

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

## Abstract

The perhaps strongest argument for scientific realism, the no-miracles-argument, has been said to commit the so-called base rate fallacy. The apparent elusiveness 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 perhaps 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.<sup>1</sup> 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. In particular, it

---

<sup>1</sup> See for example (Psillos 2009) and a reply by (Howson 2013). For another recent attempt see (Menke 2013).

will be argued 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.

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.<sup>2</sup> The view defended here may thus very well be detrimental to some of the views held by Kuhn.<sup>3</sup>

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 NO-VIRTUE-COINCIDENCE-ARGUMENT (NVC). Section 4 develops a formal apparatus for the NVC with the help of Earman's Bayesian rendering of the convergence of witness reports. Section 5 spells out this apparatus for the NVC in detail. Section 6 proposes a way to estimate the error rates regarding scientists' truth judgements. Section 7 addresses some possible objections to my argument and provides further clarification. Section 8 concludes the 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"

---

<sup>2</sup> In this sense, this paper is inspired by a recent paper by Okasha (2011).

<sup>3</sup> In particular, much of Kuhn's work is hardly reconcilable with realism, which will be defended here. In Section 3 I will outline in detail which parts of Kuhn's account I intend to use in this paper. Should any reader with a stake in the scholarship on Kuhn object to my interpretation, I invite them to consider my assumptions in abstract terms and to ignore any reference to Kuhn.

(322).<sup>4</sup> 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.<sup>5</sup> 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).<sup>6</sup> 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:

---

<sup>4</sup> Dicken (2013) has pointed out that retail arguments risk losing sight of the philosophical substance of the realism debate.

<sup>5</sup> Magnus and Callender's contribution can be seen as a synthesis of earlier points made by (Howson 2000) and (Lewis 2001).

<sup>6</sup>  $P(\neg T|e)$  and  $P(e|\neg T)$  are related by Bayes' theorem:  $P(\neg T|e) = \frac{P(e|\neg T) \times P(\neg T)}{P(e)}$ .

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).<sup>7</sup>

Although I think Magnus and Callender are correct in their diagnosis, the consequences of their insight can be alleviated and the second horn of their dilemma be rejected: for the no-miracles-argument to go through, the base rate of true theories need not be high.<sup>8</sup> In fact, they may even approach zero. In order to present the argument to this effect, we will first of all have to set up the theoretical framework in which I intend to make the argument.

### **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

---

<sup>7</sup> 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).

<sup>8</sup> See the first quotation above: “a *significant proportion* of our current theories [must be] true”. See also Magnus and Callender’s p. 325 (end of the second last paragraph).

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.<sup>9</sup>

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 weighting preferences concerning the standard theory choice criteria. Whereas some prefer simpler theories, for instance, others prefer more unified theories, and so on.<sup>10</sup> It is this subjective element that led Kuhn to the conclusion that there is ‘no neutral algorithm’ for theory choice to which all practitioners 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: theoretical 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 to qualitative parsimony when calling the theory ‘simple’. For example, one practitioner might judge the Copernican system of the planets simpler because it represents the retrogressive motion in simple terms, and another practitioner might judge the Ptolemaic system as simple as the Copernican system, because the

---

<sup>9</sup> 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*.

<sup>10</sup> 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).

Copernican system, too, made use of overall roughly the same number of epicycles (Kuhn 1957). 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).<sup>11</sup>

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 might be 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 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. Scientists with different preferences will end up choosing different theories. 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

---

<sup>11</sup> 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 of theoretical virtues 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.

facie, there is nothing intrinsic in the virtues or their relationships that would cause CONFLICT. There therefore should 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 prefer unified theories, then we will adopt different theories when there is no theory that has both of these properties. But 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 the NVC. As I will argue in the following, however, it is more powerful than the NMA: it offers an answer to 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 it being singled out as a hopeful truth-candidate. At the same time, scientists would have a hard time making their theory-choices when faced with a range of theories that would all score highly on all dimensions of theory choice. But very often, scientists do come to an agreement as to what theory to embrace. Although CONFLICT thus does seem plausible, it is of course an open question *how* frequent it is that only few theories or only one theory at a time possesses all the standard 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 Earman convergence

