DOI: 10.1017/psa.2025.22
This is a manuscript accepted for publication in *Philosophy of Science*.
This version may be subject to change during the production process.

Theoretical Stability and Scientific Realism

Johannes Nyström[1*]

[1]Stockholm University, Department of Philosophy
*Corresponding author: Johannes Nyström; johannes.nystrom@philosophy.su.se

Abstract
The frequency of major theory change in natural science is rapidly decreasing. Sprenger and Hartmann (2019) claim that this observation can improve the justificatory basis of scientific realism, by way of what can be called a stability argument. By enriching the conceptual basis of Sprenger and Hartmann's argument, this paper shows that stability arguments pose a strong and novel challenge to scientific anti-realists. However, an anti-realist response to this challenge is also proposed. The resulting dialectic establishes a level of meaningful disagreement about the significance of stability arguments for scientific realism, and indicates how the disagreement can ultimately be resolved.

1. Introduction

Scientific realists have recently sought novel ways to complement the main justificatory basis of their position, the no-miracles argument (NMA), which holds that the only plausible explanation of the predictive success demonstrated by science is that predictively successful theories are approximately true (Putnam (1975), Boyd (1984), Smart (1985)). In light of the conclusion that rationally subscribing to realism on the basis of the NMA is contingent on a fairly strong conviction about the position's prior plausibility (Howson (2000), Magnus and Callender (2004)), realists have in particular tried to identify strategies which supplement the NMA with some further independent support for their position. One interesting strategy of this kind asserts that the *stability* of scientific theories can be used in support of realism. The core claim of the strategy may be stated as follows. Periods of theoretical volatility indicate that the constraints on theory building are weak, and that an abundance of alternative theoretical structures are available to scientists. On the other hand, a period of theoretical stability, like the current stage of theory assessment in science, indicates that viable scientific theories are a scarce good. In this context, the epistemic status of one's theory at hand can be considered strong.

Fahrbach (2011*a*) was first to propose an explicit argument for scientific realism based on theoretical stability.[1] Building on Fahrbach's idea, Sprenger (2016) and Sprenger and Hartmann (2019) carry out the currently most elaborate attempt to

[1]See Wray (2013) for criticism of Fahrbach's argument.

https://doi.org/10.1017/psa.2025.22 Published online by Cambridge University Press

construct a stability argument, and also formalize the argument in Bayesian epistemology. Sprenger and Hartmann conclude that their argument significantly improves the dialectical situation for scientific realism: "Observations of theoretical stability over an extended period of time can greatly increase the range of prior probabilities for which the NMA leads to acceptance of [a theory]." (Sprenger and Hartmann (2019), 100). Despite this promise, their argument has not attracted much attention in the literature. One initial diagnosis of this situation is that disagreement about the significance of theoretical stability is already well-understood as disagreement about the significance of the historical record. The anti-realist charges the realist with overemphasizing the weight of the evidence amassed in favor of the current scientific description of the world, and the realist charges the anti-realist with overemphasizing the weight of the historical evidence against it. Ultimately, the anti-realist can account for a period of stability by pointing at the constraints imposed on theory-building by the current scientific paradigm (Kuhn (1962)) and unconceived alternatives (Stanford (2001), Stanford (2006)).

The aim of this paper is to advance discussions on theoretical stability and scientific realism in two significant steps. First, it is argued that while Sprenger and Hartmann's original argument could plausibly be rejected on the basis of the described dialectic, an *enriched* version of their argument can be constructed which significantly improves it with respect to this problem. The enriched stability argument suggests that periods of theoretical stability can have significant epistemic implications with respect to scientific realism if they are situated at the end point of a consistent pattern of theory change. In that context, the surprise value of a stable period calls for an explanation of stability that goes beyond standard accounts of scientific theory building. Scientific realism may be offered as an explanation of that kind.

The second step is the construction of a possible response to this argument by the anti-realist. The response involves another non-standard explanation of theoretical stability that does not imply realism. The anti-realist may argue that modern science finds itself in a quite special situation as a result of the uniquely large degree of empirical success demonstrated by scientific theories in the 20th and 21st century. In this situation, a conservative research strategy which sticks with the established theoretical framework(s) may typically look like the best option for realizing the core aims of science. Hence, given the exceptional success of the current scientific picture of the world, the stability that is characteristic of modern science is no surprise at all.

The resulting polemic between these two rival explanations of theoretical stability shows that stability arguments actually offer a progressive take on the realism debate which cannot be sorted under classical categories of disagreement. An enriched stability argument engages the historical record of theory change in *favor* of scientific realism, and is countered by an anti-realist line of reasoning which points at the exceptional degrees of predictive success of our current best scientific theories. This polemic also implies that advocating for a non-conservative research strategy can be a good option for boosting trust in scientific realism.

The plan of the paper is as follows. In section 2, Sprenger and Hartmann's stability argument in favor of realism is presented. The argument assumes a rational expectation on scientists to find and develop alternative theories within a roughly specified time frame, if such alternatives exist at all. In section 3.1, two core concerns about this assumption is raised on the basis of classical anti-realist perspectives on scientific theory building. It is then argued that enriching their argument by invoking a consistent prior

https://doi.org/10.1017/psa.2025.22 Published online by Cambridge University Press

record of theory change can act as a cogent and strong justification for an assumption of this kind, and that enriching the argument in this way provides an empirical basis for a significant stability argument.

In section 3.2, an anti-realist response to the enriched argument is proposed. The response is based on an alternative explanation of stability which, even given a prior record of theory change, does not imply scientific realism. The suggested explanation is based on the concept of theoretical conservatism (Stanford (2019)) but provides a plausible general reason for assuming high degrees of conservatism in the current period of theory assessment in science: The exceptional degrees of predictive success demonstrated by our current best scientific theories. In section 3.2.2, Newtonian mechanics is employed as a case study in support of this alternative explanation. It is suggested that theory assessment in physics in the 18th and 19th centuries in one important respect resembles the situation today in many scientific disciplines. The dominant theoretical framework is exceptionally successful compared to earlier theories in the field, and this success may in part explain the long-term stability of that framework through the adoption of a fairly conservative research strategy. Section 3.2.3 refers to Fahrbach's scientometric investigation of scientific activity (Fahrbach (2011*a*)) in order to motivate the claim that current scientific theories are, by and large, indeed exceptionally successful.

In section 3.3, the suggested anti-realist response is given a formal interpretation as a constraint on a probability distribution in Sprenger and Hartmann's model of a stability argument. It is shown that this constraint strongly weakens the argument, even on otherwise generous concessions to the realist. Hence, significant disagreement about the significance of stability arguments can be established on the basis of this response.The two-step contribution of this paper thereby establishes a meaningful basis for further discussions on theoretical stability. In section 3.4, some implications of the results for recent debate on theoretical conservatism in science are discussed. Section 4 sums up the main results of the paper and clarifies the current status of stability arguments in light of its contributions.

2. Sprenger and Hartmann's stability-based no miracles argument

2.1. The conceptual framework

Sprenger and Hartmann's [henceforth, 'SH'] stability argument is based on a meta-level observation about a scientific discipline. They consider a scientific theory T stable iff there is an *absence of alternative theories* to T in the discipline that persists over a long period of time (Sprenger and Hartmann propose that thirty to fifty years is an adequate measure, but this specific choice plays no decisive role in their argument). SH claim that the failure of scientists in the discipline to come up with and develop rivalling theories in this time indicates that there is a significant scarcity of possible alternative theories, or even that there are no possible alternatives at all. On this basis, they then conclude that the epistemic status of a stable theory can be considered significantly improved.

SH understand a possible alternative to an established theory T to be a theoretical structure (conceived or unconceived) that (i) satisfies a set of (context-dependent) theoretical constraints, (ii) is consistent with available evidence and (iii) makes distinguishable predictions for some future observation(s). Crucially, they do not consider theoretical structures that are empirically equivalent to T as alternatives, since they do

https://doi.org/10.1017/psa.2025.22 Published online by Cambridge University Press

not satisfy (iii). For this reason, SH do not take their stability argument to bear *directly* on the question of whether or not T is approximately true.

