DOI: 10.1017/psa.2026.10210
This is a manuscript accepted for publication in *Philosophy of Science*.
This version may be subject to change during the production process.

Parsimony and Overfitting

James H. McIntyre

Rutgers University—New Brunswick

Email: james.mcintyre@rutgers.edu

Abstract:

Philosophers often defend appeals to parsimony by invoking its central role in science. I argue that this move fails once we distinguish between two uses of parsimony: *non-ideal* and *ideal*. Non-ideal parsimony enjoys strong inductive support in science, since complex models are prone to overfit to predictively irrelevant noise. But philosophical data aren't significantly noisy in the relevant sense: when our intuitions are unreliable, their unreliability typically reflects systematic *bias* rather than noise, which parsimony doesn't mitigate. Philosophers therefore need ideal parsimony, which finds only weak support from science. Thus, the scientific analogy cannot vindicate the philosopher's use of parsimony.

Keywords: parsimony, overfitting, theoretical virtues, noise, bias, methodology

https://doi.org/10.1017/psa.2026.10210 Published online by Cambridge University Press

1 Introduction

Since empirical evidence underdetermines theory, theoreticians must turn to the so-called "non-empirical" theoretical virtues to choose between empirically adequate theories. Chief among these virtues is parsimony, which plays a central role in both scientific and philosophical methodology. All else equal (and often when all else is decidedly *not* equal), parsimonious theories, in some sense to be spelled out, are preferred.

The use of parsimony in evaluating philosophical theories remains contentious.[1] However, even many skeptics—impressed by the success of science—are happy to concede parsimony's legitimacy in that domain. This has prompted some defenders of parsimony to argue that its role in philosophy closely mirrors its role in science. My aim in this paper is to challenge that analogy, on the grounds that it conflates two distinct applications of parsimony: what I'll call *ideal* and *non-ideal* parsimony. Under conditions of noisy data—roughly, when random variation is prominent—the non-ideal use of parsimony functions as a safeguard against overfitting. Overfitting occurs when a model mistakes incidental patterns in the noise for predictively relevant ones. The non-ideal use of parsimony discounts weak patterns and is thereby more likely to capture those that are predictively relevant. The ideal use of parsimony, by contrast, operates when the data are treated as acceptably accurate and complete, and noise is not a significant issue. I'll argue that while the non-ideal use of parsimony enjoys strong support from science, that support doesn't carry over to philosophy. By contrast, the ideal use, though ubiquitous in philosophy, finds only weak support from science. In short, the analogy to science fails to support the kind of parsimony that philosophers actually need. (For ease of exposition, I will refer to the uses of parsimony as, simply, 'ideal parsimony' and 'non-ideal parsimony.')

While other defenses of parsimony's use in philosophy, such as Bayesian approaches, merit their own discussions, those go beyond the scope of this paper. Similarly, those who regard ideal parsimony as intuitively or *a priori* justified, independent of the scientific analogy, will find nothing here directly challenging that view. My aim is not to undermine ideal parsimony as such, but to deny that its use in philosophy can be supported by an analogy to science.

[1] For skeptical takes, see Huemer (2009); Ladyman (2012); Bryant (2020); Bueno & Shalkowski (2020). For optimism, see Paul (2012); Brenner (2017, 2023); Bradley (2018).

https://doi.org/10.1017/psa.2026.10210 Published online by Cambridge University Press

Here is the plan. Section 2 provides relevant background. Section 3 argues that although non-ideal parsimony finds inductive support from science, that support doesn't extend to philosophy. Section 4 argues that while ideal parsimony is exactly the tool philosophers need, it finds little support from science. Section 5 concludes the paper by briefly sketching an alternative approach to vindicating parsimony's use in philosophy.

2 Background

2.1 Parsimony as Minimizing Free Parameters

Before we can discuss parsimony's role in philosophy or science, we need to get clear about the notion under discussion. To illustrate what I have in mind, suppose we want to model the relationship between the number of hours a student studies and their score on a geology exam. We might do this with a simple linear regression, as follows: Score = m * (hours studied) + b. Here, b (the intercept) represents the predicted score for a student who studies zero hours, and m (the slope) represents the predicted increase in score for each additional hour studied. An optimistic model might set $m = 10$ and $b = 80$, predicting large gains per hour studied and an easy exam. A more pessimistic one might set $m = 2$ and $b = 50$, predicting more modest gains and a more challenging exam. In each case, both m and b are parameters: their values are fixed for a given model but can vary across different possible models of student performance. Metaphorically, m and b are the model's "moving parts." In scientific practice, the number of parameters is treated as a measure of a model's complexity—e.g., a linear model, with its two parameters, is less complex than a cubic model, which has four.

