The No Miracles Argument
without the Base Rate Fallacy

Richard Dawid[*]   Stephan Hartmann[†]

April 6, 2017

Abstract

According to an argument by Colin Howson, the no-miracles argument (NMA) is contingent on committing the base-rate fallacy and is therefore bound to fail. We demonstrate that Howson's argument only applies to one of two versions of the NMA. The other version, which resembles the form in which the argument was initially presented by Putnam and Boyd, remains unaffected by his line of reasoning. We provide a formal reconstruction of that version of the NMA and show that it is valid. Finally, we demonstrate that the use of subjective priors is consistent with the realist implication of the NMA and show that a core worry with respect to the suggested form of the NMA can be dispelled.

1 Introduction

The No Miracles Argument (NMA) is arguably the most influential argument in favour of scientific realism. First formulated under this name in Putnam (1975), the NMA asserts that the predictive success of science would be a miracle if predictively successful scientific theories were not (at least) approximately true. The NMA may be framed as a three step argument. First, it is asserted that the predictive success we witness in science does not have any satisfactory explanation in the absence of a realist interpretation of scientific theory. Predictive success is often understood (see e.g. Musgrave 1988) in terms of the novel predictive success of science. It is then argued that a realist interpretation of scientific theory can provide an explanation of scientific success. Finally, abductive reasoning is deployed to conclude that, given the first two points, scientific realism is probably true.

---
[*]Department of Philosophy, Stockholm University, 10691 Stockholm (Sweden) – richard.dawid@philosophy.su.se

[†]Munich Center for Mathematical Philosophy, LMU Munich, 80539 Munich (Germany) – s.hartmann@lmu.de.

All three steps of NMA were questioned already shortly after its formulation. It was argued that scientific success needs no explanation beyond what can be given in an empiricist framework (van Fraassen 1980), that scientific realism cannot provide the kind of explanation of predictive success aimed at by the realist (Laudan 1981) and that the use of abductive reasoning in a philosophical context already presumes a realist point of view (Fine 1986). The debate on all these points continues until this day.

In the year 2000, Colin Howson presented an interesting new line of criticism (Howson 2000) that did not look at one of the three individual steps of the NMA but questioned the overall logical validity of the argument. Howson argued that a Bayesian reconstruction of the NMA revealed a logical flaw in the argument's structure: it commits the base-rate fallacy. Howson's line of reasoning was followed and extended in Magnus and Callender (2003).

Attempts to counter Howson's argument have been of two kinds. Some defenders of the NMA have generally disputed the adequacy of a Bayesian perspective on the NMA. Worrall (2007) has argued that the abductive character of the NMA cannot be accounted for by a Bayesian reconstruction. In a similar vein, Psillos (2009) has argued that the subjective character of priors is at variance with the spirit of realist reasoning. Therefore, according to Psillos, critically analysing a realist argument in a Bayesian framework that involves prior probabilities amounts to begging the question by choosing an anti-realist point of view from the start. Howson (2013) highlighted an inherent problem of this argumentative strategy. In direct response to Psillos (2009) he pointed out that (i) a Bayesian perspective can in principle be based on an objective concept of prior probabilities, which is not anti-realist on any account, and (ii) rejecting prior probabilities amounts to rejecting any probabilistic understanding of a commitment to scientific realism. In the absence of any reference to probabilities, however, it becomes difficult to explain what is meant by the statement that one believes in scientific realism.

A second strategy has been to acknowledge the viability of Howson's analysis in principle but point to the limits of its range of applicability. As first emphasised in Dawid (2008), the NMA can be formulated in two substantially different forms. Menke (2014) and Henderson (forthcoming) have pointed out that Howson's base rate fallacy charge only applies to one of them but not to the other.[1]

In this article, we formally spell out the latter position and demonstrate that the analysis of the NMA can gain substantially from such a full formal reconstruction. We start with a rehearsal of Howson's argument in Section 2. Section 3 then specifies two distinct kinds of the NMA which we will call *individual theory-based NMA* and *frequency-based NMA*. In Section 4, we

---

[1]A different idea how to enrich the NMA in order to avoid the base rate fallacy has been presented in Sprenger (2015).

present a formalisation of frequency-based NMA that demonstrates (i) that the individual theory-based NMA is a sub-argument of the frequency-based NMA, (ii) that Howson only reconstructs this sub-argument, and (iii) that the full frequency-based NMA does not commit the base rate fallacy. In Section 5, we argue that, while there is a strand in scientific realism that reduces the NMA to the individual theory-based NMA, Putnam and Boyd, the first main exponents of the argument, clearly endorsed the frequency-based NMA. The final parts of this article utilize the full formal reconstruction of the NMA to clarify two important points on the nature of the NMA. In Section 6, it is explained why the use of subjective priors poses no problem for the frequency based NMA as a realist argument. Section 7 formally demonstrates that a core worry with respect to a frequency-based understanding of the NMA can be dispelled. Section 8 discusses a lottery analogy to support our argument. Finally, Section 9 sums up the main results of this article.

2   Howson's Argument

In his book Howson (2000), Colin Howson makes the remarkable claim that the NMA commits the base-rate fallacy and therefore is invalid on logical grounds. Howson provides the following Bayesian formalisation of this argument: Let S be a binary propositional variable with values S: Hypothesis H is predictively successful, and its negation ¬S. Let T be the binary propositional variable with values T: H is approximately true, and its negation ¬T. Next, the adherent of the NMA makes two assumptions:

A$_1$: $P(\text{S}|\text{T})$ is quite large.

A$_2$: $P(\text{S}|\neg\text{T}) < k \ll 1$.

A$_1$ states that approximately true theories are typically predictively successful, and **A$_2$** states that theories that are not at least approximately true are typically not predictively successful. Note that ¬T is the so-called *catch-all hypothesis*. For the sake of the argument, we consider both assumptions to be uncontroversial, although anti-realists objected to both of them (see, e.g., Laudan (1981) and Stanford (2006)). The adherent of the NMA then infers:

C: $P(\text{T}|\text{S})$ is large.