The reason that they nevertheless take the argument to bear on the question of scientific realism is connected to their understanding of the NMA, which they construe as a two-step argument. First, T's predictive success is taken to indicate that T is empirically adequate (i.e., consistent with all possibly collectible data). Second, T's empirical adequacy is taken to indicate that T is true. While both steps are necessary for accepting the full realist conjecture, justifying the first step of the NMA is already an important result for the realist and is in line with a realist interpretation of T. However, this step is far from trivial. According to SH, the pessimistic meta-induction (PMI) (Laudan (1981)) is a threat already at this level: "Laudan's PMI ... teaches us that there have often been false theories that explained the data well (and were superseded later). In other words, empirically non-adequate theories can be highly successful." (Sprenger and Hartmann (2019), 86)

In light of the PMI, amassing additional evidence of T's empirical adequacy is of crucial importance for justifying the first step of the NMA. According to SH, T's stability over time may act as support of this kind. Call this argument the *stability-based NMA*.[2]

The stability-based NMA can be considered a synthesis of the NMA and the so-called 'no-alternatives argument' (NAA) formalized and discussed in Dawid, Sprenger and Hartmann (2015). However, there is a crucial difference between what one may call the 'classical' NAA presented there and the version of NAA that SH deploy in the stability-based NMA. The core role of a classical NAA is to generate trust in a theory in the absence of relevant empirical data. It is based on the observation that, despite significant efforts, scientists have failed to come up with alternatives to their theory at hand. The argument infers from that observation that the spectrum of theories which makes distinguishable predictions within the current empirical horizon is probably fairly limited. Hence, the classical NAA addresses a *local* spectrum of alternatives and is therefore an argument in favor of the *viability* of a scientific theory, i.e., its empirical adequacy within the current empirical horizon.

In SH's stability-based NMA, the NAA is deployed to strengthen a realist interpretation of an already empirically well-confirmed theory. To this end, it is deployed as an argument in favor of the unrestricted empirical adequacy of T and therefore addresses the *global* spectrum of alternatives. This version of NAA is consistent with the core epistemic mechanism of the classical NAA and can therefore be considered an extension of its scope toward the question of scientific realism. The extension is interesting and is in line with interpreting scientific realism as a meta-level hypothesis about the scientific research process which makes a wide range of observational predictions about that process, rather than only as a post-hoc explanation about an instance of predictive success.

2.2. The Bayesian model

SH model the stability-based NMA with the help of propositional variables *H, S, C, A*:

[2] 'The NMA' and 'the stability-based NMA' will henceforth be taken to refer to the first inferential step of the full two-step NMA structure, as construed by SH.

https://doi.org/10.1017/psa.2025.22 Published online by Cambridge University Press

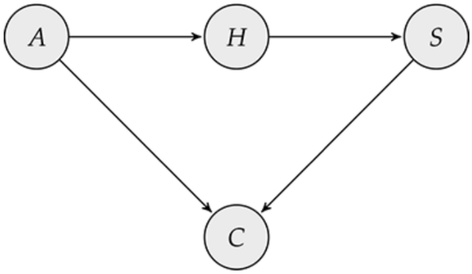

Figure 1. SH's Bayesian network representation of the stability-based NMA.

H := T is empirically adequate.
¬ H := T is not empirically adequate.

S := T is predictively successful.
¬ S := T is not predictively successful.

C := T is not stable (scientists have developed an alternative to T).
¬ C := T is stable (scientists have not developed an alternative to T).

A_j := There are exactly *j* possible alternatives to T that (i) satisfy the relevant theoretical constraints, (ii) are consistent with currently available data and (iii) give distinguishable predictions for some set of future measurements or observations.

Next, SH make the following assumptions, which they consider "equally plausible for the realist and the anti-realist" (Sprenger and Hartmann (2019), 91):

B1: The variables satisfy the (conditional) independencies in the Bayesian network structure of Figure 1.

B2: If T is empirically adequate, it will succeed: $p(\mathrm{S} \mid \mathrm{H}) = 1$.

B3: T is not more or less probable to be empirically adequate than any other possible alternative theory that is consistent with available evidence: $p(\mathrm{H} \mid \mathrm{A}_j) = 1/(j+1)$.

B4: The probability that T is stable over time is a decreasing function of the number of possible alternatives to T:[3,4]

$$p(\neg \mathrm{C} \mid \mathrm{A}_j, \mathrm{S}) = e^{-\frac{1}{2}\left(\frac{j}{\alpha}\right)^2} \tag{1}$$

[3]SH first propose the set of conditional probabilities $p(\neg \mathrm{C} \mid \mathrm{A}_j, \mathrm{S}) := 1/j+1$, but concede that this choice can be considered realist-biased, and instead propose the set given by (1). Since (1) is more neutral and more flexible, the argument discussed here is based on (1).

[4]α is supposed to measure the degree of scientific conservatism in the discipline, and is set $\alpha = 4$ in SH's model but may be varied according to circumstance. More on this below.

https://doi.org/10.1017/psa.2025.22 Published online by Cambridge University Press

B5: The probability that there exist (additional) alternatives to T is the same no matter how many alternatives to T are already known to exist:[5]

$$p\left(A > j \mid \bigvee_{k=j}^{\infty} \mathrm{A}_k\right) = p\left(A > j+1 \mid \bigvee_{k=j+1}^{\infty} \mathrm{A}_k\right) \text{ for all } j \geq 0. \tag{2}$$

Two probabilities are left unspecified by SH's assumptions, and are considered free parameters of the argument: $p(\mathrm{A}_j)$ and $p(\mathrm{S} \mid \neg \mathrm{H})$. The distribution across $p(\mathrm{A}_j)$ reflects one's prior opinion on the number of alternatives to T. Given **B3**, this prior belief determines the core prior $p(\mathrm{H})$ and is therefore considered a free parameter. $p(\mathrm{S} \mid \neg \mathrm{H})$, i.e., the probability that T is successful given that T is not empirically adequate, is considered controversial on the basis of the PMI and is therefore difficult to numerically constrain.

This situation leads to a well-known problem for the classical (non-stability-based) NMA. The inequality $p(\mathrm{H} \mid \mathrm{S}) > f$, where f is some relevant threshold like 0.5, can only be inferred given that values for these parameters are chosen carefully. In other words, the classical NMA only convinces those who take H to be sufficiently probable already to begin with (Howson (2000), Magnus and Callender (2004)). However, SH's stability-based NMA significantly improves the realist's situation with respect to this problem, as may be demonstrated by an evaluation of the argument:

$$p(\mathrm{H} \mid \neg\mathrm{C}, \mathrm{S}) = \frac{p(\neg\mathrm{C}, \mathrm{S}, \mathrm{H})}{p(\neg\mathrm{C}, \mathrm{S})} =$$

$$\frac{\sum_{j=0}^{\infty} p(\mathrm{A}_j) p(\neg\mathrm{C} \mid \mathrm{A}_j, \mathrm{S}) p(\mathrm{S} \mid \mathrm{H}) p(\mathrm{H} \mid \mathrm{A}_j)}{\sum_{j=0}^{\infty} p(\mathrm{A}_j) p(\neg\mathrm{C} \mid \mathrm{A}_j, \mathrm{S}) \left(p(\mathrm{S} \mid \mathrm{H}) p(\mathrm{H} \mid \mathrm{A}_j) + p(\mathrm{S} \mid \neg\mathrm{H}) \left(1 - p(\mathrm{H} \mid \mathrm{A}_j)\right)\right)} \tag{3}$$

