounds for defending retail arguments, they urge that “the wholesale realism debate should be dissolved”, for wholesale arguments amount to no more than “adamant, futile table thumping” (322). Their skepticism is grounded in their claim that realists and antirealists alike commit the base rate fallacy.

The base rate fallacy can be illustrated with a simple example from the medical context. Suppose we were to test the presence of some disease T in a population of subjects with a very effective test. That test, suppose, would have a very high probability of indicating to us the presence of a disease, when the disease is really present in a subject. Let us refer to a positive test result as e. Expressed formally, then, $P(e|T) \gg 0$. Suppose further that the test has a very low false positive rate. That is, the test is unlikely to indicate the presence of the disease when it is actually absent ($P(e|\neg T) \ll 1$). For concreteness’s sake, assume that $P(e|T) = 1$ and $P(e|\neg T) = .05$. Contrary to many people’s intuitions, it would then be fallacious to infer that the probability of some subject having the disease, when the
test indicates that the subject has it ($P(T|e)$), is high, for example .95. In fact, it can be rather low. If the disease is very rare in the population (i.e., $P(T) \ll 1$), for example 1/1000 then, given the effectiveness of our test, we would expect 51 subjects to test positive. Because, by assumption, only one of those actually has the disease, $P(T|e)$ would be just .02, that is, much lower than the intuitive .95.

Magnus and Callender accuse the partakers in the realism debate of having made the same mistake. That is, they accuse realists and antirealists of having neglected the base rate of true theories in the pool of all theories / the prior probability of a(ny) theory being true.\(^5\) Instead the debate has focused on the probability of a theory being false if successful $P(\neg T|e)$, and the likelihood of $e$ given $\neg T$ (i.e., the false positive rate $P(e|\neg T)$), where $P(T)$ now is to be interpreted as the probability of a theory being true and $P(e)$ as the probability of a theory being successful. Whereas antirealists have sought to increase $P(\neg T|e)$ with arguments like the Pessimistic Meta Induction, realists have tried to decrease $P(e|\neg T)$ by restricting the notion of success to novel success (327).\(^6\) But without knowledge of the base rate, engaging in arguments about the posteriors appears meaningless.

Although Magnus and Callender believe that their challenge is equally futile to realists and antirealists, they pose the following dilemma to the realist:

Either there is a way of knowing the approximate base rate of truth among our current theories or there is not. If there is, then we must have some independent grounds for thinking that a theory is very likely true; yet if we had such grounds, the no-miracles argument would be superfluous. If there is not, then the no-miracles argument requires an assumption that some significant proportion of our current theories are [sic] true; yet that would beg the question against the anti-realist. (328)

Because they see no way out of the dilemma, Magnus and Callender conclude that the entire wholesale realism debate is an irrational debate, which better be dissolved:

Without independent methods for estimating crucial base rates, there is little to do but make arguments that beg the question. Wholesale realism debates persist not due to mere stubbornness, but because there is no reason for opponents to disagree (336; original emphasis).\(^7\)

---

\(^5\) Magnus and Callender’s contribution can be seen as a synthesis of earlier points made by (Howson 2000) and (Lewis 2001).

\(^6\) $P(\neg T|e)$ and $P(e|\neg T)$ are related by Bayes’ theorem: $P(\neg T|e) = \frac{P(e|\neg T)\times P(\neg T)}{P(e)}$.

\(^7\) Similarly, when the probabilities are interpreted as subjective probabilities, Magnus and Callender also “can’t imagine how one could find a reasonable set of priors” (329).
Although I think Magnus and Callender are correct in their diagnosis, the consequences of their insight, I think, can be alleviated and the second horn of their dilemma be rejected.8

3 Kuhnian theory choice and the idea of virtue convergence

In his *The Structure of Scientific Revolutions* (1962/1996), T.S. Kuhn claimed that paradigm change, such as the change from Newtonian to relativistic mechanics, or from the phlogiston to the oxygen theory of combustion, “cannot be … forced by logic [or] neutral experience” (149). Rather each paradigm comes with its own set of evaluation criteria. Whenever scientists have to choose between paradigms, “each paradigm will be shown to satisfy more or less the criteria that it dictates for itself and to fall short of a few of those dictated by its opponent” (109). In other words, paradigm change is circular in the sense that changing a paradigm must rely on the evaluation criteria that the new paradigm identifies as important (and which will be different from the criteria identified as important by the old paradigm). About ten years after *Structure* Kuhn tried to answer those who (rightly) accused him of putting the case for relativism in a seminal paper on theory choice (Kuhn 1977). Departing from *Structure* to a degree that he probably did not quite realize, Kuhn in this paper advanced the view that there is a standard set of theoretical virtues on the basis of which theories have been assessed by scientists. Kuhn, without claiming either originality or completeness, mentions five prominent virtues: empirical accuracy, (internal and external) consistency, scope, simplicity, and fertility.9