As Howson correctly points out, the stated argument commits the base-rate fallacy: **C** is only justified if the prior probability $P(\text{T})$ is sufficiently large as, by Bayes' Theorem,

$$P(\text{T}|\text{S}) = \frac{P(\text{S}|\text{T})\,P(\text{T})}{P(\text{S}|\text{T})\,P(\text{T}) + P(\text{S}|\neg\text{T})\,P(\neg\text{T})}, \tag{1}$$

according to which $P(\mathrm{T|S})$ is strictly monotonically increasing as a function of $P(\mathrm{T})$ with $P(\mathrm{T|S}) = 0$ for $P(\mathrm{T}) = 0$. However, the condition that $P(\mathrm{T})$ is sufficiently large is not in the set of assumptions $\{\mathbf{A_1}, \mathbf{A_2}\}$. If it were, we would beg the question because the truth of a predictively successful theory would then be derived from the assumption that the truth of the theory in question is a priori sufficiently probable, which is exactly what an anti-realist denies.[2]

3 Two Versions of the NMA

Does Howson's formal reconstruction constitute a faithful representation of the NMA? It has been pointed out in Dawid (2008) that two conceptually distinct versions of the NMA have to be distinguished. They differ with respect to the issue as to what exactly has to be explained by the realist conjecture.[3]

The first version is the following: We consider one specific predictively successful scientific theory. The approximate truth of that theory is then deployed for explaining why it is predictively successful. We shall call a NMA based on individual predictive success an *individual theory-based NMA*. According to the second version, what is to be explained by the realist conjecture is the observation that theories which are developed and held by scientists tend to be predictively successful. In this version, the NMA does not rely on the observation that one specific theory is predictively successful but rather on an observed characteristic of science as a whole or at least on a more narrowly specified segment of science. Theories that are part of that segment, such as theories that are part of mature science or that are part of a specific mature research field, are expected to be predictively successful. We will call a NMA based on the frequency of predictive success *frequency-based NMA*.[4] Menke (2014) and Henderson (forthcoming) have

---

[2]Howson (2015) indeed calls the statement 'one should endorse the truth of an empirically confirmed theory if one believes $P(\mathrm{T}) > P(\mathrm{S|\neg T})$' the only valid conclusion from NMA-like reasoning. However, since no reason for believing the stated relation is given, this form of "NMA" cannot seriously be called an argument for scientific realism. As an aside, we note that Howson's point is false: $P(\mathrm{T}) > P(\mathrm{S|\neg T})$ and the fact that $P(\mathrm{T|S})$ is strictly monotonically increasing as a function of $P(\mathrm{T})$ imply that $P(\mathrm{T|S}) > P(\mathrm{S|T})/\left[P(\mathrm{S|T}) + P(\neg \mathrm{S|\neg T})\right]$. Hence, $P(\mathrm{T|S}) > 1/2$ holds only if one additionally requires that $P(\mathrm{S|T}) + P(\mathrm{S|\neg T}) > 1$. This inequality does not follow from the set of assumptions $\{\mathbf{A_1}, \mathbf{A_2}\}$.

[3]In Dawid (2008), the two forms of the NMA are called "analytic NMA" and "epistemic NMA".

[4]A related but different distinction between two forms of NMA was made in Barnes (2003). Barnes calls the argument from a theory's success the "miraculous theory argument" and contrasts it with the "miraculous choice argument" from the scientists' actual development and choice of successful theories. The miraculous theory argument is necessarily an individual theory-based NMA, since it presumes the predictive success of the individual theory under consideration. Therefore, no frequency of predictive success can

pointed out that Howson's argument only addresses what we call individual theory-based NMA. The frequency-based version of the NMA is not adequately represented by Howson's reconstruction.

In the following section, we develop a Bayesian formalisation of frequency-based NMA and therefore provide the basis for a fully formalised analysis of the NMA. This will help us to investigate how and to what extent the frequency-based version of the NMA reaches beyond Howson's reconstruction.

4   Formalising the Frequency-Based NMA

To begin with, we specify a scientific discipline or research field $\mathcal{R}$.[5] We count all n_E theories in $\mathcal{R}$ that have been empirically tested and determine the number n_S of theories that were empirically successful. We can thus state the following observation:

O: n_S out of n_E theories in $\mathcal{R}$ were predictively successful.

Now let us assume that we are confronted with a new and so far empirically untested theory H in $\mathcal{R}$. We want to extract the probability $P(\text{S}|\text{O})$ for the predictive success S of H given observation O.

We then assume that each new theory that comes up in $\mathcal{R}$ can be treated as a random pick with respect to predictive success. That is, we assume that there is a certain overall rate of predictively successful theories in $\mathcal{R}$ and that, in the absence of further knowledge, the success chances of a new theory should be estimated according to our best estimate R of that success rate:

$$P(\text{S}|\text{O}) = R \qquad (2)$$

R will be based on observation O. The most straightforward assessment of R is to use the frequentist information and to identify R with

$$R_{freq} = \frac{n_S}{n_E} \, . \qquad (3)$$

We will adopt it in the remainder.[6]

---

be specified within the framework of the miraculous theory argument. A miraculous choice argument may be either of the individual theory or of the frequency-based type.

[5]One might argue that a strong statement on predictive success in a research field $\mathcal{R}$ always requires the specification of some conditions $\mathcal{C}$ that separate more promising from less promising theories in the field. However, accounting for this additional step does not affect the basic structure of the argument, which is why, in order to keep things simple, we won't explicitly mention these conditions in our reconstruction.

[6]A fully Bayesian analysis that extracts R by updating on the individual instances of predictive success and predictive failure (and converges towards the frequentist result in the large n_E limit) is possible but requires some conceptual effort. We don't carry it out here in order to keep things simple.

To proceed with our formal analysis of the frequency based-NMA, we need to make two assumptions similar to A_1 and A_2:

$\mathbf{A}_1^O$: $P(\mathrm{S}|\mathrm{T},\mathrm{O})$ is quite large.