$p(\mathrm{A}_j)$ can be expressed (via **B5**) as a function of $p(\mathrm{A}_0)$ (Sprenger and Hartmann 2019, 102-103):

$$p(\mathrm{A}_j) = p(\mathrm{A}_0)\left(1 - p(\mathrm{A}_0)\right)^j \tag{4}$$

B1 - **B5** then generate the following expression for the conditional probability of H, given ¬C and S, evaluated in figure 2:

$$\frac{\sum_{j=0}^{\infty} e^{-\frac{1}{2}\left(\frac{j}{\alpha}\right)^2} \frac{1}{j+1} \left(p(\mathrm{A}_0)\left(1 - p(\mathrm{A}_0)\right)^j\right)}{\sum_{j=0}^{\infty} e^{-\frac{1}{2}\left(\frac{j}{\alpha}\right)^2} \left(p(\mathrm{A}_0)\left(1 - p(\mathrm{A}_0)\right)^j\right) \left(\frac{1}{j+1} + p(\mathrm{S} \mid \neg\mathrm{H}) \left(1 - \frac{1}{j+1}\right)\right)} \tag{5}$$

The results of the stability-based NMA may be understood to generate a reversal of the situation with respect to the described problem of prior commitment to realism. This argument only supports anti-realism if one very carefully chooses parameter values which retain an anti-realist position. In other words, the stability-based NMA only

[5] SH interpret this assumption as follows: "finding an alternative does not, in itself, raise or lower the probability of finding another alternative" (Sprenger and Hartmann (2019): 91). However, this assumption is not fundamentally about the probability of *finding* alternatives, as no reference is made to *C*, but rather about the *existence* of alternatives. Hence, a more formally accurate formulation is chosen here.

https://doi.org/10.1017/psa.2025.22 Published online by Cambridge University Press

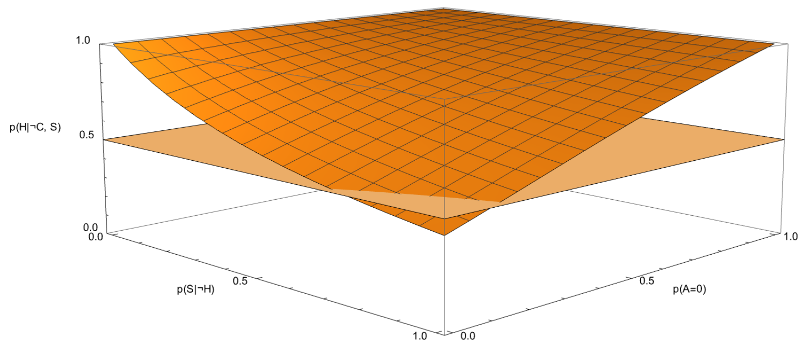

Figure 2. The results of SH's stability-based NMA, with $\alpha = 4$, contrasted with hyperplane $z = 0.5$.

.

fails to convince the strong skeptic who considers the hypothesis that T is empirically adequate to be highly improbable to begin with.[6]

3. On the significance of stability arguments

3.1. Stability and the historical record of theory change

The driver of the improved results of the stability-based NMA is SH's assumption **B4**. According to this assumption, a period of theoretical stability in a scientific discipline should only be expected given that the number of possible alternative theories in the discipline is very small. However, this is precisely the kind of assumption that an anti-realist may take issue with on the basis of her understanding of the process of scientific theory building. When faced with a period of theoretical stability, she may rather conclude that scientists are not sufficiently cognitively and technically competent (with respect to mathematical machinery, for example) to come up with an alternative at all, and that the relevant alternatives therefore remain unconceived (Stanford (2001), Stanford (2006)). Call this the problem of *cognitive capacity*. Another common explanation of stability associated with an anti-realist perspective on theory building is that scientists are working under the strong constraints of the current scientific paradigm (Kuhn (1962)) and therefore do not spend much resources on (genuinely) exploratory theorizing, or that they simply reject rival theoretical frameworks before they have the chance to develop into fully fledged alternatives. Call this the problem of *conservatism*. In this light, SH's claim that their assumptions are neutral with respect to the realist and anti-realist alike is questionable.[7]

[6]It is pertinent to note that while the stability-based NMA has a wider scope than the classical NMA, it does not solve the fundamental problem identified by Howson (2000), namely that the validity of the argument is contingent on a specific set of prior probabilities which are not constrained by the core premise of the NMA. As modeled by SH, this is still true for the stability-based NMA.

[7]A similar problem befalls the classical NAA (Dawid (2013), Menon (2019)). Classical NAA can be strengthened in light of this problem by a meta-inductive argument from the (local) viability of earlier theories in the research field (Dawid (*Forthcoming*)). However, this strategy does not work in the context of the stability-based NMA, as the latter addresses the unrestricted (global) empirical adequacy of T.

https://doi.org/10.1017/psa.2025.22 Published online by Cambridge University Press

Now, it is true that SH only offer what they call a ”possibility result” (Sprenger and Hartmann (2019), 89), and they state that one condition on the improved results of the stability-based NMA is that the scientific discipline in question is not too conservative and that the α parameter (see (1)) therefore takes a small value. In conservative disciplines, which corresponds to a larger value of α, stability will be expected even in light of a large number of possible alternatives and, therefore, ”stability is the default state of [the] discipline and will not strongly support the NMA” (Sprenger (2016), 185). However, SH do not propose any criteria which may be used to identify conservative (or cognitively limited) disciplines. In the absence of criteria of this kind, it is not clear if and how their possibility result can be realized. Hence, the argument is at risk of being sorted to already existing disagreements between realists and anti-realists about the process of theory building in science, and a relevant question is therefore if SH's argument can be understood to improve the realist's dialectical situation at all.

A core claim of this paper is that a general criterion of the described kind can be identified. This criterion is based on the brief but interesting remark of SH that, in conservative disciplines, stability is the *default state*. The function of this statement is to stress that observations of stability are not in need of any special explanation in disciplines that are conservative. As such, it reinforces SH's claim that they provide a possibility result. A stronger interpretation of this remark, however, is to take it to identify conservative disciplines with those that do not have a *demonstrated history of theory change*. Exactly in disciplines without a history of theory change, stability is default. Indeed, adopting this criterion is a natural strategy for the proponent of a stability argument, as it also addresses the problem of cognitive capacity. In disciplines with a demonstrated history of theory change, scientists were at previous points sufficiently cognitively competent to develop alternative theory.

The general importance of this criterion may be emphasized by considering two different ways to understand stability arguments. On a naïve understanding, a stability argument points at an observation of stability and asserts that realism can explain it. The issue with this understanding is that the assertion that an observation of stability is consistent with scientific realism is of marginal importance on its own. What conclusions are drawn on the basis of that observation will be determined by one's background beliefs. Are scientists in the given field cognitively and technically competent? Are they actively trying to develop alternatives? If the answer to those questions is no, or if there is no relevant data in support of a specific answer, observations of stability may be considered trivial and hence cannot underwrite a strong argument in favor of realism. On a more sophisticated understanding, the argument identifies an observation of stability as *unexpected* relative to background beliefs and asserts that those beliefs therefore fail to explain it. Only at this stage is realism proposed as an explanation. This understanding may in principle lead to a strong argument.

Pointing at a historical record of theory change, as suggested here, offers a cogent basis for putting forward the latter kind of argument. Since the current period of theory assessment in many scientific disciplines is situated at the end point of a fairly consistent pattern of theory change, this pattern can be inductively projected into the future. In other words, a historical record of theory change underwrites the belief that scientists are typically cognitively capable *enough*, and typically liberal *enough* with respect to allocating resources to exploratory theorizing, to find and develop alternative theory within some (roughly) specified amount of time. Seeing this pattern broken in the current period of theory assessment is therefore genuinely surprising. The surprise value of that observation may act as the basis of a significant stability argument. Call this an *enriched* stability argument.