Kuhn—slightly reluctantly—distinguished between an objective and subjective element of theory choice (359). The former concerns the set of virtues involved in theory choice. Kuhn writes: “the criteria or values deployed in theory choice [i.e., the virtues] are fixed once and for all, unaffected by their participation in transitions from one theory to another” (Kuhn 1977, 364). Although Kuhn assigned universality to the five standard theoretical virtues, he believed that there was a lot of room for legitimate disagreement among practitioners in deciding which theory to adopt. This constitutes the subjective element of theory choice. Each scientist, Kuhn claims, has different preferences for weighting the standard theory choice criteria. Whereas some prefer simpler theories, for instance, others prefer more unified theories, and so on.10 It is this subjective element, that led Kuhn to the conclusion that there is ‘no neutral algorithm’ for theory choice to which all practitioners

---

8 See the first quotation above: “a theory [must be] very likely true” … “a significant proportion of our current theories [must be] true”. See also their p. 325 (end of the second last paragraph).
9 Theoretical virtues are also sometimes denoted as ‘values’. In fact Kuhn himself suggested that label. I prefer ‘virtues’ because ‘values’ have ethical connotations. Recently there has been a debate about the virtues of the *scientists* making theory-choice (Ivanova 2010, Stump 2007). My discussion instead focuses on the virtues of *theories*.
10 Although Kuhn thought the five standard virtues are relevant to theory choice throughout the history of science, he thought that ‘application of these values’ and the ‘relative weights attached to them’ changed (364-5).
would be bound (Kuhn 1977 199). Thus, “two men fully committed to the same list of
criteria for choice may nevertheless reach different conclusions” (Kuhn 1977, 358). The
subjective element of theory choice, however, does not imply that theory choice would be
arbitrary. The standard criteria of theory choice are not projections; they map onto actual
theory properties. Theories really are accurate, consistent, fertile, and so on, or they are not.

Another remark by Kuhn seems to undermine the objective element in theory choice:
thoretical virtues are ‘imprecise’ or ‘ambiguous’. By that he meant that different
practitioners might refer to different properties of a theory with the same term. For example,
one practitioner might refer to quantitative parsimony and another one to qualitative
parsimony when calling the theory ‘simple’. Yet this problem should constitute no major
obstacle for theory choice: barring the much criticized Kuhnian communication failures,
practitioners should be able to specify to their peers what properties they are referring to.
Practitioners might then still disagree about how these two kinds of simplicity ought to be
weighted, of course. The ambiguity problem thus arguably reduces to the weighting
problem (Okasha 2011).  

A similar point can be made with regards to empirical accuracy. In a theory-choice
situation, one theory might be empirically accurate with regards to one set of evidence, and
another theory empirically accurate with regards to another set of evidence. In such a case
we would of course have to fine-grain the virtue of empirical accuracy. Scientists may then
disagree as to whether one or the other data set is to be given preference when it comes to
the choice between the two theories.

The weighting problem is contingent on another part of the Kuhnian picture of
theory choice. According to Kuhn, as a matter of empirical fact, the virtues “repeatedly
prove to conflict with one other” (357). In other words, as a matter of empirical fact, theories
repeatedly do better than others with regard to some criteria, but worse with regard to
others. For convenience let us refer to this claim as ‘CONFLICT’. When there is CONFLICT, and
when scientists have different weighting preferences, there will be diverging theory choices.

CONFLICT can also occur within the category of one particular virtue. Take again
empirical accuracy. One theory may perform better with regard to one set of data, and
another theory better with regard to another set of data. Scientists with different preferences
will end up choosing different theories.

11 Some readers suggested to me that Kuhn believed that the virtues were intrinsically and inextricably
vague, with no disambiguation being possible. I have found no evidence in Kuhn’s text for this
suggestion. Regardless, disambiguation may not always be unequivocal, of course. There may be
boundary cases. But boundary cases do not necessarily imply that we cannot reach agreement. In our
everyday life, for example, we pretty successfully manage to agree on what we consider to be a bald
person, in spite of baldness being a standard example for a vague predicate.
CONFLICT can also occur within the category of one particular virtue. Take for example empirical accuracy. One theory may perform better with regard to one set of data, and another theory better with regard to another set of data. This is another reason for why Kuhn thinks theory choice is often indeterminate (357). Again, the ‘subjective’ element of theory choice, i.e., the interests and preferences of the investigator, will influence which data set an individual will assign greater weight, and accordingly, which theory she will end up choosing.