Strictly speaking, it is the *free* parameters—independently adjustable quantities or commitments whose values are not fixed by other parts of the model or theory—that contribute to complexity. In the case of the geology exam, m and b are free parameters because the modeler is free to choose their values. By contrast, a *fixed* parameter's value is stipulated in advance by the modeling assumptions and is therefore the same in every admissible model. For example, suppose we want to model the results of an organic chemistry exam, again as a function of hours studied. But in this case, the professor announces that because the material is so difficult, the exams will be scored on a curve such that each student's final score will be 110% of their raw score. We might model this as follows: Score = 1.1 * (m * (hours studied) + b). Here, the multiplier 1.1 is a fixed parameter,

https://doi.org/10.1017/psa.2026.10210 Published online by Cambridge University Press

not something that can be adjusted by the modeler—i.e., any candidate model of the organic chemistry exam will include that exact multiplier.

Fixed parameters do not count toward complexity because they add no degrees of freedom: they cannot be tuned to fit the data differently or to capture new patterns. Also consider that we could always add a fixed parameter by multiplying an equation by 1, but this would do nothing to increase the model's complexity. With this in mind, by "parameter," I will henceforth mean "free parameter."

The parameter-based measure of complexity can be understood to roughly covary with forms of complexity more familiar to philosophers, such as ontological complexity. Adding new kinds of entities to a theory, for instance, typically also requires new parameters to characterize their behavior. Consider that a theory positing only neutrons will require parameters to encode their mass and behavior, whereas a theory that also posits protons and electrons will require an additional parameter for charge. In some cases, more ontologically extravagant theories may require fewer parameters, if the behavior of the richer ontology's entities can be modeled more simply. But all else equal, more ontological commitments entail more parameters. Similarly, enriching a theory's ideology will require adding parameters. For example, adding the notion "spin" requires adding a parameter to encode whether spin is up or down. With this in mind, I will henceforth use 'more parsimonious' to mean "has fewer parameters," unless otherwise specified.

One might object that ontological parsimony cannot be understood in terms of minimizing free parameters, because the number of parameters depends on how a theory is represented.[2] For example, suppose one theory posits electrons, protons, and neutrons, while another posits only a single kind of particle. As Lewis pointed out, in principle, we could represent the richer ontology with a single predicate, "F," that applies to all and only the actual particles.[3] On that representation, both theories require only one parameter. But the richer ontology's apparent economy is merely a representational artifact, not a reflection of its genuine simplicity. For this reason, following Williamson, I treat the counting of parameters as only a rough-and-ready measure of parsimony.[4] As I explain in the following Section, the scientific practice of relying on parameter-based metrics has been extraordinarily successful at preventing overfitting, despite technical issues regarding representation

[2] Thanks to Jill North for helpful exchanges on this point.
[3] For discussion, see Lewis (1983, 367).
[4] See Williamson (2024, 61-62).

https://doi.org/10.1017/psa.2026.10210 Published online by Cambridge University Press

dependence. This raises difficult questions: *why* should a representation-dependent measure reliably avoid overfitting in scientific practice? And what features do our representational schemes have that enable this outcome? While pressing in their own right, I cannot tackle these questions here. For present purposes, what will matter is that the practice of measuring parsimony by counting parameters enjoys significant inductive support as a safeguard against overfitting.

Still, one might argue that parameter-based measures of parsimony don't covary with philosophical notions like ontological parsimony, because the usefulness of representation-dependent measures hinges on whether our representational scheme incorporates the "right" categories. Using the wrong ones—like the single predicate "F"—would undermine parameter minimization's truth-conduciveness. By contrast, traditional philosophical measures of parsimony, we might suppose, are not representation-dependent and thereby avoid this worry. To successfully establish that parameter-based scientific measures of parsimony covary with representation-independent philosophical ones, then, one must first assume that our scientific languages have already installed the correct categories. I set this concern aside, because even if we grant the point for the sake of argument, it offers no help to my opponents. If our scientific language does not make use of the correct categories—whatever precisely that entails—then the analogy between scientific and philosophical parsimony is undermined from the start.

With all of that said, it's worth noting that my subsequent arguments do not depend on a parameter-based notion of parsimony. Even bracketing talk of parameters, adding a new ontological or ideological commitment to one's theory risks overfitting in the manner I describe in Section 3, for the very same reasons I discuss there. However, because the notion of parameters facilitates precise numerical analyses, I will follow working scientists in helping myself to it.