The enriched argument provides an empirical basis for assuming a fairly high degree of theoretical creativity in a scientific discipline and therefore offers a plausible justification for **B4**. According to SH's formalism, this amounts to assuming a small value for the parameter α such that the conditional probabilities $p(\neg \mathrm{C} \mid \mathrm{A}_j, \mathrm{S})$ decrease quickly in j (see (1)). Conceptually, this means to assume that observations of stability would be surprising if a large number of possible alternatives were in principle available to scientists at the current stage of theory assessment. Hence, the surprise value of an observation of stability can be naturally interpreted as being based on a probability distribution of that kind. Of course, the exact specification of this distribution will vary depending on context and prior beliefs. Whether or not this distribution is sufficient, given some specified set of priors, for generating posteriors that support a realist position will depend on the specifics of each situation and therefore lies beyond the scope of this paper. The core function of relying on an enriched stability argument is to adopt an argument structure where the expectation on scientists to find and develop alternative theories within a roughly specified time frame can in principle be provided an empirical basis.

The enriched argument also involves a better motivated choice of time frame for assessments of theoretical stability. SH's choice of thirty to fifty years is based on concerns about unfit theories being artificially sustained: "to rule out preservation of a theory by a series of degenerative accommodating moves, [stability] should be evaluated over a longer period, e.g. thirty to fifty years" (Sprenger and Hartmann (2019), 90). However, this choice is fairly arbitrary and, moreover, does not take into account the specific context within which the stability claim is put forward. The enriched argument instead relativizes the time frame across which stability is assessed to the historical record of theory change in the discipline. If earlier periods of theory assessment all lasted well over fifty years before a rival theory was developed, concluding that a current theory is stable just because it survived for fifty years fails to underwrite the claim that the lack of alternatives is surprising. Only when a clear discontinuity with respect to the historical pattern of theory change is observed can the enriched argument get off the ground.

The plausibility of the enriched stability argument may be motivated by general considerations on the history of science which support the claim that theory change has historically been a fairly frequent occurrence. Some versions of the PMI may lend themselves well for this task; Stanford's record of theory change in physics, chemistry, medicine and biology (Stanford (2001)) and Laudan's list of theories discarded in instances of theory change Laudan (1981)) are promising candidates.[8] Both Stanford and Laudan take their lists to be representative of 'a seemingly endless array of theories' (Stanford (2001), 9) that 'could be extended *ad nauseum*' (Laudan (1981), 33).

While these considerations motivate the initial plausibility of enriched stability arguments, the realist who wants to rely on an argument of this kind is still tasked with providing an assessment of her chosen discipline or research field which supports the claim that this field exhibits a consistent historical pattern of novel theory building. This is a non-trivial task for several reasons. Theories are seldom developed as a complete

[8] Lyons (2002) and Vickers (2013) make substantial additions to Laudan's list.

https://doi.org/10.1017/psa.2025.22 Published online by Cambridge University Press

and individuated package and presented to the scientific community at a given time, but may be better understood as constructed over a long series of steps from a core concept to fully developed theory. Hence, a clear pattern of novel theory building may be difficult to describe in terms of definite temporal units. Furthermore, theory change may occur in one sub-domain of the discipline while stability is maintained in another. However, the enriched stability argument does not imply that the realist must be in the position to provide a complete account of the history of the relevant discipline, but rather that an overall appraisal of the history of that discipline should justify the conclusion that there has been a fairly consistent pattern of novel theory building. Of course, whether or not that appraisal in fact supports that conclusion or not is still an important and complex empirical question.

Furthermore, even if the realist is able to provide an assessment of the described kind, a critic may still point at the complex web of social and institutional factors in play in scientific theory building and hold that those factors must be taken into consideration when projecting past theoretical creativity towards the future.[9] For example, she may argue that her meta-level assessment of the research strategy in a given discipline motivates ascribing a small probability of a theoretical shift in the near future, even if the discipline in the past has instantiated a fairly consistent pattern of theory change.[10] On the one hand, an enriched stability argument is based on a finite series of prior instances of theoretical creativity and is therefore structurally incapable of responding to this argument. Hence, this strategy is always a live option for the critic. On the other hand, it is based on a description of a research field that must be empirically motivated on a case-by-case basis. In this light, the proponent of an enriched stability argument may just suggest that, in the absence of a strong motivation of this kind, the statistical basis of her argument is sufficient to justify a future projection of theory change.

3.2. Predictive success and theoretical conservatism

In the previous section, an enriched stability argument in favor of scientific realism was described. The argument claims that observations of stability become epistemically significant with respect to realism when theory change is expected based on a historical record of theory change. As proposed above, the anti-realist may reject this expectation on the basis of a discipline-specific assessment of the factors at play in theory building in the discipline. However, the question arises if the anti-realist can also give a more general explanation of the current period of stability in science that is consistent with her anti-realist position. In other words, is there some *structural* explanation of theoretical stability that (i) cannot be rejected by pointing at the historical record of theory change but that (ii) does not speak strongly in favor of realism? The aim of this section is to show that an explanation of exactly this kind can be provided.

The suggested explanation is based on the concept of theoretical conservatism but provides a strong and highly general basis for assuming high degrees of conservatism in the current scientific landscape. This explanation can be constructed as a three-step line of reasoning. First, a correlation between predictive success and theoretical conservatism is asserted. Hence, large degrees of predictive success in a scientific discipline

[9]Thanks to an anonymous reviewer for this journal for pressing this point.

[10]For example, she could point at peer-review bias (Lee et al (2013)), an aversion towards risky research (Braben (2004)), and a structural inability to identify innovative research proposals (Luukkonen (2012)).

https://doi.org/10.1017/psa.2025.22 Published online by Cambridge University Press

are accompanied by an expectation that the discipline will tend to constrain the distribution of resources for exploratory theorizing. Second, it is claimed that this correlation is substantial. In effect, theories in that discipline will be expected to be stable. Finally, the point is made that since the currently endorsed theories in science enjoy significantly higher degrees of predictive success compared to theories of the past, today's scientific disciplines are probably much more conservative and the stability of the current stage of theory assessment is therefore not very surprising. On this basis, the anti-realist then concludes that this stability cannot underwrite a strong argument in favor of scientific realism.

Now, granting strongly increasing degrees of predictive success may look to the anti-realist like conceding too much to the realist. However, if the realist is forced to retreat back to classical NMA-type reasoning, she is no longer able to point at an additional independent line of justification for her position, which is the core role of a stability argument. Asserting a strong correlation between predictive success and stability is therefore an effective counterargument to the enriched stability-based NMA, as will be demonstrated formally below (Section 3.3). Call the problem identified by this line of reasoning the *new problem of conservatism*. The following three sections fleshes out and further motivates the problem.

3.2.1. Professionalization, peer-review, and hierarchy

Recently, Stanford (2019) has argued for the following conclusion: "Todays scientific communities are almost certainly more effective vehicles for testing, evaluating, and applying theoretical conceptions of various parts of the natural world than were their historical predecessors, but ... we have compelling reasons to believe that they are actually less effective than those same predecessors in conceiving, exploring, or developing fundamentally novel theoretical conceptions of nature in the first place." (Stanford (2019), 3931)

Stanford identifies three core reasons for taking this claim seriously. First, the professionalization of natural science have led to stronger constraints on what kind of research questions are deemed genuinely scientific, since those constraints delimit science as a professional enterprise. Today, departure from those constraints come with the risk of losing one's livelihood. Second, securing funding for research projects today typically requires approval from a committee of peers who assess the proposal from the perspective of theoretical orthodoxy. The scarce resources available in a scientific discipline encourages those committees to prioritize research that already from the start is understood to probably generate progress, even if minor, over risk-taking exploratory theorizing. Finally, the hierarchical structure of modern science ensures that research is conducted primarily under the leadership of senior researchers. Therefore, creativity and intellectual independency is constrained by existing frameworks of conducting research, both by content and method. Skills passed on to younger generations of scientists are specifically designed to prepare them for careers in the established institutional and theoretical context.