Interestingly, and somewhat counterintuitively, CONFLICT, can explain theory choice convergence. Again, CONFLICT is an empirical thesis. That is, at least prima facie, there is nothing intrinsic in the virtues or their relationships that would cause CONFLICT. Because that is so, there should therefore be situations in which there are theories that do better with regard to any virtue. In that case the subjective element of theory choice, which Kuhn was so keen to stress, simply cancels out. If I prefer simple theories and you unified theories, then we will adopt different theories when there is no theory that has both of these properties. Yet when there is a theory that is both simple and unified and its competitors are not, then we will end up choosing the same theory despite our diverging preferences.

Perhaps even more interestingly, CONFLICT offers a new argument for realism. Roughly, it goes like this: if CONFLICT is true and there are only sometimes theories that exhibit several or even all of the standard virtues, then it would be a strange coincidence if a theory had all the five virtues and not be true. Further below we will see that the fully fleshed out argument will involve converging judgements by scientists concerning a theory’s truth on the basis of that theory’s virtues, despite scientists’ different weighting preferences. The argument presented here will thus involve both the ‘objective’ and ‘subjective’ aspects of Kuhnian theory choice. It is easy to see: just like the standard NMA for realism, this is a no-coincidence (NC) argument. I will therefore refer to it as VIRTUE-NC. As I will argue in the following, however, it is more powerful than the NMA: it offers an escape from Magnus and Callender’s challenge.

Before proceeding, however, let us note that CONFLICT does seem to possess a good deal of prior plausibility. If it was easy to construct theories that possessed all of the standard virtues, and most of the theories we come up with possessed all of the standard virtues, then a theory possessing all of the virtues wouldn’t warrant singling out as a hopeful truth-candidate. At the same time scientists would have a hard time making their theory-choices. But very often, they don’t. Of course, although CONFLICT thus does seem plausible, it is an open question how frequent it is that theories do not possess all of the virtues. This question is an empirical one and beyond the scope of this article. For the purposes of this paper we shall see, however, the answer to this question will not matter.
4 Convergence, witnesses, and Bayes

Although clearly related, **VIRTUE NC** differs from the NMA in that it appeals to the persuasive power of the convergence of several independent information sources. Whereas the NMA exploits the fact that there are so many ways in which a theory could have been wrong, the class of arguments of which **VIRTUE NC** is a member banks on the fact there are so many ways in which each information source could have produced a result inconsistent with the other sources. Several philosophers have used related ‘convergence’ arguments. Arguing against the thesis of theory-ladenness of observation, for instance, Hacking (1983) pointed out that it would be a strange coincidence if several of our instruments (e.g. the light and the electron microscope), presupposing different background theories, were to produce the same data, if the data were not correct. Likewise, Salmon (1984) pointed out that it would be an inexplicable coincidence if J.-B. Perrin’s half-dozen experiments in 1911 all had produced the same value of Avogadro’s number and that number had not been correct (see also Cartwright 1983, van Fraassen 2009, Chalmers 2011, Psillos 2011). These ‘convergence’ arguments are in fact analogous to arguments for the trustworthiness of witness reports in the case of several independent witnesses reporting the same murderer: we’d be compelled to believe that several witnesses tell the truth if they independently from each other (i.e., without coordinating their beliefs) report the same murderer—even when the individual reliability of the witnesses is poor. C.I. Lewis (1946) concludes that “this agreement [between witness reports] is highly unlikely; the story any one false witness might tell being one out of so very large a number of equally possible choices” (246). In other words, the probability that any one witness would report, for example, the same murderer as all the other witnesses do, if the witness report were not correct, is very low. I will argue below that a theory possessing all of the Kuhnian virtues, and being judged true on the basis of its virtues by the scientific community accordingly, is analogous to independent witnesses all reporting the same murderer.