Although clearly related, the NVC 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 the NVC 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).<sup>12</sup> 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 of each other report the same murderer (i.e., without coordinating their beliefs)—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 there is convergence in the witness reports makes it unlikely that the witness reports are unreliable. 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 a crime  $T$  happened when that crime actually happened,  $P(V_i|\neg T)$  the probability that a witness reports a crime when the

---

<sup>12</sup> 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.

crime did not happen, and  $P(T)$  the prior probability of the crime itself.<sup>13</sup> Assuming that the witnesses are equally reliable and independent, the following equalities hold

$$P(V_1 \& \dots \& V_n) = P(V_1) \times P(V_2) \times \dots \times P(V_n) = P(V)^n$$

$$P(V_1 \& \dots \& V_n | T) = P(V_1 | T) \times P(V_2 | T) \times \dots \times P(V_n | T) = P(V | T)^n$$

$$P(V_1 \& \dots \& V_n | \neg T) = P(V_1 | \neg T) \times P(V_2 | \neg T) \times \dots \times P(V_n | \neg T) = P(V | \neg T)^n.$$

The posterior probability of the truth of a report given  $n$  witnesses making the same observations, by Bayes's theorem, is

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

which, assuming equally reliable witnesses, reduces to

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

Let us refer to this equation as the *Earman convergence equation*.

Earman points out that for the occurrence of a crime to be likely given all the witness reports, the witnesses need not be reliable in the absolute sense. All that is required, rather, is that the witness reports be reliable in the relative sense, so that  $P(V_i | \neg T) < P(V_i | T)$ , since in that case as  $n \rightarrow \infty$ ,  $\left[ \frac{P(V_i | \neg T)}{P(V_i | T)} \right]^n \rightarrow 0$ , and  $P(T | V^n) \rightarrow 1$ , 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 let  $P(T)$  stand 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  $V$

---

<sup>13</sup> Earman also conditionalises on the witnesses background knowledge and evidence, which I'll leave out here for the sake of simplicity. Earman also develops the argument with regards to the occurrence of miracles, instead of crimes.

by  $n$  scientists. If we now take into consideration that theories can be virtuous along different dimensions, 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. We will also assume that scientists' judgements about a theory being true on the basis of this theory's virtue are largely independent. I will say more about this in a moment. First consider the crux of the argument: just like in the witness example, for  $n \rightarrow \infty$ ,  $P(T|V^n)$ , i.e., the probability of a theory being true, will converge towards 1, *regardless* of the value of the base rate of true theories  $P(T)$ , so long as scientists are relatively reliable in judging a theory true based on its virtues, i.e., so long  $P(V_i|\neg T) < P(V_i|T)$ . With regards to Magnus and Callender's challenge, this means that the base rates need not be high for no-miracles arguments like the NVC to have any traction. What's more, if  $n$  would really approach infinity, the base rates could be neglected, as they will be diminishingly small.

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 *some* small value for  $P(T)$  the antirealist must grant; otherwise, the Bayesian formalism is simply ill-defined. We will consider further objections in Section 7.

We should note that presuming the Kuhnian framework of theory choice is relevant for making the NVC in that it must be possible for a theory to possess all five virtues. If this 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 might deem those theories true which are fertile (but not simple), etc. (cf. Section 4). There could

therefore be no converging judgements and thus no argument for any particular theory to be likely true.

There is another instance where we might have to appeal to the Kuhnian framework of theory choice for the NVC to be successful. Above we assumed that judgements about a theory being true based on the virtues are largely independent. I believe a significant amount of independence is in fact plausible. 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 to account for a variety of phenomena on the basis of some basic fundamental principles, rather than merely conjoining them. Regardless of these considerations, I think it's quite clear that on the Kuhnian picture of theory choice the virtues *have* to be largely independent, as otherwise they could not "repeatedly prove to conflict with one other", as Kuhn supposes (cf. Section 2). We can therefore treat the independence of scientists' judgements about a theory being true on the basis of its virtues as an assumption of the Kuhnian framework adopted here.

## 6 Estimating the error rates and relative truth-conduciveness

As we saw in the last section, for the NVC to go through, scientists' virtue judgements need not be absolutely reliable. They only need to be relatively reliable, that is, the true positive rate  $P(V_i|T)$  must only be larger than the false positive rate  $P(V_i|\neg T)$ , but those rates need not both be low. Is that the case?