$\mathbf{A}_2^O$: $P(\mathrm{S}|\neg\mathrm{T},\mathrm{O}) < k \ll 1$

Note that scientific realists assume that the truth of H is the dominating element in explaining the theory's predictive success. If that is so, then S is roughly conditionally independent of O given T and we have $P(\mathrm{S}|\mathrm{T},\mathrm{O}) \approx P(\mathrm{S}|\mathrm{T})$ and $P(\mathrm{S}|\neg\mathrm{T},\mathrm{O}) \approx P(\mathrm{S}|\neg\mathrm{T})$. The conditions $\mathbf{A}_1^O$ and $\mathbf{A}_2^O$ can then be roughly equated with the conditions $\mathbf{A}_1$ and $\mathbf{A}_2$. For the sake of generality, we will nevertheless use conditions $\mathbf{A}_1^O$ and $\mathbf{A}_2^O$ in the following analysis.

Let us now come to our crucial point, viz. *to show that accounting for observation O blocks the base-rate fallacy.* The base-rate fallacy in individual theory-based NMA consisted in disregarding the possibility of arbitrarily small priors $P(\mathrm{T})$. In frequency-based NMA, however, the crucial probability is $P(\mathrm{T}|\mathrm{O})$ rather than $P(\mathrm{T})$. Updating the probability of S on observation O has an impact on $P(\mathrm{T}|\mathrm{O})$. To see this, we start with the law of total probability,

$$P(\mathrm{S}|\mathrm{O}) = P(\mathrm{S}|\mathrm{T},\mathrm{O})P(\mathrm{T}|\mathrm{O}) + P(\mathrm{S}|\neg\mathrm{T},\mathrm{O})P(\neg\mathrm{T}|\mathrm{O}), \tag{4}$$

and obtain

$$P(\mathrm{T}|\mathrm{O}) = \frac{P(\mathrm{S}|\mathrm{O}) - P(\mathrm{S}|\neg\mathrm{T},\mathrm{O})}{P(\mathrm{S}|\mathrm{T},\mathrm{O}) - P(\mathrm{S}|\neg\mathrm{T},\mathrm{O})} \, . \tag{5}$$

(The proof is in Appendix A.1.) Two points can be extracted from eq. (5) and assumptions $\mathbf{A}_1^O$ and $\mathbf{A}_2^O$. First, we conclude that

$$P(\mathrm{S}|\mathrm{T},\mathrm{O}) \geq P(\mathrm{S}|\mathrm{O}). \tag{6}$$

In particular,
$$P(\mathrm{S}|\mathrm{T},\mathrm{O}) = P(\mathrm{S}|\mathrm{O}) \quad \text{iff} \quad P(\mathrm{T}|\mathrm{O}) = 1. \tag{7}$$

Second, we obtain that

$$P(\mathrm{T}|\mathrm{O}) > P(\mathrm{S}|\mathrm{O}) - k = R - k \, . \tag{8}$$

(The proof is in Appendix A.2.) Hence, $P(\mathrm{T}|\mathrm{O})$ is bounded from below if $k < R$. Since k is very small by assumption $\mathbf{A}_2^O$, this puts only moderate constraints on $R \approx n_S/n_E$. (See also the discussion on the size of R in Section 7). If a supporter of the NMA analyses a research context where $k < R$ is satisfied, then the base-rate fallacy is avoided. Note that the first and crucial inference in frequency-based NMA is made before accounting for the predictive success of H itself and relies on relating $P(\mathrm{S}|\mathrm{O})$ to $P(\mathrm{T}|\mathrm{O})$ based on the law of total probability.

Once H has been empirically tested and found to be empirically successful, we can, just as in the case of individual theory-based NMA, update on S, the predictive success of H.[7] Using eqs. (6) and (7) and the identity

$$P(\text{T}|\text{S},\text{O}) = \frac{P(\text{S}|\text{T},\text{O})}{P(\text{S}|\text{O})} \cdot P(\text{T}|\text{O}), \tag{9}$$

it is easy to see that Bayesian updating from $P(\text{T}|\text{O})$ to $P(\text{T}|\text{S},\text{O})$ further increases the probability of T for predictively successful theories as long as $P(\text{T}|\text{O}) < 1$.

Comparing this formalization of frequency-based NMA with Howson's reconstruction, we see that Howson only reconstructs the second part of frequency-based NMA, which leads from $P(\text{T}|\text{O})$ to $P(\text{T}|\text{S},\text{O})$. He leaves out the part where $P(\text{T}|\text{O})$ is extracted from the observed success frequency n_S/n_E based on the prior $P(\text{T})$. Howson's base rate fallacy charge crucially relies on the understanding that the specification of the probability of theory H before updating under the predictive success of H is *not* part of the NMA. Howson's reconstruction thus is insensitive to the observed frequency of predictive success and amounts to a formalization of individual theory-based NMA.

We want to conclude this section with a clarificatory note on the status of the suggested formalisation. If a scientific realist asserts that trusted theories in a given research field (see footnote 5) show a significant frequency of predictive success, she thereby expresses her understanding that an overall appraisal of the field justifies such a claim. Her assessment is typically based on a blend of her knowledge of the way scientists perceive the reliability of trusted theories in their field and her own historical knowledge about that field. She normally is not in the situation to back up her claim by providing a complete count of the successful and failed theories in the field based on precise criteria for what counts as a theory. The presented formalisation does not suggest that scientific realists must provide an actual count of theories any more than Bayesian confirmation theory suggests that scientists must carry out Bayesian updating from explicitly specified priors of their theories. The role of a formalisation of the presented kind can only be to clarify the logical structure of a given line of reasoning by offering a well-defined formal model of it.

5 Do Realists Use Frequency-Based NMA?

The NMA has a long and chequered history. Who among its exponents endorsed individual theory-based NMA and who endorsed its frequency-

---

[7]Note that the predictive success of H will change the value of R. However, this change will be small if n_E is large.

based counterpart? Here is the precise wording of Hilary Putnam's famous first formulation of NMA:

> The positive argument for realism is that it is the only philosophy that doesn't make the success of science a miracle. That terms in mature science *typically* refer [...], that the theories in a mature science are *typically* true, that the same term can refer to the same thing even when it occurs in different theories – these statements are viewed by the scientific realist not as necessary truths but as the only scientific explanation of the *success of science* and hence as part of any adequate scientific description of science and its relations to its objects. (Putnam 1975, our emphasis)

Note that Putnam speaks of the success of science rather than of the success of an individual scientific theory. He clearly understands the *success of science* as a general and observable phenomenon. Since he obviously would not want to say that each and every scientific theory is always predictively successful, he thereby asserts that we find a high success rate n_S/n_E based on our observations of the history of (mature) science. He then infers from the success of science that mature scientific theories are *typically* approximately true. That is, he infers $P(\text{T}|\text{O})$ from the observed rate n_S/n_E. We conclude that Putnam presented a clear-cut frequency-based version of NMA.