The intuitive persuasive power of convergent witness reports can be made precise by employing Bayes’s theorem (Earman 2000). Let \( P(V_i|T) \) represent the probability that a witness \( i \) gives a report \( V \) that an event \( T \) happened when that event actually happened, and \( P(V_i|\neg T) \) the probability that a witness reports an event when the event did not happen, and \( P(T) \) the prior probability of the event itself. Assuming that the witnesses are equally reliable and independent, the following equalities hold

---

12 For a more general discussion about robustness arguments in science see (Hudson 2013). Some antirealists have taken the view that coincidences need not be explained (cf. van Fraassen 1980). In the face of there being a plausible explanation being available, I personally would regard such a move irrational. Yet the issue would lead us too far astray to take this up in any more depth on this occasion. I thank one of the anonymous referees for pressing me on this point.

13 Earman also conditionalises on the witnesses background knowledge and evidence, which I’ll leave out here for the sake of simplicity.
\[ P(V_1 \& \ldots \& V_n) = P(V_1) \times P(V_2) \times \ldots \times P(V_n) = P(V)^n \]

\[ P(V_1 \& \ldots \& V_n | T) = P(V_1 | T) \times P(V_2 | T) \times \ldots \times P(V_n | T) = P(V | T)^n \]

\[ P(V_1 \& \ldots \& V_n | \neg T) = P(V_1 | \neg T) \times P(V_2 | \neg T) \times \ldots \times P(V_n | \neg T) = P(V | \neg T)^n. \]

The posterior probability of the truth of a report given \( n \) witnesses making the same observations, by Bayes’s theorem, is

\[
P(T | V^n) = \frac{1}{1 + \left[ \frac{1 - P(T)}{P(T)} \right] \times \left[ \frac{P(V_1 | \neg T) \times P(V_2 | \neg T) \times \ldots \times P(V_n | \neg T)}{P(V_1 | T) \times P(V_2 | T) \times \ldots \times P(V_n | T)} \right]^{n}}
\]

which, assuming equally reliable witnesses, reduces to

\[
P(T | V^n) = \frac{1}{1 + \left[ \frac{1 - P(T)}{P(T)} \right] \times \left[ \frac{P(V_1 | \neg T)}{P(V | T)} \right]^{n}}
\]

Earman points out that for the occurrence of the event to be credible, the witnesses need not be reliable in the absolute sense. All that is required, rather, is that they be reliable in the relative sense, so that \( P(V_i | \neg T) < P(V_i | T) \), since in that case as \( n \to \infty \), \[ \left[ \frac{P(V_i | \neg T)}{P(V_i | T)} \right]^{n} \to 0, \] and \( P(T | V^n) \to 1, \text{regardless of how low} P(T). \)

5 Converging virtue judgements

In the context of theory-choice, I suggest we interpret the above probabilities in the following way. Let \( P(V_i) \) stand for the probability of a scientist \( i \) deeming a theory \( T \) true on the basis of some virtue \( V \) of \( T \) and \( P(T) \) for the probability of \( T \) being true. Let \( P(V_i | T) \) then be the conditional probability that \( T \) would be correctly judged true on the basis of \( V \) by scientist \( i \), and \( P(V_i | \neg T) \) the conditional probability of \( T \) being incorrectly judged true on the basis of \( V \) by scientist \( i \). \( P(T | V^n) \) is then the posterior probability of \( T \) being true given that it was judged true on the basis of \( V \) by \( n \) scientists. If we now take into consideration that theories can be virtuous along different dimensions, and that scientists will judge a theory virtuous or not independently from how other scientists judge that theory, then:

\[
P(T | (E \cap C \cap S \cap U \cap F)^n) = \frac{1}{1 + \left[ \frac{1 - P(T)}{P(T)} \right] \times \left[ \frac{P(E_{i} \cap C_{i} \cap S_{i} \cap U_{i} \cap F_{i} | \neg T)}{P(E_{i} \cap C_{i} \cap S_{i} \cap U_{i} \cap F_{i} | T)} \right]^{n}}
\]

where \( E_{i}, C_{i}, S_{i}, U_{i}, F_{i} \) is a scientist \( i \)'s judgement about a theory being true based on that theory’s empirical accuracy, consistency, simplicity, and fertility, respectively, in line with the Kuhnian framework of theory choice. Here we allow that the individual probabilities of a theory being judged true on the basis of a virtue are all dependent on each other to some (unspecified) degree.
Now here is the crux: just like in the witness example, as \( n \to \infty \), \( P(T|V^n) \) will converge towards 1, \textit{regardless} of the value of \( P(T) \), so long as \( P(V_i|\sim T) < P(V_i|T) \). What this means is this: so long as scientists are relatively reliable in judging theories true on the basis of their virtues, the realist has good grounds to assert that a theory that is embraced by the scientific community on the basis of it being very virtuous (i.e., it possessing all five virtues) is very likely true, regardless of how low the actual base rate is. Contrary Magnus and Callender, the realist need not make any assumptions about the base rate being high.