These points suggest that today's scientific communities are quite conservative with respect to allocating resources for exploratory theorizing. However, it is not clear to what extent they identify a clear discontinuity with respect to earlier periods of theory building in science. It seems equally plausible to assume that the professionalization of science has been a gradual process over a long period of time, that peer-review has, in

https://doi.org/10.1017/psa.2025.22 Published online by Cambridge University Press

some form, always been a core part of the scientific process, and that the process of preparing younger generations of scientists for independent research is almost by definition constrained by the current understanding of the theoretical and methodological situation. Moreover, it may be equally plausible to claim that the current stage of theory assessment in science allocates *more* resources to exploratory theorizing, simply due to the sheer growth in size of the scientific enterprise (Forber (2008), Godfrey-Smith (2008)). On its own, Stanford's claim therefore does not seem to be a sufficient reply to the realist who takes the stability of the current stage of theory assessment in science to be genuinely surprising.

One way to strengthen Stanford's argument in light of this complexity is to identify some feature of the current period of theory assessment which can (i) be connected to conservatism but (ii) was not prominent during past periods of instability. In other words, is there some reason to believe that conservatism has strongly accelerated during the current period of theory assessment in science?

Pointing at the uniquely large degrees of predictive success exhibited by currently endorsed theories in science offers a cogent way of identifying a feature of this kind. This claim is based on the idea that the level of conservatism of a scientific discipline varies with the degree to which that discipline builds theory that can satisfactorily account for the continuous assembly of empirical data. Hence, the accelerated development of the scientific institutions that foster conservatism - professionalization, peer-review, and hierarchy - are at least partly the result of the high success rate of scientific theory in the last century, rather than simply the natural progression of a maturing scientific enterprise or sociological degeneration.

In fact, this claim may be understood to follow directly from the mission statement of natural science. One of its main goals is to construct empirically successful theories. Strategies that propel science toward that goal are understood to be fruitful and productive. Hence, if currently endorsed theories are persistently successful with respect to that goal, the risk/reward-calculation of exploratory theorizing in an environment with limited resources will naturally favor theories that are already successful. The same understanding implies that if the theories currently endorsed in science had been less successful in accounting for empirical data, the same disciplines would have been less conservative. In that context, the risk/reward-calculation looks quite different.

3.2.2. Theoretical stability in Newtonian mechanics

The new problem of conservatism may also be supported by an example from the history of natural science. Newtonian mechanics (NM) is widely considered a cornerstone of anti-realist rejections of the NMA (e.g., Laudan (1981), Vickers (2013)). Despite being tremendously successful, NM eventually proved incapable of accounting for the continuous assembly of empirical data and was eventually replaced by general relativity. Hence, it nicely captures the intuitive and formal motivation behind rejecting the classical NMA. SH note that NM also constitute a (probabilistic) counterexample to the stability-based NMA, since it was a highly stable (according to their definition) and predictively successful theory that was nevertheless eventually proven empirically non-adequate (Sprenger and Hartmann (2019), 89). The proponent of an enriched argument fares better with respect to this problem, because she must not assume that realism provides a good explanation of the apparent stability of NM. She may claim that, in

https://doi.org/10.1017/psa.2025.22 Published online by Cambridge University Press

the absence of a consistent prior pattern of theory chance in physics, cognitive limitations and conservatism are better explanations of this stability. However, in conjunction with the new problem of conservatism, NM presents a problem also for the enriched argument.

The theoretical situation in physics in the 18th and 19th century may, in one important respect, be understood to resemble the situation of the current period of theory assessment in science. The acceptance of NM was motivated by its unparalleled empirical track record. The theory was far more successful than any other theoretical structure, past or present, known at that time, in particular with respect to its novel predictions. The existence of Neptune, the return of Halley's comet, earth's oblate shape (Lyons (2002)), to name a few of those predictions, reassured physicists that the NM paradigm would continue to deliver successful predictions. Hence, empirical anomalies were not typically treated directly as reasons to reject NM and look for alternatives, but mainly invoked solutions within the framework of NM. Some solutions of that kind were successful, while others, like LeVerrier's hypothesized bodies of mass between the sun and Mercury responsible for the anomalous advance of the perihelion of the latter, were not. However, the continuous lack of progress on the latter issue was not taken as a reason for major revision of the research strategy, but rather provoked more and more elaborate solutions within the existing theoretical framework, and much theoretical and observational work was being devoted to solutions of that kind (Baum and Sheehan (1997), 127-144). Physicists at the time had, in light of their experience with NM's record of predictive success, no strong reasons to believe that the limits of NM were surpassed at that point, and were therefore not inclined to devote substantial resources to searching for solutions to the existing anomalies on the basis of a fundamentally novel theoretical framework.[11]

The process of theory assessment in NM is consistent with the perspective on theoretical stability offered by the new problem of conservatism. The historically exceptional degrees of predictive success enjoyed by this theoretical framework led the physics community to the belief that solutions to new and existing scientific problems could probably be continuously devised within that framework. This belief in turn informed a fairly conservative research strategy. While the details of this case study can be further nuanced, the overall picture is consistent with the new problem of conservatism and may therefore be understood to motivate this problem as a serious explanatory strategy in the context of a stability argument in the current period of theory assessment in science.

[11] A brief mention of a more recent example is also motivated here. Already when quantum field theory became the most popular framework for developing theories about particle interactions in the early-to-mid 20th century, physicists recognized foundational issues with the theory connected to its break-down at high energy scales, and much theoretical work was devoted to finding possible alternatives. However, after the standard model of particle physics, a product of quantum field theory, was exposed to empirical testing in the 1970's and turned out to be successful, most of the theoretical and experimental work in physics was carried out within the framework of quantum field theory. "As the 1980's began ... all experimental results were henceforth to be compared to the standard model - the new "paradigm" of the field. Where anomalies cropped up, they were usually resolved by moderate elaboration of the theory (or by uncovering errors made by the experimenters)" (Hoddeson et al (1997), 29).

https://doi.org/10.1017/psa.2025.22 Published online by Cambridge University Press

3.2.3. Are today's theories exceptionally successful?

At this stage, a genuine concern about the problem described above may be raised. Are currently endorsed theories in science, at large, in fact exceptionally predictively successful compared to theories of the past? Can predictive success really be quantified and measured in a way which may serve as the basis of the new problem of conservatism?

One answer to this question is based on scientometric data. Fahrbach (2011*a*,b) estimates that due to the exponential growth of scientific activity in last century, about 80% of scientific work ever carried out has been conducted after 1950, after which "refutations among [our current best scientific theories] have practically not occurred" (Fahrbach (2011*a*), 17). He then assumes that, at a general level, "more scientific work results in the discovery of more phenomena and observations, which, in turn, can be used for more varied and better empirical tests of theories. More varied and better empirical tests of theories, if passed, amount to more empirical success of theories" (Fahrbach (2011*a*), 14). By these estimations, which seem fairly reasonable, one may conclude that the current period of theory assessment in science is in fact exceptionally successful.

For these reasons, the line of reasoning followed here leads to a general statement that the critic of an enriched stability argument may understand to be fully cogent and fairly plausible. The new problem of conservatism emphasizes the correlation between predictive success and conservatism. On this basis, it asserts that the fact that currently endorsed theories perform so well over repeated attempts to refute them in itself explains their stability. In other words, *stability is exactly what one should expect* to observe in scientific disciplines that work with theoretical structures that turn out to be exceptionally successful.

3.3. Formal analysis

How should a proponent of the new problem of conservatism update her beliefs about scientific realism in light of observations of theoretical stability? In order to formally analyse this question, the concept of *degrees of predictive success* must first be offered a formal interpretation. Ideally, these degrees would be represented in their entirety. However, in the interest of simplicity and clarity (and keeping the structure of the formal model similar to SH's), let S^E be a binary propositional variable such that $\mathrm{S}^E := \mathrm{T}$ demonstrates *exceptional* degrees of predictive success, and let $\neg\mathrm{S}^E$ be its negation.