The other early main exponent of the NMA, Richard Boyd (see, e.g., Boyd (1984) and Boyd (1985)), is committed to frequency-based NMA as well. Boyd emphasises that only what he calls the "predictive reliability of well-confirmed scientific theories" and the "reliability of scientific methodology in identifying predictively reliable theories" provides the basis for the NMA. In other words, an individual case of having predictive success could be explained by good luck and therefore would not license a NMA. Only the reliability of generating predictive success in a field, i.e. a high frequency of predictive success, provides an acceptable basis for the NMA.

Later expositions of NMA at times are not sensitive to the distinction between individual theory-based NMA and frequency-based NMA and therefore are not clearly committed to one or the other version of the argument. In some cases, Putnam's phrase "success of science" is used at one stage of the exposition while the thrust of the exposition seems to endorse an individual theory-based NMA perspective. A classic example of this kind is Musgrave (1988). Psillos (2009) defends individual theory-based NMA in his attempt to refute Howson's criticism of the NMA.

6 Subjective Priors and Scientific Realism

At this point we want to return to the issue of subjective priors. As described in the introduction, Psillos has argued that the reliance of Bayesian

confirmation theory on subjective priors renders it inherently anti-realist from the start. A similar point has been raised by Brian Skyrms (private communication) and arguably has generated doubts about the relevance of Howson's argument among a number of exponents of the scientific realism debate.

The core of the argument is the following. Bayesian confirmation relies on prior probabilities of the hypotheses under scrutiny. In other words, in a Bayesian framework even the most convincing set of empirical data leads to an endorsement of the tested hypothesis only for a given range of priors. This means that a (subjective) Bayesian analysis always allows for avoiding a given conclusion by choosing sufficiently low priors for the hypothesis in question. Why should this be a lethal problem for (individual theory-based) NMA when it isn't for scientific reasoning, which can be reconstructed in Bayesian terms as well?

In order to answer this objection to Howson's argument, one has to take a look at the reason why, from a Bayesian perspective, science can reach stable and more or less objective conclusions despite the necessity of starting from subjective priors. The reason is that, in standard scientific contexts, posteriors converge under repeated empirical testing. Therefore, if one wants to test a hypothesis, any prior probability that may be chosen in advance (apart from dogmatic denial, which corresponds to a prior probability of zero), can lead to posteriors beyond a probability threshold set for acknowledging conclusive confirmation if a sufficient amount of data is collected. The fact that science is modelled as allowing for an infinite series of empirical testing therefore neutralises the threat of subjective priors to the reliability of scientific claims.

Howson's claim that the NMA involves a base rate fallacy in this light amounts to the claim that the evidence structure that enters the NMA is not of the kind to be found in scientific testing. It does not project a series of updatings under a sequence of observations like in scientific testing. Rather, it relies on acknowledging exactly one characteristic of a given theory: its novel predictive success. In Howson's reconstruction of the NMA, there is only one updating under the observation of predictive success. This, Howson argues, is no adequate structure for establishing the hypothesis under scrutiny as long as no limits to the priors are assumed. Since no anti-realist would subscribe to such limits, the (individual theory-based) NMA cannot establish scientific realism.

Understood in this way, Howson makes a crucial point about the individual theory-based NMA. The problem of individual theory-based NMA is not that the sample size happens to be 1. The problem runs deeper: individual theory-based NMA does not provide a framework for the sample testing of a rate of predictive success at all. The logic of the argument starts with selecting a theory that is known to have made correct novel predictions. The argument is not based on an observation about the research process

that has led up to developing and endorsing scientific theories but only on an observation about the theory's relation to empirical data. One could face a situation where the sheer number of theories that were developed statistically implies a high probability that, among all those theories, one predictively successful can be found. To use an analogy, if one selects the winning ticket of a lottery after the draw, the fact that it won is no surprise. Since the individual theory based NMA selects the predictively successful theory ex post, it does not distinguish whether or not a situation of this kind applies. Therefore, it provides no basis for establishing whether the predictive success of the given theory is "a miracle" or, given that it was selected for its predictive success, is no surprise at all.

Frequency-based NMA, to the contrary, relies on an observation about the research process: within a given field, scientific theories that satisfy a given set of conditions happen to be predictively successful with a significantly high probability. This observation is grounded in a series of individual observations of the predictive successes of individual theories. The empirical testing of each theory serves as a new data point. The observation $\mathbf{O}$ of a certain frequency of predictive success therefore is the result of a quasi-scientific testing series. The probability $P(\mathrm{T}|\mathrm{O})$, which is extracted from $\mathbf{O}$ based on the law of total probability, can be decoupled from the prior probability $P(\mathrm{T})$ as it is inferred from a quasi-scientific testing series. The mechanism of objectifying results that work in a scientific framework is also applicable in the case of NMA: anyone who takes the available data to be inconclusive can resort to further testing.

Specifically, if a certain number of data points are available for extracting $\mathbf{O}$ and an anti-realist observer (who has a very low prior $P(\mathrm{T})$ before observing predictive success in the field but, for the sake of the argument, accepts conditions $\mathbf{A_1^O}$ and $\mathbf{A_2^O}$) considers the available amount of data too small for overcoming her very small prior probability $P(\mathrm{T})$, one can just wait for new theories in the research context to get tested. If the additional data also supports a high frequency of predictive success, our anti-realist would at some stage acknowledge that the empirical data has established realism.[8] The deep reason why frequency-based NMA avoids the base rate fallacy therefore lies in the fact that it provides a framework in which the convergence behaviour of posteriors under repeated updating can be exploited.