We note that, on the Kuhnian framework, it is a precondition of the above argument that the theory in question possesses all five virtues. If that were not the case and the pool of theories to choose from offered only theories with some virtues but not others, then scientists, with their different weighting preferences of the virtues, would end up deeming different theories true: some might deem true theories that are simple (but not fertile), others those that are fertile (but not simple), etc. (cf. Section 4). There would be no converging judgements and thus no argument for any particular theory being likely to be true. Summarizing then (and relating back to Section 4), it would be a ‘miracle’ / unlikely, if a theory was very virtuous and deemed true by numerous scientists on the basis of its virtues, if it wasn’t true.

Of course, in any realistic scenario, the number of scientists \( n \) will not converge to infinity. Thus \( P(T) \) cannot be arbitrarily small. Yet \( P(T) \) may still be so small that the realist may argue that despite our ignorance about the precise value of the base rates, the chances are good that a very virtuous theory is true, if only a low \( P(T) \) is granted by the antirealist. It will not win the realist the argument against a very hard-headed antirealist, but it will make the realist’s argument much more unassuming. And \textit{some} small value for \( P(T) \) the antirealist must grant; otherwise, the Bayesian formalism is simply ill-defined.\(^{14} \) Whilst the realist won’t win the argument against a hard-headed antirealist, it may be argued that the realist is on firmer grounds rationally speaking. That is, although the antirealist can of course always insist that, however low \( P(T) \) as determined by the formalism for \( n \), her \( P(T) \) will be lower, this would seem badly \textit{ad hoc}: the antirealist simply assumes that value for \( P(T) \) which suffices to rebut the realist argument.

Let’s turn to a different aspect. It is of course true that scientists’ judgements about a theory being true may be construed deflationary, namely as true with regard to the theory’s empirical consequences, and not necessarily with regard to the unobservables postulated by the theory. The latter would then amount to a commitment only to a theory’s empirical adequacy, understood as the truth of a theory’s observable consequences in past, present,

\(^{14} \) Even more importantly, perhaps, setting \( P(T) \) to zero is no live option for the antirealist anyway, for the antirealist is a skeptic, not a dogmatist. All that she claims is that we have no sufficient grounds for believing that they are true, and that we furthermore need not commit ourselves to our theories being true in order to interpret science correctly (van Fraassen 1980). What she does not claim is that our best theories are indeed, or even must be, false.
and future (van Fraassen 1980). All that we will require in for our argument, however, is that scientists embrace the theory in question in some epistemic sense. That is, we require that scientists judge a theory true with regard to observables and unobservables or with regard to observables only. In the next section we will see that we can constrain the range of $P(V_i)$, whether construed instrumentally or in realist terms, on independent grounds.

Above we assumed that judgements about a theory being true based on the virtues have some effect on each other; they are not independent. And yet, a significant amount of independence seems not too implausible. For example fertility, construed as novel success, i.e., a theory’s successful predictions of novel phenomena, seems independent of empirical accuracy: for two theories may both be empirically accurate but one theory may have mere accommodative success whereas the other may have some novel success, presumably resulting in the latter being better confirmed than the former (but of course, no novel success without empirical accuracy). Or take unifying power. It is one thing to accommodate as many phenomena as possible, and accordingly to increase empirical accuracy. It is quite another thing, though, to account for a variety of phenomena on the basis of some basic fundamental principles, rather than merely conjoining them. So theoretical virtues, and accordingly the judgements about them, may be more independent than presumed here. Incidentally, for Kuhn’s aforementioned statement to be true that the virtues “repeatedly prove to conflict with one other” (cf. Section 2), it must be the case that the virtues are independent at least to some degree. But we shall not assume that they are entirely independent either; that will make our argument stronger.