2.2 The State of the Literature

While participants in the debate over parsimony's place in philosophy have appealed to a wide range of considerations, here I am interested in the analogy to science. For present purposes, I will assume some form of realism, because if science does not provide reasons to believe its theories, then the analogy to science would be of little use to philosophers. (Philosophical theses, after all, are meant to be believed.) Since science relies on parsimony, realists may take a theory's parsimony as some reason to believe it, or to assign it a higher

https://doi.org/10.1017/psa.2026.10210 Published online by Cambridge University Press

credence than more complicated rivals. The question, then, is whether parsimony's epistemic grounds extend beyond science. I will follow the literature in centering this discussion around metaphysics, but there is no reason why these issues shouldn't apply in other philosophical areas.

As Brenner puts it, in support of the scientific analogy, those who would deny parsimony's role in metaphysics "should feel compelled to point toward some relevant difference between science and metaphysics which makes it the case that theoretical simplicity is truth conducive in the former context, but not the latter context."[5] It is not enough to simply highlight difficulties for parsimony in metaphysics, because many of those difficulties apply equally well in science, where we are happy to grant its force.

To meet Brenner's challenge, Willard cites parsimony's unique inductive support in science. As she puts it, "scientific theories may be able to justify the use of simplicity by appealing to the progress that science has made, but metaphysics lacks such a theory of progress."[6] This argument depends on the claim that simpler scientific theories have been more successful than their more complex rivals.

Brenner objects to Willard's idea by noting that, although complex theories are often woefully inaccurate, it's always possible to supplement a simple theory with complexifying, *ad hoc* add-ons to salvage its predictive accuracy in the face of would-be counterevidence.[7] Thus, a simple theory cannot be more successful than *all* its more complex rivals, because many of them will make identical predictions to the simpler one. With that in mind, Willard's argument from the unique success of science needs an additional conceptual resource.

We find such a resource in Ladyman's distinction between *strong* and *weak* underdetermination.[8] The former obtains when rival theories are empirically equivalent with respect to all observations for all time; the latter arises only relative to the observations made thus far. Rather than appealing to the "success of science" in general, we can appeal to its success in resolving weak underdetermination *specifically*. Because strong underdetermination involves theories that agree about every actual data point for all time, a

[5] Brenner (2017, 2692).
[6] Willard (2014, 172).
[7] See Brenner (2017, 2701).
[8] According to Ladyman (2012, 43-45), metaphysics confronts strong overdetermination while science deals only with weak underdetermination, severing the analogy. Brenner (2023, 53-54) objects that, in fact, scientists routinely dismiss complex theories, such as the Cartesian evil demon, in the context of strong overdetermination. Even if this suffices to answer Ladyman, it will fail to address the Willard–Ladyman hybrid argument, which specifically concerns predictive success.

https://doi.org/10.1017/psa.2026.10210 Published online by Cambridge University Press

simpler theory *cannot* have a better predictive track record than its strongly underdetermined competitors. Therefore, insofar as we are interested in predictive success, it seems apt to restrict Willard's argument to cases of weak underdetermination.

My aim in the next section is to develop this idea further. The strategy is as follows. Even if, for every simple theory, there exists an equally or more predictively successful complex one, we have no way of knowing in advance *which complex theory that will be*. Suppose, for instance, that we have a set of 100 scientific theories that have so far been equally successful: 99 complex and one simple. Drawing on principles from model selection, I will argue that in any head-to-head comparison, the simpler theory should be preferred, since it is likely to better predict future data—at least under conditions of noisy evidence. This view is perfectly compatible with Brenner's suggestion that one of the complex theories might outperform the simple one in the next round of data. What he overlooks is that preferring a complex scientific theory is akin to buying a lottery ticket that will probably lose.

3 Non-ideal Parsimony

3.1 Noise and Overfitting

Suppose we want to predict an ice cream truck's daily sales. To keep the discussion tractable, assume that temperature is the dominant driver of sales, with higher temperatures generally leading to higher sales. Because this association will generalize beyond the observed data to new cases, it is precisely the sort of predictively relevant structure our model should aim to capture. However, in practice, the observed data exhibit variation around the dominant association, with some hotter days yielding lower sales than cooler days. These departures may arise from a variety of sources. Some, such as a modest weekend-sales effect, may also generalize to future data such that a more sophisticated model might do well to include them. But much of the remaining variation, such as that arising from sales miscounts, is unstable: it does not recur in a way that a model can systematically track. Of course, that does not mean that we cannot find patterns in such variation—patterns are cheap. The problem is that these patterns track idiosyncrasies of the sample rather than structure that supports reliable prediction beyond it. Accordingly, modelers must distinguish between *signal*—predictively exploitable structure in the data—and *noise*—residual variation that fails to support reliable prediction of new data.