---

[8]The present analysis implies that the proposal by Menke (2014) to use NMA only with respect to theories that show multiple predictive successes is not satisfactory. Since Menke's suggestion remains within the framework of individual theory-based NMA, it does not solve the structural problem Howson is pointing to. A theory that happens to make two novel predictions and is successful in both cases may be more likely to be true than one with only one case of novel predictive success (albeit one may object to treating instances of novel predictive success statistically as independent picks as it is proposed by Menke). But individual theories just give us the predictions they happen to imply. There is no perspective of a series of novel predictions that can be understood in terms of an open series of empirical testing. Howson's core objection thus remains valid.

7 Does NMA need a high Frequency of Predictive Success?

References to success frequencies in the context of the NMA have been avoided by a number of philosophers of science for one reason: it seemed imprudent to ground an argument for realism on a claim that looked questionable. Given the many failures of scientific theories, it looked unconvincing to assert a high frequency of predictive success in any scientific field. In this light, it is important to have a clear understanding of the success frequencies that are actually required for having a convincing NMA. In the following, we address this question within our formalised reconstruction.

The strength of the NMA is expressed by $P(\mathrm{T}|\mathrm{S},\mathrm{O})$. In order to make a strong case for scientific realism, we demand that

$$P(\mathrm{T}|\mathrm{S},\mathrm{O}) > K, \tag{10}$$

where K is some reasonably high probability value. $K = 1/2$ may be viewed as a plausible condition for taking the NMA seriously. We now ask: How does a condition on $P(\mathrm{T}|\mathrm{S},\mathrm{O})$ translate into a condition on $P(\mathrm{S}|\mathrm{O})$ and therefore on the (observed) success frequency R? Obviously, this depends on the posited values of $P(\mathrm{S}|\mathrm{T},\mathrm{O})$ and $P(\mathrm{S}|\neg\mathrm{T},\mathrm{O})$. Eqs. (5) and (9) imply that

$$P(\mathrm{T}|\mathrm{S},\mathrm{O}) = \frac{P(\mathrm{S}|\mathrm{T},\mathrm{O})}{P(\mathrm{S}|\mathrm{O})} \cdot \left(\frac{P(\mathrm{S}|\mathrm{O}) - P(\mathrm{S}|\neg\mathrm{T},\mathrm{O})}{P(\mathrm{S}|\mathrm{T},\mathrm{O}) - P(\mathrm{S}|\neg\mathrm{T},\mathrm{O})} \right). \tag{11}$$

We know already that the NMA works only for $P(\mathrm{S}|\mathrm{O}) < P(\mathrm{S}|\mathrm{T},\mathrm{O})$. From eq. (11) we can further infer that, for fixed $P(\mathrm{S}|\mathrm{O})$ and fixed $P(\mathrm{S}|\neg\mathrm{T},\mathrm{O}) < P(\mathrm{S}|\mathrm{O})$, $P(\mathrm{T}|\mathrm{S},\mathrm{O})$ decreases with increasing $P(\mathrm{S}|\mathrm{T},\mathrm{O})$. (The proof is in Appendix A.3.) Therefore, it is most difficult for $P(\mathrm{T}|\mathrm{S},\mathrm{O})$ to reach the value K if $P(\mathrm{S}|\mathrm{T},\mathrm{O})$ has the maximal value 1. It thus makes sense to focus on this case, which gives

$$P(\mathrm{T}|\mathrm{S},\mathrm{O}) = \frac{1}{P(\mathrm{S}|\mathrm{O})} \cdot \left(\frac{P(\mathrm{S}|\mathrm{O}) - P(\mathrm{S}|\neg\mathrm{T},\mathrm{O})}{1 - P(\mathrm{S}|\neg\mathrm{T},\mathrm{O})} \right). \tag{12}$$

Condition (10) then turns into

$$\frac{1}{P(\mathrm{S}|\mathrm{O})} \cdot \left(\frac{P(\mathrm{S},\mathrm{O}) - P(\mathrm{S}|\neg\mathrm{T},\mathrm{O})}{1 - P(\mathrm{S}|\neg\mathrm{T},\mathrm{O})} \right) > K, \tag{13}$$

which leads to the following condition for $P(\mathrm{S}|\mathrm{O})$:

$$P(\mathrm{S}|\mathrm{O}) > \frac{P(\mathrm{S}|\neg\mathrm{T},\mathrm{O})}{1 - K + K \cdot P(\mathrm{S}|\neg\mathrm{T},\mathrm{O})} \tag{14}$$

(The proof is in Appendix A.4.) For small $P(\text{S}|\neg\text{T}, \text{O})$, a good approximation of the condition expressed in eq. (14) is

$$P(\text{S}|\text{O}) > \frac{P(\text{S}|\neg\text{T}, \text{O})}{1 - K}. \qquad (15)$$

For $K = 1/2$, we thus obtain

$$P(\text{S}|\text{O}) > 2\, P(\text{S}|\neg\text{T}, \text{O}). \qquad (16)$$

In the large n_E-limit, this implies

$$n_S/n_E > 2\, P(\text{S}|\neg\text{T}, \text{O}). \qquad (17)$$

We thus see that we don't need a high rate of predictive success in a scientific field for having a significant argument in favour of scientific realism. The ratio n_S/n_E may be small as long as it is larger than the assumed value of $P(\text{S}|\neg\text{T}, \text{O})$. In a sense, the understanding that the NMA needs a high frequency of predictive success is based on the inverse mistake to the one committed by the endorser of the individual theory-based NMA. While the latter only focuses on the updating under the novel predictive success of an individual theory, the former does not take this updating into account.

The possibility to base a NMA on a small ratio n_S/n_E is significant. Many arguments and observations which lower our assessments of n_S/n_E in a scientific field at the same time enter our assessment of $P(\text{S}|\neg\text{T}, \text{O})$. Take, for example, the following argument against a high ratio of predictive success. "Scientists develop many theories which they don't find promising themselves. Clearly, those theories have a very low frequency of predictive success." It is equally clear, though, that most of these theories are false and therefore lower $P(\text{S}|\neg\text{T}, \text{O})$. Examples of this kind show that lines of reasoning that isolate segments of the theory space where both n_S/n_E and $P(\text{S}|\neg\text{T}, \text{O})$ are very small provide a framework in which the claim $n_S/n_E > P(\text{S}|\neg\text{T}, \text{O})$ can make sense even if the overall value n_S/n_E in a research field is rather unimpressive.