For our argument to be successful, it is of course crucial that scientists make their judgements about a theory’s truth on the basis of some virtue independently from other scientists. Is this plausible? And would the Kuhnian picture not support a view according to which there is strong mutual influence between scientists, within a particular paradigm? What Kuhn would be happy to say, and what I think carries some plausibility, is that there is one important way in which scientists’ decision making is very strongly influenced by their peers: throughout the history of science scientists have chosen (and will choose in the future?) theories on the basis of the five standard theoretical virtues. On the other hand, Kuhn does concede a fair deal of independence to scientists when it comes to their preference regarding the virtues. Kuhn says that “when scientists must choose between competing theories, two men fully committed to the same list of criteria for choice may nevertheless reach different conclusions.” (359). Indeed he thought that scientists’ divergent weighting preferences would go some way to guarantee a right balance between conservatism and innovation (363). Since I’m operating within the Kuhnian framework in this article, I will simply adopt this ‘conditional autonomy’ of decision making, as one may call it.

One may object that virtue judgements, contrary to what I assumed above, should be sensitive to degrees of ‘virtuosity’. That is, in reality, theories are not just simple or not
simple, empirically accurate or not, unifying or conjunctive, etc., but rather more or less simple, empirically accurate etc. But we can happily admit that our assumption that judgements are binary is indeed an idealization. Reality is regularly more complex than our representations of it. I think it would be foolish, however, to conclude that our argument for this reason alone is therefore meaningless. That would make large parts of science, in which idealization looms large, meaningless too. The crux is that scientists are judging theories true on the basis of the theory’s virtues. And that we represent in our formalism. The judgements are likely to be more fine-grained than we suppose here, but that, I think, is secondary.

Lastly, one may fear that the argument proposed here on the basis of the Kuhnian framework of theory-choice invites relativism. My argument suggests that we be realists about theories that are held to be true by scientists on the basis of the theory’s virtues. Does that not subject us to the risk that scientists, at some point in time, embrace a theory as true that turns out false in the end? More importantly, does that not render realism relative to a social group? My answer to both questions is ‘yes’. Yet relativism does not follow. First of all, I, like most philosophers, subscribe to fallibilism. That is, we should never think that we possess the ultimately true theory that will remain with us forever. There is always the possibility that nature will teach us better. And for that, my argument clearly allows: it’s a probabilistic argument. A theory embraced by scientists on the basis of it being very virtuous makes it very likely that the theory is true. It doesn’t guarantee it. Second, although my argument is relative to scientists’ judgements, it is so in an unproblematic way. For one, the judgements are, as I explained in Section 3, grounded in the actual virtues of the theories. So there clearly is an objective basis for these judgements. For another, my account puts constraints on the nature of these judgements. It is not any community’s judgements that can serve as basis for my argument. The judgements need to be rationally constrained ones. In the next section we will see to what those amount to.

6 Error rates

The probabilities \( P(V_i|\neg T) \) and \( P(V_i|T) \) in our formalism, i.e., the probabilities of a scientist \( i \) correctly (or incorrectly) judging a theory \( T \) true on the basis of virtue \( V \), indicate the error rates. Error rates are the rates of false positives and false negatives, also known as type I and type II error rates. There is an interesting relationship between false positives \( P(V_i|\neg T) \) and true negatives \( P(\neg V_i|\neg T) \) on the one hand, and false negatives \( P(\neg V_i|T) \) and true positives \( P(V_i|T) \) on the other hand. A high false positive rate implies a low true negative rate, and vice versa (since \( P(V_i|\neg T) = 1 - P(\neg V_i|\neg T) \)). A high false negative rate implies a high true positive rate, and vice versa (since \( P(\neg V_i|T) = 1 - P(V_i|T) \)).\(^{15}\) We shall interpret \( P(\neg V_i|\neg T) \) as the probability for a scientist \( i \) to correctly judge a theory \( T \) to be false on the basis of it not

\(^{15}\) \( P(V_i|T) \) is also known as the sensitivity, and \( P(\neg V_i|\neg T) \) as the specificity of a test.
being virtuous, and \(P(\neg V_i|T)\) as the probability for scientist \(i\) to correctly judge false a theory \(T\) on the basis of it not being virtuous. In the following we will exploit these relationships.

For virtue judgements to be absolutely reliable indicators for a theory being true, the error rates of false positives \(P(V_i|\neg T)\) and false negatives \(P(\neg V_i|T)\) would have to be both low. But, as we saw in the last section, the error rates need not be low for our Earman-type argument to go through. All that is required is that they be relatively reliable, i.e., that the true positive rate is bigger than the false positive rate.\(^{16}\) Whether this is plausible to assume will be explored in the following. We should just also add that although, in accord with the Kuhnian picture of theory choice, the probabilities with which a theory is judged to be true on the basis of some virtue \(V\) (i.e., \(P(V_i)\)) will differ from one scientist to another, there are certain general constraints on what kinds of judgements a rational scientist would make. More specifically, these constraints include, but are not restricted to, the interdependencies of false-positive/true-negative and the false-negative/true-positive rates described above.