https://doi.org/10.1017/psa.2026.10210 Published online by Cambridge University Press

Of course, if the universe is deterministic, then no influences on the data are truly unpredictable *in principle*: nothing surprises Laplace's Demon.[9] But given limited information and other practical constraints, many such influences are, from a modeler's point of view, indistinguishable from objective randomness. For my purposes, variation counts as noise if and only if, given the modeling context, it is effectively random and therefore cannot support reliable prediction by epistemically limited agents. (Of course, variation may fail to support reliable prediction for other reasons, such as if the underlying data-generating process changes.) By contrast, signal consists in patterns that epistemically limited modelers can reliably capture and exploit for prediction.

Overfitting occurs when a model mistakes noise for signal. We can see this clearly in the following fictional models (based on a fictional dataset found in the online-only Appendix) of how an ice cream truck's sales relate to temperature:

The red (overfitted) model fits the existing data better than the linear model, at great cost to simplicity: it has (eight!) more parameters than the linear model. By treating all fluctuations in the data as meaningful, the overfitted model encodes noise as part of the signal. This will impede its predictive capacities, because noise is, by hypothesis, predictively irrelevant. For example, the overfitted model implausibly predicts that ice cream sales will generally be lower when the temperature is 20°C than when it's 18°C.

Simpler models filter out noise. Because noise is, for practical purposes, random, any accidental patterns it contains will likely be weak and unstable, whereas genuine relationships among variables will be stronger and more robust. As a result, a model that ignores weaker patterns in the data is more likely to discard noise than signal, thereby reducing the risk of overfitting.

Of course, simplicity isn't everything—it must be weighed against goodness of fit. There are various metrics for doing this. The following table contains the above models' parameter counts, regression results (for goodness of fit), and two standard criteria that balance fit against complexity:

Unsurprisingly, the linear model performs much worse on the regression. However, both the AIC and BIC favor the linear model anyway, because these criteria penalize complexity. Note that although both criteria favor the linear model, the BIC imposes a stronger

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

https://doi.org/10.1017/psa.2026.10210 Published online by Cambridge University Press

complexity penalty than the AIC. It is unclear whether there are principled reasons to prefer one criterion over the other: the more we penalize complexity, the more likely we are to incidentally penalize models that capture subtle but predictively relevant patterns in the data.[10] Still, all such metrics agree that simplicity must matter in some form, since without it we would be left with hopelessly overfitted models like the ninth-degree polynomial considered here.

I call this use of parsimony "non-ideal" because it is relevant under non-ideal conditions—that is, when the data are noisy. It enjoys strong inductive support: because unnecessarily complex models are prone to encoding random noise as signal, they tend to predict new data poorly, even though they accommodate the existing data extremely well. We should therefore conclude that under noisy conditions, in a head-to-head matchup, the simpler model is more likely to capture predictively relevant patterns than a complex model, all else equal. And even when all else is *not* equal, as the ice cream case illustrates, simplicity can outweigh goodness of fit such that the simpler model is still more likely to track the predictively relevant associations. In sum, under conditions of weak underdetermination, we have reason to believe the simpler model will be more predictively successful. Again, this is fully consistent with the fact that *some* complex model, unidentifiable in advance, would better predict the next data point.

While this discussion has so far concerned models, similar considerations apply to theories. Indeed, according to semantic accounts of theories, "theory" is itself a model-theoretic notion, in which case the foregoing remarks straightforwardly extend to theories.[11] But even if theories are not understood in terms of models, the same basic dynamic recurs. As with models, scientific theories are constructed and evaluated under noisy conditions, requiring us to safeguard against overfitting. That is why scientists include error bars in their figures rather than attempting to vindicate every data point. The presence of noise is also among the reasons that non-conforming data aren't immediately taken to falsify an otherwise good theory.