The new problem of conservatism asserts that the theoretical stability demonstrated in the current period of theory assessment is not very surprising, but in fact expected, given the exceptional degrees of predictive success exhibited by our currently endorsed scientific theories. Expectancy, at the very least, must be considering something more probable than not. 'Alice expects rain today' plausibly expresses (at least) that Alice considers it more probable that it rains today than that it does not rain today. Hence, a proponent of the new problem of conservatism will subscribe to the claim that, given exceptional degrees of predictive success, the probability of stability is larger than 0.5. Formally:

$$p(\neg\mathrm{C} \mid \mathrm{S}^E) = \sum_{j=0}^{\infty} p(\neg\mathrm{C} \mid \mathrm{A}_j, \mathrm{S}^E)p(\mathrm{A}_j) > 0.5 \tag{6}$$

At first glance, a natural way to modify the stability-based NMA in order to satisfy this inequality is by increasing the value of the α parameter as proposed by SH (see (1)).

https://doi.org/10.1017/psa.2025.22 Published online by Cambridge University Press

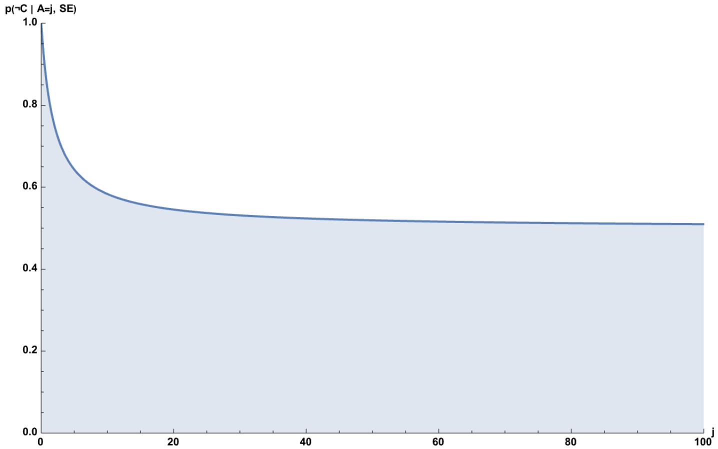

Figure 3. A graphical visualization of (7) with a visual cut-off at $j = 100$.

However, it turns out that α is not sufficiently flexible for this purpose. For all positive integer values of α, $e^{-\frac{1}{2}\left(\frac{j}{\alpha}\right)^2}$ converges to zero in $\lim_{j\to\infty}$. Therefore, the sum $\sum_{j=0}^{\infty} p(\neg C \mid A_j, S^E)p(A_j)$ also converges to zero in $\lim_{j\to\infty}$. Hence, in order to retain (1), a proponent of the new problem of conservatism is forced to assume that the number of alternatives to T is limited already from the start. However, by accepting (6), one makes the claim that, due to theoretical conservatism at least in part caused by T's exceptional success, T's stability remains quite probable even if an infinite amount of possible alternatives could in principle be found. For this reason, a subscriber to (6) cannot be plausibly modeled as accepting (1).[12]

In order to evaluate the implications of the problem for the stability-based NMA, a new set of conditional probabilities $p(\neg C \mid S^E, A_j)$ must therefore be adopted. First, this choice should be consistent with SH's core assumption **B4** that $p(\neg C \mid S^E, A_j)$ are decreasing in j with the extreme case $p(\neg C \mid A_0, S^E) = 1$. Second, it should satisfy (6) given any possible probability distribution across A. Finally, it should be maximally generous to the realist who adopts an enriched stability argument, in order to ensure that the problem is distinguished from other considerations. Hence, $p(\neg C \mid S^E, A_j)$ should decrease quickly in j. The following distribution may be understood as the simplest formally adequate way of satisfying the described conditions:[13]

$$p(\neg C \mid A_j, S^E) = 1/(j+2) + 0.5 \tag{7}$$

[12] As an aside, note that this result contradicts SH's claim that their assumptions are neutral with respect to the realist and anti-realist: here is a plausible anti-realist position which cannot be represented by their assumptions.

[13] Of course, these conditions can be satisfied with other choices, which may side more to the anti-realist or even more to the realist. However, the same is true for (1). It is therefore important to recognize a general limitation of both SH's original argument and the modified argument evaluated here. Both reveal the coarse implications for scientific realism of adopting a certain view on theoretical conservatism, rather than numerical values for actual credences.

https://doi.org/10.1017/psa.2025.22 Published online by Cambridge University Press

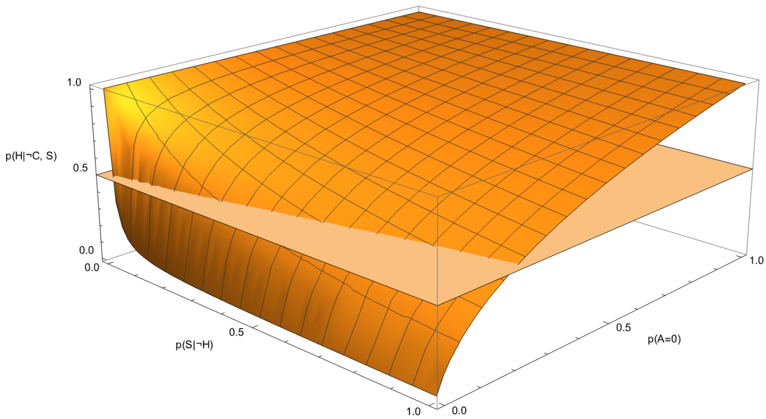

Figure 4. The results of the modified stability-based NMA.

Adopting this distribution leads to the following expression for the posterior probability of H, evaluated in figure 4:

$$p(\mathrm{H} \mid \neg\, \mathrm{C}, \mathrm{S}^E) =$$

$$\frac{\sum_{j=0}^{\infty}\left(\frac{1}{j+2}+0.5\right)\frac{1}{j+1}\left(p(\mathrm{A}_0)\left(1-p(\mathrm{A}_0)\right)^j\right)}{\sum_{j=0}^{\infty}\left(\frac{1}{j+2}+0.5\right)\left(p(\mathrm{A}_0)\left(1-p(\mathrm{A}_0)\right)^j\right)\left(\frac{1}{j+1}+p(\mathrm{S}^E \mid \neg\mathrm{H})\left(1-\frac{1}{j+1}\right)\right)} \tag{8}$$

An apt comparison is with the classical NMA, evaluated in figure 5, which can be extracted from SH's model with the following expression:

$$p(\mathrm{H} \mid \mathrm{S}^E) = \frac{\sum_{j=0}^{\infty}\frac{1}{j+1}\left(p(\mathrm{A}_0)\left(1-p(\mathrm{A}_0)\right)^j\right)}{\sum_{j=0}^{\infty}\left(p(\mathrm{A}_0)\left(1-p(\mathrm{A}_0)\right)^j\right)\left(\frac{1}{j+1}+p(\mathrm{S}^E \mid \neg\mathrm{H})\left(1-\frac{1}{j+1}\right)\right)} \tag{9}$$

The results show that modifying the stability-based NMA according to the new problem of conservatism generates an argument that does not amount to a substantial improvement over the classical NMA. Looking at an example makes the comparison more concrete. A fairly skeptical anti-realist may assign a quite small value to $p(\mathrm{A}_0)$, say 0.01. In SH's model based on (1) with $\alpha = 4$, this assignment still allows significant flexibility with respect to the other free parameter, $p(\mathrm{S}^E \mid \neg\, \mathrm{H})$, which is just required to be in the interval [0, 0.75] in order to conclude $p(\mathrm{H} \mid \neg\mathrm{C}, \mathrm{S}^E) > 0.5$. In the modified argument, based on (7), an interval of [0, 0.13] is required, and for the classical NMA, the interval required is [0, 0.08].[14] Hence, while the modified stability-based NMA does amount to

[14] At this point, the realist may retaliate: $p(\mathrm{S}^E \mid \neg\mathrm{H})$, i.e., the probability that T is exceptionally successful given that T is not empirically adequate, may be considered smaller than $p(\mathrm{S} \mid \neg\mathrm{H})$, i.e., the probability

https://doi.org/10.1017/psa.2025.22 Published online by Cambridge University Press