In order to address this question, we shall exploit an interesting inverse relationship between the 'error rates': a high false positive rate implies a low true negative rate and vice versa, since  $P(V_i|\neg T) = 1 - P(\neg V_i|\neg T)$ , and a high true positive rate implies a low false negative rate and vice versa, since  $P(V_i|T) = 1 - P(\neg V_i|T)$ .<sup>14</sup> We shall interpret the true negative rate  $P(\neg V_i|\neg T)$  as the probability of a scientist  $i$  to *correctly* judge a theory  $T$  to be *false* on the basis of it not being virtuous, and the false negative rate  $P(\neg V_i|T)$  as the probability of scientist  $i$  to *correctly* judge

---

<sup>14</sup>  $P(V_i|T)$  is also known as the sensitivity, and  $P(\neg V_i|\neg T)$  as the specificity of a test.

*false* a theory  $T$  on the basis of it not being virtuous. In what follows, we shall use these relationships to estimate the error rates.

Although the probabilities scientists assign to a theory based on its virtues will of course differ from one individual to the other (in agreement with the Kuhnian picture of theory choice), an individual scientist judging a theory likely to be true when there is no good reason to (and vice versa), would amount to this scientist being rational. The error rates can thus be viewed as constraints on rational theory-choice decisions.

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. So if a theory is true, it would have to be empirically accurate, and accordingly be judged empirically accurate by scientists. 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 it 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, and we can therefore conclude that empirical accuracy is indeed relatively truth-conducive. That means that although empirical accuracy is by no means a guarantee for truth (after all, we're dealing here with probabilities), and although empirical accuracy must not make it even likely that a theory is true (i.e., it must not even be the case that  $P(E_i|T) > .5$ , although here it appears to be), we determined on the basis of the aforementioned inverse relationships of the error rates that it is still more likely that a theory is empirically accurate when it is true than when it is false. Importantly, this does not necessarily mean the converse, namely that a theory which is empirically accurate is likely to be true. This probability is the probability of the left hand side of our Earman equation derived in Section 4, namely  $P(T|V^n)$ , and is not to be confused with the error rate  $P(E_i|T)$ .

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 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. On the other hand, van Fraassen's arguments of course do not establish that false theories are bound to be simple (which would translate to  $P(S_i|\neg T) = 1$ ), or even that false theories are more likely to be simple than true theories (which would translate to  $\frac{P(S_i|T)}{P(S_i|\neg T)} < 1$ ). If 1 is not only the upper, but also the lower limit, then  $\frac{P(S_i|T)}{P(S_i|\neg T)} = 1$ , which would cancel it out of our *Earman equation* (cf. Section 5).

Despite first appearances, the realist might have reason for optimism with regards to simplicity. In the context of model selection it has recently been argued that simplicity is an epistemic theory property after all: simpler models are more likely to accurately predict data than more complex ones (Forster and Sober 1994, Hitchcock and Sober 2004). However, assessing the merits of this proposal and whether it applies also to theories goes beyond the purposes of the present paper.

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 the *Earman equation*.

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:  $C_{I_i}$ ), i.e.,  $P(C_{I_i}|T)$  must be 1, which means that the false negative rate must be zero. On the other hand, the false positive rate  $P(C_{I_i}|\neg T)$  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  $P(\neg C_{I_i}|\neg T) = 1 - P(C_{I_i}|\neg T)$ , so  $P(\neg C_{I_i}|\neg T)$  and  $P(C_{I_i}|\neg T)$  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  $P(C_{I_i}|\neg T) < P(C_{I_i}|T)$  in conformity with the requirement of the *Earman equation*.

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 Earman equation underlying our NVC 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 and that therefore our NVC, based on the *Earman convergence equation*, against the antirealist succeeds: for  $n \rightarrow \infty$ ,  $\left[\frac{P(V_i|\neg T)}{P(V_i|T)}\right]^n \rightarrow 0$ , and therefore  $P(T|V^n) \rightarrow 1$ , regardless of the value of  $P(T)$ . Less formally, since 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 very 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 indeed truth-conducive in this minimal sense.