One might object that the foregoing remarks unduly suggest that we should model the world as simple even when we have reason to believe that it is complex.[12] According to this line of argument, there is nothing intrinsically worrisome about an overfitted model when the

[10] See Hitchcock & Sober (2004, 12-15) for more on the AIC and BIC.
[11] For proponents of the semantic view, see van Fraassen (1980); Giere (1988); Weisberg (2013).
[12] Thanks to an anonymous reviewer for pressing me on this point.

https://doi.org/10.1017/psa.2026.10210 Published online by Cambridge University Press

target system itself is complex. Indeed, recent work in learning theory on "benign overfitting" suggests that, in certain high-dimensional, noisy contexts (e.g., object classification), models that very closely fit the observed data can nonetheless generalize well to new data without any explicit simplicity constraint.[13] If overfitting can be benign—or even optimal—then appeals to non-ideal parsimony as a safeguard against noise may be only conditionally justified rather than generally so.

I agree that when we have good reason to believe the world is complex, we should model it as such. Still, even if we acknowledge "benign" overfitting and permit highly complex models, there must be *some* constraint on complexity. Otherwise, the optimal model would degenerate into a mere list of the observed input-output pairs (along with some unobserved ones), thereby achieving perfect fit with the observed data. The phenomenon of benign overfitting therefore does not show that complexity is cost-free.

Rather than a general vindication of unconstrained model complexity, benign overfitting is a highly selective phenomenon that depends both on the inductive biases of particular learning procedures and on how well those biases happen to match the underlying data-generating process. A model that perfectly fits the observed data can be understood as a solution to multiple learning problems that agree on the observed data but diverge off-sample. Such a model will only generalize beyond the observed data when the learning problem implicitly solved during training coincides with the learning problem that determines reliable prediction in new cases. Outside very specific contexts, highly complex models achieve perfect fit by tracking incidental patterns in the observed data, thereby solving a learning problem that diverges from the one relevant to prediction of new data. Even if the world is highly complex, an overfitted model is likely to make poor predictions, unless the complexity it encodes happens to correspond to stable structure in the data-generating process, not idiosyncrasies of the observed sample. In the absence of independent reason to expect such a correspondence, increasing model complexity increases the risk of fitting noise.

3.2 Biased Philosophical Data

So far, we have established that non-ideal parsimony is truth conducive in science. But that isn't enough to vindicate its use in philosophy, because unlike scientific data, philosophical data aren't significantly noisy. While any conception of philosophical data is apt to be controversial, pre-theoretic intuitions and judgments about thought experiments

[13] See Shamir (2022) for discussion.

https://doi.org/10.1017/psa.2026.10210 Published online by Cambridge University Press

clearly play a central role.[14] For example, an adequate theory of modality should either vindicate pre-theoretically plausible counterfactuals like "if I jumped out the window, I would sustain injuries" or explain away their apparent truth. To justifiably apply non-ideal parsimony to philosophy, our intuitions about such counterfactuals would need to be subject to significant noise. If they were, we would expect substantial random variation in our judgments. But in many cases, including this one, our judgments are strikingly consistent. Indeed, it is precisely the *settled* character of pre-theoretic intuitions that constrains philosophical theorizing. Put simply, intuitive consensus isn't subject to significant noise.[15] Thus, appeals to non-ideal parsimony are generally unwarranted in philosophy.

One might immediately object that intuitive consensus does not entail an absence of significant noise.[16] Even when individual judgments are subject to such noise, aggregation can yield strikingly stable patterns, including reliable averages. According to this line of objection, philosophical intuitions may be no less noisy than scientific measurements, with the apparent stability of our judgments merely reflecting the smoothing effects of aggregation rather than a lack of noise in the underlying data.

This objection misconstrues the (supposed) epistemic role of thought experiments. While this matter is contentious, it's clear that they are not meant to function by aggregating noisy individual responses in the way opinion polls do. To illustrate, suppose we ask a thousand philosophers for their intuitions about whether the protagonist in an epistemological thought experiment knows that *p*, allowing responses on a scale from 1 ("very confident that no") to 10 ("very confident that yes"). Suppose further that 500 respondents answer 1, while the remaining 500 answer 6, resulting in an average of 3.5. Philosophers would not conclude from this that the protagonist doesn't know that *p*. Rather, they would conclude that the thought experiment fails to generate intuitive consensus and is thus dialectically ineffective. The takeaway is that although highly noisy data can generate reliable averages, it cannot generate true intuitive consensus. In its idealized form, an effective thought experiment is meant to elicit broadly convergent judgments, not merely a reliable average across respondents who disagree among themselves.

[14] For a prominent account of philosophical data that vindicates this claim, see Bengson, Cuneo, & Shafer-Landau (2022, 35-76).

[15] As an anonymous reviewer points out, noise is a matter of degree. However, I am not arguing that the philosophical data are *entirely* immune from noise. Rather, I am arguing that whatever noise is present is not sufficient to trigger the methodological role non-ideal parsimony plays in scientific contexts—namely, guarding against overfitting.