8   An Analogy

In conjunction, the analysis of Section 6 and the possibility to base a frequency-based NMA on low success frequencies have an important consequence: lines of reasoning that at first glance look like exemplifications of individual theory-based NMA may allow for a frequency based understanding and can be made valid on that basis. In the following, we discuss this case in detail by spelling out the analogy between the NMA and the lottery case that was already mentioned in Section 6.

Let us look at the case of a lottery with one draw and 10^6 numbered tickets. Each individual participant i has a fair winning chance of $P(\text{W}_i) = 10^{-6}$.

Let us assume that the lottery had the outcome W_I that participant I got the winning ticket. Let us further assume that some observers develop the hypothesis T that the lottery is rigged and an employee has given away the winning number to someone in advance. One of those observers addresses person I and asks him: "the probability that you would win without fraud was 10^{-6}. What other explanation than fraud could explain your win?"

The correct formal Bayesian response to that observer is: "You fell prey to the base rate fallacy. Since you have not specified the prior probability of fraud, your argument is invalid." Another way to respond is: "You did not even check whether the outcome of the lottery was unlikely. In order to do so you have to look at the overall winning rate for sold tickets. If all 10^6 tickets were sold that rate is 10^{-6} and there is nothing unusual about the process whatsoever."

The two responses can be related to each other by considering an assessment of the probability of fraud in the lottery. Someone who takes lottery fraud to be a priori unlikely assumes that the prior probability $P(\mathrm{T})$ that the winning number was given away at all is small, let us say 1%. The prior probability for the hypothesis T_i that a randomly specified individual buyer i was involved in a fraud is therefore of the order $P(\mathrm{T_i}) = 10^{-2}/(\text{number of buyers})$. If all tickets were sold, that prior probability is $P(\mathrm{T_i}) = 10^{-2} \times 10^{-6} = 10^{-8}$. In other words, the very high number of sold tickets leads to a very low prior $P(\mathrm{T_i})$ of fraudulent behaviour by an individual buyer. This turns committing the base rate fallacy into a disastrous mistake. We assume $P(\mathrm{W_i}|\mathrm{T_i}) = 1$: if the person fraudulently got the winning number, she is sure to win; and we have already specified $P(\mathrm{W_i}|\neg\mathrm{T}) = 10^{-6}$: if the lottery is fair, the winning chance of a participant i is 10^{-6}. We then extract the total probability $P(\mathrm{W_i}) = P(\mathrm{W_i}|\mathrm{T_i})\,P(\mathrm{T_i}) + P(\mathrm{W_i}|\neg\mathrm{T_i})\,P(\neg\mathrm{T_i}) \simeq 10^{-8} + 10^{-6} \simeq 10^{-6}$. Bayesian updating under the observation W_I that buyer I won the lottery then gives $P(\mathrm{T_I}|\mathrm{W_I}) = P(\mathrm{T_I})/P(\mathrm{W_I}) = 10^{-2}$. Given that only one ticket won and $P(\mathrm{T_i}|\neg\mathrm{W_i}) = 0$, this is equal to the posterior probability $P(\mathrm{T}|\mathrm{W_I})$ that fraud occurred in this lottery at all. We find $P(\mathrm{T}|\mathrm{W_I}) = P(\mathrm{T}) = 10^{-2}$. Committing the base rate fallacy thus led our observer to believe that the outcome of the lottery strongly supported fraud even though fraud remained as improbable as it had been before the lottery results were announced.

Let us now transform our example into a case of the individual theory based NMA. We carry out the following transformation:

- Buyer $i \to$ theory h

- (Number of lottery tickets)$^{-1} \to$ chances of a false theory's predictive success

- Number of sold tickets $\to$ number of theories developed

- W_i: i is drawn in the lottery $\to$ S_h: theory h is predictively successful

- T: The lottery is rigged → T: the theory is approximately true

To keep the analogy perfect, we assume $P(S_h|T_h) = 1$: the approximately true theory is certain to be predictively successful. What we get is an explication of how the individual theory based NMA commits the base rate fallacy (with arbitrary choices of numbers, obviously).

Let us assume that a theory H has predictive success to an extent that had the probability $P(S_H|\neg T_H) = 10^{-6}$ if the theory were not true,. Scientific realists develop the hypothesis T that a predictively successful theory H is most often approximately true. Exponents of the individual theory-based NMA point to theory H and ask: "The probability that the theory is predictively that successful without being approximately true is 10^{-6}. What other explanation than the theory's approximate truth could explain its predictive success?"

The correct formal Bayesian response to the exponent of individual theory-based NMA is: "You fell prey to the base rate fallacy. Since you have not specified the prior probability of the theory's approximate truth, your argument is invalid." Another way to respond is: "You did not even check whether the occurrence of predictive success of the observed degree is unlikely. In order to do so you have to look at the overall rate of predictive success. If the number of theories that have been developed in science is 10^6 there is nothing unusual about the fact that a theory with the observed degree of predictive success is among them."

The two answers can be related to each other by considering an assessment of the probability of approximately true theories in science. Let us take the scientific anti-realist to assume that the probability $P(T)$ that there is a known theory in science that is approximately true is small, let us say 1%. The probability $P(T_h)$ that a random theory, irrespectively of its success, is true therefore is of the order $10^{-2}/(number\ of\ theories\ that\ have\ been\ developed)$. If the number of theories that were developed is 10^6, $P(T_h) = 10^{-8}$. Based on $P(S_h|T_h) = 1$ and $P(S_H|\neg T_H) = 10^{-6}$, we can write the total probability $P(S_h) = P(T_h)P(S_h|T_h) + P(\neg T_h)P(S_h|\neg T_h) \simeq 10^{-8} + 10^{-6} \simeq 10^{-6}$. Bayesian updating under the observation S_H that theory H was predictively successful then gives $P(T_H|S_H) = P(T_H)/P(S_H) = 10^{-2}$. If H was the only theory that had novel predictive success, our assumption $P(T_h|\neg S_h) = 0$ implies that $P(T|S_H) = P(T) = 10^{-2}$. Committing the base rate fallacy thus has lead the exponent of the individual theory based NMA to believe that the predictive success of H strongly supports scientific realism even though scientific realism has remained as improbable as it was before the predictive success of H was observed.