## 7 Objections and clarifications

It should be obvious by now that the correct estimates of the error rates are crucial for the NVC argument for realism to have any force. Although we sought to set the error rates with a great deal of charity (consider e.g. simplicity), the antirealist might nevertheless have specific objections to how we set the error rates. Such objections are not necessarily bad news for the purposes of this paper. On the contrary, as this paper sought to undermine Magnus and Callender's claim that the realism debate is irrational, pointed objections against the way the error rates are set would actually support the conclusions of the current paper, as they would be evidence for a rational debate. And although it might be over-optimistic to expect that the realist and the antirealist will come to any agreements (as it is usually the case in philosophy), arguments about the error rates would be decidedly more tractable than arguments about the base rates.

It is worth recalling that we already established that the antirealist cannot undermine the NVC by simply setting the base rates to zero: that would render the formalism ill-defined. It is however widely accepted that the NMA is suitably expressed in Bayesian terms. There is another reason why the antirealist cannot help herself to such an escape: the antirealist is a skeptic, not a dogmatist. That is, antirealists think we have no good reason to believe that our theories are true, not that we have good reason to believe that our theories are false. However, setting the base rates to zero would mean insisting that virtuous theories are *bound to be* false.

There is a more severe problem for the argument presented in this paper. As mentioned above, the number of scientists  $n$  can of course not go to infinity, but is some finite number. Accordingly, there will be values for the base rate  $P(T)$  for which it will no longer be the case that it is likely that a theory is true when very virtuous, i.e., it will no longer be the case that  $P(T|V^n) > .5$ —as the realist would like to have it. The antirealist could thus simply adopt values for  $P(T)$ , which prevent the NVC to go through. It would seem that Magnus and Callender's observation would then still hold. At the very least, however, such a move by the antirealist would be utterly ad hoc, if not question-begging: the antirealist would have no other motivation for setting her base rates than to avoid the realist's argument to go through. The antirealist would thus damage the possibility of a rational debate about the error rates, when such debate is clearly available.

Another objection to the NVC might be to question whether the community of scientists would really be diverse enough for the individual scientists to make independent theory-choice decisions. Isn't it in fact the case that on the Kuhnian picture of theory choice, a scientific community is under the strong influence of a paradigm, ensuring concord amongst the scientists' decisions? First of all, in Kuhn's view, scientists are under the influence of a paradigm only in periods of normal science where theory-choice is not relevant. Only in revolutionary periods, scientists must make theory choices. In these periods, there clearly is diversity of judgements in Kuhn's view. Second, as pointed out above, Kuhn thought that scientists were indeed very independent with regards to their weighting preferences of the virtues in theory choice. Indeed he thought that scientists' divergent weighting preferences would go some way to guarantee a right balance between conservatism and innovation (Kuhn 1977, 363).

Let's turn to a different aspect. The NVC assumes that scientists make judgements about the *truth* of a theory on the basis of its virtues. But why should we presume that scientists make such judgements rather than merely acceptance

judgements, which do not require any commitment from them regarding the truth of a theory? On the contrary, is it not more plausible that scientists normally do not make such commitments? In fact, I'm happy to say that scientists need not have any commitments about the truth of a theory, so I'm happy to accept that what I referred to as "scientists' judgements about the truth of a theory" *can* be construed minimally so that they amount to no more than judgements about empirical adequacy in van Fraassen's sense, that is, a theory's truth concerning observables in the past, present, and future (van Fraassen 1980). I therefore want to keep apart the truth in the context of scientists' judgements, i.e.,  $P(V_i)$ , from the truth of the theory  $P(T)$ . So a scientist might judge a theory only empirically adequate (and not true) on the basis of that theory's virtue, but still we might, as we did, take this judgement as evidence for the truth of a theory, namely  $P(T|V^n)$ . More fundamentally, it might still be justified to think that a theory is likely to have certain virtues if true (and accordingly, be judged true (in a deflationary sense), i.e.,  $P(V_i|T)$ ), as we argued in Section 6.

A further objection might be that scientists' judgements about a theory's truth based on its virtues, contrary to what I assumed above, should be sensitive to *degrees* of 'virtueness'. 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. Although the judgements the NVC is based on are perhaps more fine-grained than supposed here, I can't see that this would undermine the basic argument.