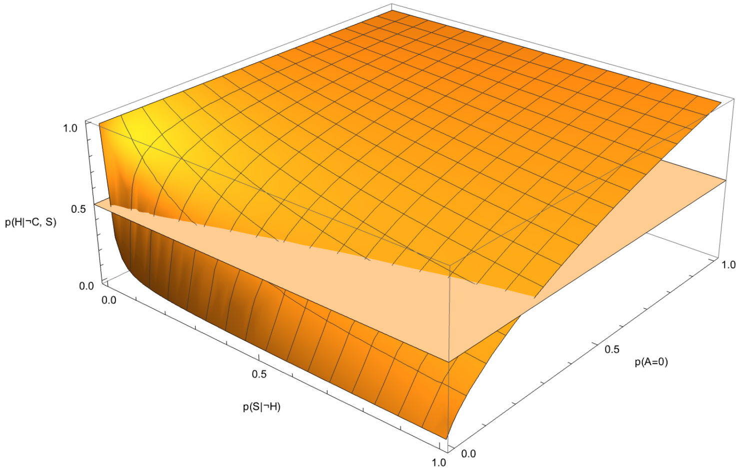

Figure 5. An evaluation of the classical NMA, as extracted from SH's model.

a slight upgrade to the classical NMA, the upgrade is not of substantial significance. The basic uncertainties associated with attaching numerical values to the free parameters of the NMA may be understood to offset a single minor improvement. The important conclusion is that, just like the classical NMA, the modified stability-based NMA requires commitment to quite specific parameter values in order to generate results that support scientific realism.

3.4. Who should advocate for non-conservatism?

Before concluding, a brief aside is motivated. The results of the analysis carried out above offer a contribution to a recent exchange between Stanford (2015) and Dellsén (2018) on theoretical conservatism and scientific realism. Stanford argues that realists should embrace conservatism, since they consider it probable that the currently endorsed theoretical picture of the natural world is, by and large, approximately the correct one. Conversely, an anti-realist perspective on science supports a liberal strategy of allocating resources to exploratory theory building. Against Stanford, Dellsén claims that scientific realists should embrace non-conservatism even if they do not consider it probable that exploratory theorizing will generate promising new theory. Failed exploration of that kind amounts to evidence in favor of the endorsed theory, since this failure is consistent with the understanding that the endorsed theory is the (approximately) correct one. The conclusion of the analysis in the present paper is consistent with the view put forward by Dellsén. A conservative scientific discipline cannot act as a good basis for a strong stability argument in the context of scientific realism. Hence, for the realist, advocating

that T is 'normally' successful given that T is not empirically adequate. Hence, a smaller $p(S^E \mid \neg H)$ could perhaps be motivated. In fact, Fahrbach (2011*b*) may be understood to argue this exact point. However, (i) the acceptable interval ([0, 0.13]) is very small, and (ii) whether assigning a very small value $p(S^E \mid \neg H)$ can be motivated at all is ultimately decided by the dynamic between the PMI and the classical NMA. Hence, the stability argument as such does not significantly contribute to the realist's justification in this case, but merely shifts the goalposts slightly in her favor.

https://doi.org/10.1017/psa.2025.22 Published online by Cambridge University Press

non-conservatism will in the long run generate a stronger basis for an argument of that kind.[15]

It is important to recognize, however, that this implication is connected to scientific realism. And one should not assume that the aims of science and the philosophy of realism are always fully aligned. If an emergence of conservatism in science is considered fully the result of a sociological development that is detrimental to the research process, it may look natural to advocate for non-conservatism. However, the line of reasoning followed in section 3.2 suggests that the situation is more complicated. A reason why one may propose a recent emergence of conservatism in science is the increasing degree of predictive success demonstrated by scientific theories. In this situation, conservatism may be considered a fruitful research strategy, given that a core aim of scientific research is to produce empirically successful theory.

This understanding is consistent with Tambolo and Cevolani (2023) more general claim that whether or not theoretical conservatism should be preferred in a scientific discipline depends on what the 'main cognitive aim' of the discipline is, and the degree to which scientists in the discipline expect currently endorsed theory to continue realizing that aim. From this perspective, it is not at all clear that advocating non-conservatism leads to a more successful scientific enterprise. Whether or not it does depends on what one considers to be part of the core aims of science against which that success is evaluated, and to what degree the current theoretical description of the world is successful in realizing those aims.

4. Conclusions

The line of reasoning followed in this paper has served two core aims. First, an enriched form of a stability argument for scientific realism has been proposed. Enriching the argument addresses a core worry about SH's original argument: Why are observations of theoretical stability in science in need of any special explanation at all? In section 3.1, it was suggested that restricting stability arguments to disciplines which have a historical record of theory change offers a cogent answer to this question. A long period of stability is genuinely surprising, because scientists are typically creative enough to develop novel theory in those disciplines. On this basis, enriched stability arguments can be meaningfully employed in defense of scientific realism and offer a novel challenge to classical anti-realist accounts of scientific theory building.

While the enriched argument is a dialectical upgrade for the realist, a second aim of this paper was to show that the anti-realist is not without resources to defend herself. Against the enriched argument, the *new problem of conservatism* was proposed in section 3.2. The problem is based on the claim that predictive success is strongly correlated

[15] Oriti (2019) and Dawid (2022) discuss a similar issue in the context of classical NAA. Oriti claims that the NAA supports conservatism, since inferring that there probably are no alternatives implies that a search for such alternatives will be fruitless. Against Oriti, Dawid argues that the NAA actually supports a non-conservative strategy: "the more energy has been invested in alternative research programs without success, the more powerful a NAA can become. In this light, it is in the epistemic interest of the dominant research program that alternative research programs are pursued with vigor and in sufficient breadth" (Dawid (2022): 70). This discussion shows interesting parallels between the problem of realism and situations of scarce evidence in science. In both cases, the core problem is that observational data in favor of a theory at hand is insufficient for justifying a given level of trust in the theory. And in both cases, a non-conservative research strategy can be the best option for boosting that trust.

https://doi.org/10.1017/psa.2025.22 Published online by Cambridge University Press

with conservatism. Hence, periods of theory assessment that enjoy exceptional levels of predictive success compared with earlier periods can be expected to be more stable, and it was suggested that an anti-realist may explain periods of stability on this basis and therefore reject even an enriched stability argument in favor of realism. The plausibility of this explanatory strategy was supported by a case study from fundamental physics. It was then suggested, based on Fahrbach's scientometric work on scientific activity, that our current best scientific theories are, by and large, indeed exceptionally predictively successful. The formal analysis carried out in section 3.3 showed that subscribing to this problem strongly undermines the improved results of the stability-based NMA. This result also implies that the realist who wishes to rely on a stability argument to justify her position should advocate a non-conservative research strategy.

The analysis of the new problem of conservatism should not be taken to suggest that science proceeds in exactly the manner described by that problem. Rather, it concludes that understanding the research process on the basis of this problem amounts to a cogent and defensible position for a critic of stability arguments. Ultimately, the problem indicates that to the extent that stability arguments are put forward in order to offer a case for realism that is convincing also to the anti-realist, they are in need of further substantiation. One strategy to that end is to offer an empirically based description of the current stage of theory assessment in science which strongly supports the conclusion that it this period is fairly liberal with assigning resources to exploratory theorizing. Whether or not such a description in facts supports that conclusion is of course an open question.

The general message of the paper is that realists and anti-realists alike access viable strategies for advancing the debate on stability beyond classical disagreements of the realism debate. Notably, these strategies show that the historical record of theory change does not unequivocally support an anti-realist position, and, on the other hand, that the empirical evidence in favor of the current scientific picture of the world does not unequivocally support a realist position. The resulting dialectic is promising because it implies that the disagreement can ultimately be dissolved by the continuous assembly of empirical data. If the current period of stability extends far into the future, the anti-realist can at some point no longer hold that the lack of alternative theory is not genuinely surprising without committing to an absurd description of science as a theoretical preserve fully devoid of traces of creativity. Hence, she would need to retreat to other philosophical disagreements in order to retain her position. On the other hand, the emergence of just a small set of alternatives to our currently endorsed theories should already be cause for alarm for the proponent of a stability argument in favor of realism. The anti-realist may then, in a PMI-style move, suggest that realism is not the only plausible explanation of theoretical stability, even in disciplines that exhibit a high degree of theoretical creativity.