We know already that the base rate fallacy can be avoided by formulating the NMA in a frequency-based way. What we can see in our example, however, is that this may even happen if the number of known predictively successful theories is just 1. Going back to the lottery case, sophisticated

observers might indeed check how many tickets were sold and find out that the number was just 10. This means that the outcome that there was a winning ticket at all was indeed very improbable, namely 10^{-5}. Updating on the observation that there was a winning ticket then does strongly increase the fraud probability. Assuming an a priori overall fraud probability of 10^{-2} now indeed leads to the conclusion that fraud is very probable. A "scientific" test of the fraud hypothesis can only be carried out, however, if the fraud is systematic and occurs again in future lotteries. Only then a staunch believer in the purity of lotteries could be convinced of the fraud hypothesis by establishing a rate of wins in the lottery that is systematically too high.

Translating this into the NMA scenario, the exponent of the frequency-based NMA may argue that the probability of a false theory's stunning predictive success would be so small that such stunning success is unlikely to have occurred at all in the history of science. Even if this just involves one example of predictive success, it is a frequency-based argument and thus avoids the base rate fallacy. In order to be trustworthy, however, the defender of the NMA must be ready to wait for further examples of similarly stunning predictive success in science. If no other case ever occurred, the significance of the first instance of stunning predictive success would at some stage wane and the early stunning success would have to be viewed as a statistical fluke.

9 Conclusions

The following picture of the role of the base rate fallacy with respect to the NMA has emerged. Howson's argument decisively destroys the individual theory-based NMA, which has been endorsed by some adherents to NMA and clearly was not understood to be logically flawed by many others. This is an important step towards a clearer understanding of the NMA and the scientific realism debate as a whole. However, the individual theory-based NMA is only one part of the NMA as it was presented by Putnam and Boyd. The full argument, which we call the frequency-based NMA, does not commit the base rate fallacy.

A clearer understanding of what the base rate fallacy amounts to in the context of the NMA can be achieved by phrasing it in terms of the convergence behaviour of posteriors. An argument commits the base rate fallacy if it (i) ignores the role of the subjective priors and (ii) does not offer a perspective of convergence behaviour under an always extendable sequence of updatings under incoming data. The individual theory-based NMA is structurally incapable of providing such a sequence of updatings because it addresses only the spectrum of novel predictions provided by one single

theory.[9] Even worse, it does not consider the success frequency at all and therefore offers no basis whatsoever for taking the observed predictive success to be unlikely. The frequency-based NMA, to the contrary, is based on a general observation about the research process (the frequency of predictive success in a research field) that can be tested by collecting a sequence of data points where each data point corresponds to the observed novel predictive success of an individual theory. That sequence can always be extended by taking new theories into account. Therefore, the process of testing the hypothesis "theories that have novel predictive success are probably true" under the assumptions $\mathbf{A}_1^{\mathbf{O}}$ and $\mathbf{A}_2^{\mathbf{O}}$ is of the same type as scientific testing and does not commit the base rate fallacy.

One worry about the frequency-based NMA is related to the understanding that the high frequencies of predictive success necessary for having a convincing NMA cannot be found in actual science. The formalisation of the frequency-based NMA demonstrates that such high frequencies are not necessary for achieving high truth probabilities with respect to theories with novel predictive success.

To end with, let us emphasise one important point. This article has shown that the NMA, if correctly reconstructed, does not commit the base-rate fallacy and does not require a high overall frequency of predictive success. Those results are not sufficient for demonstrating the soundness of the NMA, however. A supporter of the frequency-based NMA must justify assumptions $\mathbf{A}_1^{\mathbf{O}}$ and $\mathbf{A}_2^{\mathbf{O}}$ and must still explain on which grounds she takes a *sufficiently* high frequency of predictive success to be borne out by the data. Whether or not that can be achieved lies beyond the scope of this article.

Acknowledgements

---

[9]Formally, one might, in cases of particularly successful theories, divide the theory's novel predictive success into several separate instances the testing of which may in fact be spread out over a period of time. However, this does not improve the case for the individual theory NMA since the degree of improbability of the theory's overall novel predictive success must remain unchanged under this operation. What is gained by increasing the number of instances of novel predictive success is lost due to the correspondingly diminished improbabilities of the individual instances.

A Appendix

A.1 Derivation of Eq. (5)

We use the law of total probability and obtain:

$$
\begin{aligned}
P(\text{S}|\text{O}) &= P(\text{S}|\text{T},\text{O})P(\text{T}|\text{O}) + P(\text{S}|\neg\text{T},\text{O})P(\neg\text{T}|\text{O}) \\
&= P(\text{S}|\text{T},\text{O})P(\text{T}|\text{O}) + P(\text{S}|\neg\text{T},\text{O}) \cdot [1 - P(\text{T}|\text{O})] \\
&= P(\text{T}|\text{O}) \cdot [P(\text{S}|\text{T},\text{O}) - P(\text{S}|\neg\text{T},\text{O})] + P(\text{S}|\neg\text{T},\text{O})
\end{aligned}
$$

Hence,

$$
P(\text{T}|\text{O}) = \frac{P(\text{S}|\text{O}) - P(\text{S}|\neg\text{T},\text{O})}{P(\text{S}|\text{T},\text{O}) - P(\text{S}|\neg\text{T},\text{O})} \,.
$$

Here we have used that $P(\text{S}|\text{T},\text{O}) > P(\text{S}|\neg\text{T},\text{O})$, which follows from assumptions $\mathbf{A}_1^Q$ and $\mathbf{A}_2^Q$. □