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, as I argued in Section 6, my account puts constraints on the nature of these judgements. It is therefore not *any* community's judgements that can serve as basis for my argument. Rather, the judgements are rationally constrained ones.

## 8 Conclusion

In this paper I argued, with the help of a point made by Earman in the case of converging witness reports and by presuming the Kuhnian framework of theory choice, that a convergence of scientists' judgements about a very virtuous theory being true on the basis of its virtues will make it very likely that the theory is true, almost regardless 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, my argument still tilts the balance in favour of the realist.

## References

- Boyd, Richard N. 1983. "On the current status of the issue of scientific realism." *Erkenntnis* 19 (1-3):45-90.
- Cartwright, Nancy. 1983. *How the laws of physics lie*. Oxford: Oxford University Press.
- Chalmers, A. 2011. "Drawing philosophical lessons from Perrin's experiments on Brownian motion: A response to van Fraassen." *The British Journal for the Philosophy of Science* 62 (4):711-732.
- Dicken, Paul. 2013. "Normativity, the base-rate fallacy, and some problems for retail realism." *Studies in History and Philosophy of Science Part A* 44 (4):563-570.
- Earman, John. 2000. *Hume's Abject Failure: The Argument Against Miracles*. Oxford: Oxford University Press.
- Forster, Malcolm, and Elliott Sober. 1994. "How to tell when simpler, more unified, or less ad hoc theories will provide more accurate predictions." *The British Journal for the Philosophy of Science* 45 (1):1-35.
- Hacking, I. 1983. *Representing and intervening*: Cambridge University Press.
- Hitchcock, Christopher, and Elliott Sober. 2004. "Prediction versus accommodation and the risk of overfitting." *The British journal for the philosophy of science* 55 (1):1-34.
- Howson, Collin. 2000. *Hume's Problem: Induction and the Justification of Belief*. Clarendon Press.

- Howson, Collin. 2013. "Exhuming the No-Miracles Argument." *Analysis* 73 (2):205-211.
- Hudson, Robert G. 2013. *Seeing Things: The Philosophy of Reliable Observation*. Oxford: Oxford University Press.
- Ivanova, Milena. 2010. "Pierre Duhem's good sense as a guide to theory choice." *Studies in History and Philosophy of Science Part A* 41 (1):58-64.
- Kuhn, Thomas S. 1957. *The Copernican revolution: planetary astronomy in the development of western thought*. Harvard: Harvard University Press.
- Kuhn, Thomas S. 1962/1996. *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.
- Kuhn, Thomas S. 1977. "Objectivity, Value Judgment, and Theory Choice." In *The Essential Tension*, 320-333. Chicago: University of Chicago Press.
- Lewis, C I. 1946. *An analysis of knowledge and valuation*. LaSalle: Open Court.
- Lewis, Peter J. 2001. "Why the pessimistic induction is a fallacy." *Synthese* 129 (3):371-380.
- Lipton, Peter. 1991/2004. *Inference to the best explanation*. London: Routledge.
- Magnus, PD, and Craig Callender. 2004. "Realist Ennui and the Base Rate Fallacy." *Philosophy of Science* 71 (3):320-338.
- Menke, Cornelis. 2013. "Does the miracle argument embody a base rate fallacy?" *Studies in History and Philosophy of Science Part A*.
- Okasha, Samir. 2011. "Theory choice and social choice: Kuhn versus Arrow." *Mind* 120 (477):83-115.
- Psillos, Stathis. 1999. *Scientific realism: How science tracks truth*. London: Routledge.
- Psillos, Stathis. 2009. *Knowing the structure of nature: Essays on realism and explanation*: Palgrave Macmillan.
- Psillos, Stathis. 2011. "Making contact with molecules: On Perrin and Achinstein." *Philosophy of Science Matters: The Philosophy of Peter Achinstein*:177.
- Salmon, Wesley. 1984. *Scientific Explanation and Causal Structure of the World*.
- Stump, David J. 2007. "Pierre Duhem's Virtue Epistemology." *Studies in History and Philosophy of Science* 18 (1):149-159.
- van Fraassen, Bas. 1980. *The scientific image*. Oxford: Oxford University Press.
- van Fraassen, Bas. 2009. "The perils of Perrin, in the hands of philosophers." *Philosophical Studies* 143 (1):5-24.