[16] Thanks to an anonymous reviewer for pressing me on this point.

https://doi.org/10.1017/psa.2026.10210 Published online by Cambridge University Press

At this point, experimental philosophers will loudly object that pre-theoretic intuitions are in fact far from convergent: survey data often reveal striking variation across cultures, genders, socioeconomic groups, and even the order in which cases are presented.[17] For example, responses to ethical dilemmas vary systematically with wording: people react differently to scenarios framed in terms of 'killing' compared to 'saving,' even when the scenarios are logically identical.[18] Furthermore, experiments have identified large cultural influences on reactions to a variety of canonical philosophical cases, including Gödel cases, the experience machine, and even whether we can know that a zebra before us is not a cleverly painted mule. Far from reflecting a stable consensus, our judgments about thought experiments appear to be highly unreliable, suggesting that philosophical data are noisy after all.

For the sake of argument, I'll grant the experimental philosophers' contention that the philosophical data are highly susceptible to these irrelevant influences. However, such influences are *systematic*. Isn't that exactly why experimental philosophers are so worried about them? Machery even characterizes the practice of reasoning from our pre-theoretic intuitions as "enshrining our prejudices."[19] If the philosophical data are indeed subject to such systematic influences, then the resulting variation doesn't qualify as noise: systematic influences generate stable patterns rather than effectively random variation. In other words, even if experimental findings undermine the epistemic utility of certain thought experiments, they do not thereby indicate the presence of significant noise in the philosophical data.

I will refer to the stably patterned variation resulting from systematic irrelevant influences as *biased*. This terminology is meant to reflect the ordinary use of the term: it is natural to say that a sports fan is biased when their opinions systematically favor their own team for epistemically irrelevant reasons. Similarly, it is natural to say that philosophical data are biased when they systematically track epistemically irrelevant features, such as framing effects. Importantly, although I have been discussing modeling, I do not mean to invoke "bias" as it is used in model selection theory.[20] In that context, bias arises when the same data are used both to fit a model and to assess its predictive accuracy—a form of "double-dipping" that leads to systematically distorted evaluations. Nothing like that is at issue here. My

[17] For a comprehensive discussion of these findings and their implications, see Machery (2017).
[18] See Petrinovich & O'Neill (1996).
[19] See Machery (2017, 126-148).
[20] Thanks to an anonymous reviewer for pushing me to clarify this point.

https://doi.org/10.1017/psa.2026.10210 Published online by Cambridge University Press

concern is instead with systematic, irrelevant influences on the inputs of philosophical theorizing.

Although non-ideal parsimony can effectively safeguard against noise, it offers little protection against bias. To illustrate this point, suppose you manage an ice cream truck enterprise with locations in many cities. You have reliable sales data from only one truck, and you want to use it to estimate how temperature affects demand so as to make informed decisions about where to place new trucks. Inconveniently, the manager of the observed truck has a standing policy of offering substantial discounts on unusually cold days to maintain foot traffic. This policy is not representative of how other locations will be run and, relative to your modeling aim of generalizing across locations, is not predictively useful. Because this irrelevant effect is systematic rather than effectively random, it counts as bias rather than noise.[21]

When a bias is small, simplifying the model may smooth it away. But when a bias is large, parsimony backfires: rather than eliminating the bias, a simpler model will treat it as genuine signal. To illustrate, consider the following fictional model (whose dataset can be found in the online-only Appendix):

Here we see that despite its relative complexity, model selection criteria prefer the cubic model. This shouldn't surprise us, because a bias that large demands accommodation by the model. The upshot: whereas noise can often be mitigated by favoring simpler models, bias calls for other forms of correction, such as counterbalancing, reweighting, or other strategies that directly address the source of distortion.[22]

These considerations apply to philosophical bias, too. If intuitions are biased enough to raise genuine concerns about philosophical methodology, as many experimental philosophers would have it, then those biases must be substantial. Certainly, many of the reported effect sizes are either moderate or large. Taking those data at face value, parsimony ceases to be a safeguard and instead becomes a mechanism for enshrining bias. The proper response to large biases is not to model their outputs more simply, but to correct the distortion at its source or to discard the data altogether.

[21] As an anonymous reviewer points out, this case doesn't involve bias in the technical sense used in model selection, because the new data that we want to predict is sampled from a different probability distribution from the observed data. To reiterate, I use 'bias' here in a more ordinary sense, to refer to systematic, irrelevant influences on the data.