A.2 Derivation of Eq. (8)

We start with eq. (5) and obtain

$$
\begin{aligned}
P(\text{T}|\text{O}) &= \frac{P(\text{S}|\text{O}) - P(\text{S}|\neg\text{T},\text{O})}{P(\text{S}|\text{T},\text{O}) - P(\text{S}|\neg\text{T},\text{O})} \\
&\geq \frac{P(\text{S}|\text{O}) - P(\text{S}|\neg\text{T},\text{O})}{1 - P(\text{S}|\neg\text{T},\text{O})} \\
&\geq P(\text{S}|\text{O}) - P(\text{S}|\neg\text{T},\text{O}) \qquad (18) \\
&> P(\text{S}|\text{O}) - k. \qquad (19)
\end{aligned}
$$

Here eq. (18) follows from assumptions $\mathbf{A}_1^Q$ and $\mathbf{A}_2^Q$ and eq. (19) follows from assumption $\mathbf{A}_2^Q$. □

A.3 Proof of the Remark Following Eq. (11)

We start with eq. (11) and assume that $\alpha := P(\text{S}|\text{O})$ and $\beta := P(\text{S}|\neg\text{T},\text{O})$ are fixed with $\alpha > \beta$. Next we set $y := P(\text{T}|\text{S},\text{O})$ and $x := P(\text{S}|\text{T},\text{O})$. Then:

$$
y = \frac{\alpha - \beta}{\alpha} \cdot \frac{x}{x - \beta} \qquad (20)
$$

We now differentiate y by x and obtain

$$
\frac{dy}{dx} = -\frac{\alpha - \beta}{\alpha} \cdot \frac{\beta}{(x - \beta)^2} < 0. \qquad (21)
$$

Hence y (i.e. $P(\text{T}|\text{S},\text{O})$) is a decreasing function of x (i.e. $P(\text{S}|\text{T},\text{O})$). □

A.4 Derivation of Eq. (14)

We start with eq. (13) and obtain in a sequence of transformations:

$$\frac{1}{P(\text{S}|\text{O})} \cdot \left(\frac{P(\text{S}|\text{O}) - P(\text{S}|\neg\text{T}, \text{O})}{1 - P(\text{S}|\neg\text{T}, \text{O})} \right) \; > \; K$$

$$\frac{1}{P(\text{S}|\text{O})} \cdot \left(P(\text{S}|\text{O}) - P(\text{S}|\neg\text{T}, \text{O}) \right) \; > \; K \left(1 - P(\text{S}|\neg\text{T}, \text{O}) \right)$$

$$1 - \frac{P(\text{S}|\neg\text{T}, \text{O})}{P(\text{S}|\text{O})} \; > \; K \left(1 - P(\text{S}|\neg\text{T}, \text{O}) \right)$$

$$1 - K \left(1 - P(\text{S}|\neg\text{T}, \text{O}) \right) \; > \; \frac{P(\text{S}|\neg\text{T}, \text{O})}{P(\text{S}|\text{O})}$$

Hence,

$$P(\text{S}|\text{O}) > \frac{P(\text{S}|\neg\text{T}, \text{O})}{1 - K + K \cdot P(\text{S}|\neg\text{T}, \text{O})} \, . \quad \square$$

References

Barnes, E. (2003). "The Miraculous Choice Argument for Realism", *Philosophical Studies* 111: 97–120.

Boyd, R. (1984). "The Current Status of Scientific Realism", in: J. Leplin (ed.): *Scientific Realism.* Berkeley: University of California Press, pp. 41–82.

Boyd, R. (1985). "Lex Orandi and Lex Credendi", in: P. Churchland and C. Hooker (eds.): *Images of Science: Essays on Realism and Empiricism with a Reply from Bas C. van Fraassen.* Chicago: University of Chicago Press, pp. 3–34.

Dawid, R. (2008). "Scientific Prediction and the Underdetermination of Scientific Theory Building", PhilSci-Archive 4008.

Fine, A. (1986). "The Natural Ontological Attitude" in: *The Shaky Game.* Chicago: The University of Chicago Press.

Henderson, L. (forthcoming). "The No-Miracles Argument and the Base Rate Fallacy", to appear in *Synthese.* DOI: 10.1007/s11229-015-0995-7

Howson, C. (2000). *Hume's Problem: Induction and the Justification of Belief.* Oxford: The Clarendon Press.

Howson, C. (2013). "Exhuming the No Miracles Argument", *Analysis* 73 (2): 205–211.

Howson, C. (2015). "David Hume's No Miracles Argument Begets a Valid No Miracles Argument", *Studies in History and Philosophy of Science* 54: 41–45.

Laudan, L. (1981). "A Confutation of Convergent Realism", *Philosophy of Science* 48: 19–49.

Magnus, P. D. and C. Callender (2003). "Realist Ennui and the Base Rate Fallacy", *Philosophy of Science* 71: 320–338.

Menke, C. (2014). "Does the Miracles Argument Embody a Base Rate Fallacy?", Studies in History and Philosophy of Science 45: 103–108.

Musgrave, A. (1988). "The Ultimate Argument For Scientific Realism", in: R. Nola (ed.): *Relativism and Realism in Science.* Dordrecht: Kluwer Academic Publishers, pp. 229–252.

Psillos, S. (2009). *Knowing the Structure of Nature.* New York: Palgrave McMillan.

Putnam, H. (1975). "What is Mathematical Truth?" in: H. Putnam: *Mathematics, Matter and Method*, Collected Papers Vol. 2. Cambridge: Cambridge University Press.

Sprenger, J. (2015). "The Probabilistic No Miracles Argument", *European Journal for Philosophy of Science* online first DOI 10.1007/s13194-015-0122-0.

Stanford, K. (2006). *Exceeding Our Grasp: Science, History, and the Problem of Unconceived Alternatives.* Oxford: Oxford University Press.

van Fraassen, B. (1980). *The Scientific Image.* Oxford: Oxford University Press.

Worrall, J. (2007). "Miracles and Models: Why Reports of the Death of Structural Realism May Be Exaggerated", *Royal Institute of Philosophy Supplement* 61: 125–154.