Lastly, let us remind ourselves that, in accordance with the Kuhnian framework of theory-choice, we are assuming that theoretical virtues are real theory properties, not just subjective projections. We can therefore allow ourselves to consider the probabilities of a theory possessing / not possessing a virtue when true / false as a short-cut for gauging the range of probabilities with which a rational scientist would embrace a theory on the basis of a theory’s virtues.

First consider empirical accuracy. If a theory is not empirically accurate, it presumably is no candidate for being true: empirical accuracy is a necessary condition for truth. Scientists therefore will, by necessity, judge a theory as empirically accurate, if it is true. Thus \(P(E_i|T)\) must be 1. On the other hand, empirically accurate theories may of course simply save the phenomena without being true. Thus, \(P(E_i|\neg T)\), the false positive rate, could in principle also be 1. However as a matter of fact it cannot, for if \(P(E_i|\neg T) = 1\), then \(P(\neg E_i|\neg T)\), i.e., the probability of a theory being correctly judged false on the basis of its not being empirically accurate would be zero, because \(P(\neg E_i|\neg T) = 1 - P(E_i|\neg T)\) (see above). But that is extremely implausible. Indeed, a lack of empirical accuracy is probably the best criterion scientists can go by when judging a theory to be false. Thus, \(P(\neg E_i|\neg T)\) must have some positive value and \(P(E_i|\neg T)\) accordingly a value lower than 1. Since, as determined above, \(P(E_i|T) = 1\) and \(P(E_i|\neg T) < 1\), the latter is lower than the former, in accordance with they Earman-type argument constraint for realism.

Next simplicity. Van Fraassen (1980) and others have argued that a theory’s being simple gives us no grounds for thinking that the theory is true or likely to be true. Although this is generally accepted, I have provided reasons for the contrary view in the previous chapter. If we were to assume that there were indeed no good grounds for simplicity being

\(^{16}\) Note, again, that the error rate of false negatives is simply the inverse of the true positive rate.
truth-indicative, it would not be the case that $P(S_i|T) > P(S_i|\neg T)$, i.e., it would not be the case that the probability of a scientist $i$ correctly judging theory $T$ to be true on the basis of it being simple could be larger than the probability of $i$ falsely judging theory $T$ to be true on the basis of it being simple. With the arguments provided in the previous chapter, however, we may very well say that $P(S_i|T) > P(S_i|\neg T)$.

Let us now turn to unifying power. Theories that have been singled out as approximately true by the realists generally manage to unify the phenomena. Einstein’s theory of relativity, the standard model of particle physics, the modern synthesis in evolutionary biology, plate tectonics etc. are cases in point. Thus a true theory, for the realist, is likely to unify and therefore $P(U_i|T)$, i.e., the probability of a theory being correctly judged true on the basis of it unifying the phenomena should approach 1. Of course, the antirealist may want to doubt that a theory is likely to unify if true, i.e., she may want to insist that $P(\neg U_i|T)$ is low, which would, according to the error-rate relations mentioned above, drive down the “realist’s” $P(U_i|T)$. But that would imply that it is likely that true theories will not be recognized as true by virtue of their unifying power. Since the realist holds high unifying power as a mark of truth, it may be that there are many more true theories out there that are unrecognized. But that wouldn’t be a problem for the realist. She only wants to make sure that she has good grounds for believing in the truth of those theories to which she does commit. The antirealist would therefore need to give us some reason for driving down $P(U_i|T)$ other than that it wins her the argument. I can’t see such an independent argument. The realist, on the other hand, seems to have an argument for $P(U_i|T)$ being high: it simply makes sense (not only from a historical perspective) that theories will unify the phenomena, if they are supposed to be genuine truth-candidates. With regards to the false positive rate, $P(U_i|\neg T)$, it seems glaringly obvious that false theories have only rarely, if ever, achieved the kinds of unification that have regularly been accomplished by theories like the ones above. It seems then safe to assume that $P(U_i|\neg T) < P(U_i|T)$, as required by our Earman-type argument.