[22] For discussion of such techniques, see Shadish, Cook, & Campbell (2002).

https://doi.org/10.1017/psa.2026.10210 Published online by Cambridge University Press

3.3 Philosophical Misapplications of "Overfitting"

Previous discussions of "overfitting" in philosophy have largely overlooked the distinction between noise and bias. Weinberg, for instance, explicitly conflates them:

> "This is where experimental philosophy becomes highly salient, for the last decade and a half of 'negative program' x-phi indicates a substantial amount of noise in our intuitive data across many areas in philosophy. I won't bother rehearsing it all here: order effects, ethnicity effects, personality trait effects, framing effects, font effects, and so on and on."[23]

Yet none of these factors creates noise, strictly speaking, since they are systematic.

Powers' discussion of moral overfitting reflects a similar problem. Consider her remarks about the possibility that, in his sexism, Kant might have made the categorical imperative sensitive to gender:

> "This new parameter might be the claim of the superiority of the husband to the wife. By adding this parameter, we could input descriptive facts into our model about the ways that women are subjugated within marriages, and the model would output the prediction that this state of affairs is fine, morally speaking.
>
> "Now, clearly we ought not to add this new parameter. And of course, one way—I think the best way!—for us as contemporary readers of Kant to diagnose the problem is to say, look, the output here is just plain wrong, as gendered inequality is morally impermissible. But Kant seems to have lacked epistemic access to this particular moral fact. So perhaps he could have noticed that there seem to be some methodological pitfalls that come with the addition of extra parameters to models. We should be wary of fitting models too closely to all data in general… And we can note that some additions make models complex where they were previously simple, specific where they were general, and flexible where they were more rigid. We have good reasons to at least entertain the possibility that our machinery might be better without such additions."[24]

[23] Weinberg (2017, 259).
[24] Powers (2025, 2729-2730).

https://doi.org/10.1017/psa.2026.10210 Published online by Cambridge University Press

Of course, I agree that we should reject a gender-sensitive categorical imperative. But notice how Powers' reasoning proceeds: she moves directly from the fact that this proposal fits Kant's (sexist) intuitive data too closely to the conclusion that it likely constitutes a case of overfitting. But that inference is unwarranted. We, of course, know that Kant's data had a bias that could be corrected by ignoring the sexist data points. But outright ignoring data we know are bad has nothing to do with parsimony! From Kant's ignorant perspective, those data were legitimate and robustly patterned. Because the effect is so large, there is no guarantee that penalizing complexity would rule out a sexist categorical imperative, so long as the sexist data aren't simply removed (as they should be).[25]

Williamson's discussion of overfitting likewise makes no substantive mention of noise. In arguing that hyperintensional theories in metaphysics are overfitted, he writes that "overfitting is a typical effect of over-confidence in one's data."[26] His mistake is the same as Powers': overconfidence in systematically faulty data may produce bad models, but it does not produce overfitting that is properly addressed with non-ideal parsimony. Perhaps he's right that we should discount the intuitive evidence in favor of hyperintensional theories on the grounds that those intuitions derive from imperfect cognitive heuristics poorly suited to handle such baroque cases. But as in the case of Kant's sexism, the solution there is one we're already familiar with: discount the bad data. Merely simplifying our theories will not magically prioritize fit with the *good* data unless the patterns in the bad data are weak.

I suspect that some philosophers who incorrectly invoke overfitting are motivated by the idea that needing to add a parameter is itself evidence that the data prompting us to do so are faulty. If this were true, then as Powers suggests, Kant should have begun to doubt his sexist intuitions once he realized accommodating them would require adding a parameter to this ethical theory. But before making such inferences, we must ask ourselves: *why* would the need to add a parameter itself be evidence that the data are faulty? I can think of only two plausible reasons. The first is that the data are noisy, but that does not apply in philosophical cases. The second is a broader conviction that simplicity itself tracks truth, but that's a matter of ideal rather than non-ideal parsimony, to be taken up in the next Section.

[25] In fairness, Powers does explicitly say that the solution to this sort of problem is, roughly, to sideline overly complicated thought experiments. She thereby correctly recommends targeting the source of the data itself rather than merely simplifying the model of that data. However, she fails to explain why data from complicated thought experiments should give rise to overfitting in the absence of significant noise.
[26] Williamson (2024, 118).

https://doi.org/10.1017/psa.2026.10210 Published online by Cambridge University Press

Let's take stock. Non-ideal parsimony is justified as a safeguard against overfitting when the data are noisy. In science, we have strong inductive evidence that this strategy is truth-conducive: overfitted models generally yield poorer predictions than simpler ones. But because philosophical data are not noisy in this sense, appeals to non-ideal parsimony in philosophy are generally unwarranted. Anyone who wishes to invoke it must first show that the relevant data exhibit genuine noise rather than systematic bias.