Given the current state of the scientific realism debate, any development which opens the opposing camps to the risk of being at variance with observational evidence must be considered a welcome addition. Exposing scientific realism and anti-realism against new empirical data is of crucial importance for the status of the debate as a rational disagreement about a genuinely testable philosophical hypothesis. For this reason, stability arguments in the enriched form presented here are clearly progressive and should be paid their due attention by realists and anti-realists alike.

Conflicts of interest: None to declare.

https://doi.org/10.1017/psa.2025.22 Published online by Cambridge University Press

Funding information: Open access funding is provided by Stockholm University.

Acknowledgements: I am grateful to Jan Sprenger, Richard Dawid and James Nguyen for feedback on earlier versions of this draft, and to audiences at the Bergen Philosophy of Science Workshop 2023, the Stockholm Higher Seminar in Theoretical Philosophy, and PSA 2024. I also thank two anonymous reviewers for this journal for helpful comments and suggestions.

References

Baum, Richard, and Sheehan, William. 1997. *In search of planet Vulcan: the ghost in Newton's clockwork universe*, New York: Springer New York https://doi.org/10.1007/978-1-4899-6100-6

Boyd, Richard. 1984. 2The current status of scientific realism." In *Scientific realism*, ed. J. Leplin, 41–82. Berkeley: University of California Press. https://doi.org/10.1525/9780520337442

Braben, Donald W. 2004. *Pioneering research: A risk worth taking*. Hoboken, NJ: Wiley-Interscience

Dawid, Richard. 2013. *String theory and the scientific method*. Cambridge: Cambridge University Press https://doi.org/10.1017/CBO9781139342513

Dawid, Richard. 2022. "Meta-empirical confirmation: addressing three points of criticism." *Studies in the history and philosophy of science* 93:66-71. https://doi.org/10.1016/j.shpsa.2022.02.006

Dawid, Richard. (Forthcoming). "Does the No Alternatives Argument Need Gerrymandering to Be Significant?" *British Journal for the Philosophy of Science* 717081. https://doi.org/10.1086/717081

Dawid, Richard, Hartmann, Stephan, and Sprenger, Jan. 2015. 2The No Alternatives Argument." *British Journal for the Philosophy of Science* 66 (1):213-234. https://doi.org/10.1093/bjps/axt045

Dellsén, Finnur. 2018. "Should Scientific Realists Embrace Theoretical Conservatism?" *Studies in History and Philosophy of Science* 76 (Part A):30-38. https://doi.org/10.1016/j.shpsa.2018.09.005

Farhbach, Ludwig. 2011a. "How the Growth of Science Ends Theory Change." *Synthese* 180 (2):139-155. https://doi.org/10.1007/sl 1229-009-9602-0

Farhbach, Ludwig. 2011b. "Theory Change and Degrees of Success." *Philosophy of Science* 78(5):1283-1292. https://doi.org/10.1086/662280

Forber, Patrick. 2008. "Forever beyond our grasp?" *Biology and Philosophy* 23:135–141. https://doi.org/10.1007/s10539-007-9074-x.

Godfrey-Smith, Peter. 2008. "Recurrent transient underdetermination and the glass half full." *Philosophical Studies* 137:141–148. https://doi.org/10.1007/s11098-007-9172-2.

Hoddeson, Lillian, Brown, Laurie, Riordan, Michael, and Dresden, Max. 1997. The Rise of the Standard Model: 1964-1979. In *The Rise of the Standard Model: A History of Particle Physics from 1964 to 1979*, eds. Hoddeson, Lillian, Laurie Brown, Michael Riordan, Max Dresden. Cambrige: Cambrige University Press https://doi.org/10.1017/CBO9780511471094

Howson, Colin. 2000. *Hume's problem: induction and the justification of belief*. New York: Oxford University Press. https://doi.org/10.1093/0198250371.001.0001.

Kuhn, Thomas. 1962. *The Structure of Scientific Revolutions*, Chicago: University of Chicago Press

Laudan, Laurens. 1981. "A confutation of convergent realism." *Philosophy of Science* 48:19–49. https://doi.org/10.1525/9780520337442-012

Lee, Carole, Sugimoto, Cassidy, Zhang, Guo, and Cronin, Blaise. 2013. "Bias in peer review." *Journal of the American Society for Information Science and Technology* 64:2–17. https://doi.org/10.1002/asi.22784

Luukkonen, Terttu. 2012. "Conservatism and risk-taking in peer review: Emerging ERC practices." *Research Evaluation* 21:48–60. https://doi.org/10.1093/reseval/rvs001

Lyons, Timothy D. 2002. "Scientific Realism and the Pessimistic Meta-Modus Tollens." In *Recent Themes in the Philosophy of Science: Scientific Realism and Commonsense*, eds. Steve Clarke and Timothy D. Lyons, 63-90. Dordrecht: Springer https://doi.org/10.1007/978-94-017-2862-14

Magnus, P. D. and Callender, Craig. 2004. "Realist ennui and the base rate fallacy." *Philosophy of Science* 71:320-338. https://doi.org/10.1086/421536

Menon, Tushar. 2019. "On the viability of the No alternatives argument." *Studies in History and Philosophy of Science* 76 (A):69-75. https://doi.org/10.1016/j.shpsa.2018.10.005

Oriti, Daniele. 2019. "No alternative to proliferation." In *Why trust a theory? -Epistemology of fundamental physics*, eds. R. Dardashti, R. Dawid, K. Thebault, 125-153. Cambridge: Cambridge University Press https://doi.org/10.48550/arXiv.1705.09858

Putnam, Hilary. 1975. ”What is mathematical truth?” In *Mathematics, matter and method*, Collected Papers Vol. 2. Cambridge: Cambridge University Press https://doi.org/10.1017/CBO9780511625268

Smart, John J. C. 1985. ”Laws of nature and cosmic coincidences.” *Philosophical Quarterly* 35 (140):272-280. https://doi.org/10.2307/2218906

Sprenger, Jan. 2016. ”The probabilistic no miracles argument.” *European Journal for Philosophy of Science* 6 (2):173-189. https://doi.org/10.1007/s13194-015-0122-0

Sprenger, Jan, and Harmann, Stephan. 2019. *Bayesian Philosophy of Science*. Oxford: Oxford University Press https://doi.org/10.1093/oso/9780199672110.001.0001

Stanford, Kyle P. 2001. ”Refusing the Devil’s bargain: What kind of underdetermination should we take seriously?” *Philosophy of Science* 68 (S3):1-12. https://doi.org/10.1086/392893

Stanford, Kyle P. 2006. *Exceeding our grasp*. Oxford: Oxford University Press. https://doi.org/10.1093/0195174089.001.0001

Stanford, Kyle P. 2015. ”Catastrophism, Uniformitarianism, and a Scientific Realism Debate That Makes a Difference.” *Philosophy of Science* 82 (5):867-878. https://doi.org/10.1086/683325

Stanford, Kyle P. 2019. ”Unconceived alternatives and conservatism in science: the impact of professionalization, peer-review, and Big Science.” *Synthese* 196 (10):3915-3932. https://doi.org/10.1007/s11229-015-0856-4

Tambolo, Luca, and Cevolani, Gustavo. 2023. ”Realism, antirealism, and theoretical conservatism.” *Synthese* 201 (1):1-18. https://doi.org/10.1007/s11229-022-04022-0.

Vickers, Peter. 2013. ”A Confrontation of Convergent Realism.” *Philosophy of Science* 80 (2):189-211. https://doi.org/10.1086/670297

Wray, Brad. 2013. ”The pessimistic induction and the exponential growth of science reassessed.” *Synthese* 190 (18):4321-4330. https://doi.org/10.1007/s11229-013-0276-2