Consider consistency next. Kuhn lumps together internal and external consistency. And yet they are best treated separately. Let’s start with internal consistency. Clearly, a true theory must be internally consistent (for short: $C_i$), i.e., $P(C_i|T)$ must be 1, which means that the false negative rate must be zero. On the other hand, the false positive rate $P(C_i|\neg T)$ is much more difficult to assess. There are probably indefinitely many false theories out there that are internally consistent. But there will also be as least as many false theories that are not even consistent. We also know that $P(\neg C_i|\neg T) = 1 - P(C_i|\neg T)$, so $P(C_i|\neg T)$ and $P(C_i|\neg T)$ cannot both be equal to 1, or be close to 1. In the face of our ignorance about the precise values, it seems most reasonable to set both expressions to .5. That would give us $P(C_i|\neg T) < P(C_i|T)$ in conformity with the requirement of our Earman-type argument.
Now consider external consistency ($C_{E_i}$). True theories must be consistent with our background knowledge (Psillos 1999, Lipton 1991/2004, Boyd 1983). Thus $P(C_{E_i}|T) = 1$ and accordingly the false negative rate $P(\neg C_{E_i}|T)$ would have to be zero. What about the false positive rate $P(C_{E_i}|\neg T)$, i.e., the probability that a theory is judged true on the basis of it being externally consistent if it is actually false? Although there are probably many false theories that are consistent with our background knowledge, there are at least as many (but probably many more) false theories that are inconsistent with our background knowledge, which would mean that $P(\neg C_{E_i}|\neg T)$ will approach 1. But since $P(C_{E_i}|\neg T) = 1 - P(\neg C_{E_i}|\neg T)$, $(P(C_{E_i}|\neg T)$ would then approach 0. In that case our condition of $P(C_{E_i}|\neg T) < P(C_{E_i}|T)$ seems to be well satisfied.

Lastly, consider fertility. As mentioned above, fertility is standardly construed in terms of novel success. And the capacity to successfully predict novel phenomena, as we noted above, is indeed the virtue most cherished by the realist. In fact, it is often treated as a necessary condition for a theory being true. Thus $P(F_i|T)$ should be equal to 1. Conversely, realists consider false theories very unlikely to generate successful novel predictions. $P(F_i|\neg T)$ thus ought to be close to zero. The condition for our Earman-type argument for realism is again well satisfied. If the antirealist wishes to challenge this, she would have to present evidence for false theories being capable of producing novel success. Magnus and Callender are right to stress that a handful of cases where false theories allegedly produced novel success won’t suffice to challenge the realist. Not only need there be a more substantial data base, but, as we noted above, the antirealist must also make a case for the theory’s false posits being responsible for novel success (Psillos 1999). Whether or not this has been achieved in even just a single case is still up for debate (cf. Section 2). In any case, the above reasoning requires us to set $P(F_i|\neg T)$ to a low level. And once again, this gives us $P(F_i|\neg T) < P(F_i|T)$ in accord with our requirement.

In sum then, it appears that for almost all theoretical virtues considered here (except perhaps simplicity), $P(V_i|\neg T) < P(V_i|T)$ holds true and therefore our argument against the antirealist succeeds: for $n \to \infty$, $\left[\frac{P(V_i|\neg T)}{P(V_i|T)}\right]^n \to 0$, and therefore $P(T|V^n) \to 1$, regardless of the value of $P(T)$. Less formally, so long as the false positive rate is smaller than the true positive rate, the more scientists embrace a theory on the basis of its virtues, the more likely it is that it is true, even when the base rates of true theories is small. As $P(V_i|T)$ need not be >.5 for this argument to go through, this means that the realist can assume that the theoretical virtues are only minimally truth-conducive (so that $P(V_i|\neg T) < P(V_i|T)$). In fact, we argued in this section that the theoretical virtues are truth-conducive in this minimal sense.
7 Conclusion

In this paper I argued, with the help of a point made by Earman in the case of converging witness reports, that a convergence of scientists’ truth judgements about a very virtuous theory will make it very likely that the theory is true, almost regardless of the value of the base rates. Although, for reasons to do with the finite amount of scientists embracing a theory at any particular time, this cannot be a blanket victory for the realist, it still helps to tilt the balance in favour of the realist. The only avenues left open to the antirealist are now either to cough up, in an ad hoc fashion, base rates that are designed to save her face, or she can try to argue with us about the error rates. In the latter case we would have ourselves a rational debate, in spite of what Magnus and Callender had feared. In the former case, the antirealist, although sure to win the argument, would win it arbitrarily.

References

Lewis, C I. 1946. An analysis of knowledge and valuation. LaSalle: Open Court.
Menke, Cornelis. 2013. "Does the miracle argument embody a base rate fallacy?" Studies in History and Philosophy of Science Part A.