4 Ideal Parsimony

4.1 Ideal Parsimony in Science

Underdetermination by evidence isn't only a feature of noisy contexts. Even if God himself gifted us with a set of unassailable data, there would still be numerous theories compatible with it. Indeed, much philosophical discussion focuses on how to best accommodate the data, whose accuracy is largely taken for granted. Non-ideal parsimony cannot help us in such cases, because these data aren't noisy. Thus, philosophers need some other notion of parsimony—namely, *ideal* parsimony. By way of illustration, consider an influential application of parsimony in the metaphysics of mind: physicalism is widely considered more parsimonious than dualism, since dualism posits an additional kind of state or entity, depending on the version.[27] But the physicalist's invocation of parsimony in no way depends on noise. Rather, they rely on a notion according to which more parsimonious metaphysical theories are more likely to be true even when the data are taken for granted.[28] The situation is much the same for other philosophical topics.

We can now ask: does ideal parsimony find support from science? Sort of. I will argue that appeals to ideal parsimony in science are primarily pragmatic, with only a minor and indirect epistemic dimension. This position remains fully compatible with realism: so long as non-ideal parsimony performs significant epistemic work in science, as I have argued, relegating ideal parsimony to a pragmatic role does not undermine the realist position. Of course, a full treatment of this topic would have to be much more thorough, so my remarks will be selective and schematic, intended primarily to motivate an initially plausible position.

Before I attempt to explain away the uses of ideal parsimony in science, let's get a few on the table. Consider that even after accounting for noise, cognitive scientists have increasingly favored predictive-processing frameworks, which model the brain as a

[27] See Smart (1959, 142); Melnyk (2003, 300).
[28] For further discussion about the relationship between physicalism and parsimony, see Zanotti (2021).

https://doi.org/10.1017/psa.2026.10210 Published online by Cambridge University Press

hierarchically organized system minimizing prediction error between sensory input and top-down expectations. The most ambitious formulations of this idea, such as active inference, seek to unify perception, action, and learning under a single principle of prediction-error minimization.[29] (Of course, proponents of active inference also claim predictive advantages.) The framework thereby houses a diverse range of well-established phenomena—ranging from visual illusions to motor control—under a single, elegant explanatory scheme.

Similar considerations apply to paradigm shifts. Roughly, a paradigm is a comprehensive framework that includes not just theories, but also methods, standards of evidence, and shared assumptions about legitimate scientific practice.[30] Periods of "normal science" proceed within a paradigm, but when evidence accumulates that cannot be reconciled with it, crisis ensues, and a rival paradigm may displace the old one. Crucially, anomalous data consist not of "bad" or objectionably noisy data, but instead of persistent, well-attested findings that resist assimilation into the existing paradigm.[31] In such cases, paradigm shifts aren't a response to noise, so overfitting concerns are irrelevant.

Just as evidence underdetermines theory, it also underdetermines paradigm. Thus, scientists may rely (in part) on ideal parsimony to select a new one. While it is not easy to identify historical cases in which a paradigm was consciously selected over otherwise plausible competitors mainly for its simplicity, note that the paradigms of history—geocentrism, heliocentrism, miasma theory, germ theory, etc.—are all far simpler than many potential alternatives. Perhaps, then, parsimony has played an important implicit role in paradigm selection. If so, then it operates not as a safeguard against noise, but as a criterion for choosing between empirically adequate but competing frameworks.[32]

Scientists may also rely on ideal parsimony when deciding which hypotheses to test. Unexplained deviations in Uranus's orbit generated two broad possibilities. One was parsimonious: posit a single, as-yet-undiscovered planet whose gravitational influence accounted for the anomalies. The other was more complex: modify Newton's law of gravitation or introduce multiple perturbing bodies. Le Verrier and Adams pursued the

[29] For discussion, see Friston (2010); Clark (2013).
[30] For further discussion of paradigms, see Kuhn (1962); Masterman (1970); Wray (2011).
[31] Sometimes, scientific crises arise from persistent discrepancies between measurement techniques. The ongoing tension between local and cosmological measurements of the Hubble constant provides a recent example: while both sides rely on robust, independently verified datasets, their results remain statistically inconsistent. Some explanations of this discrepancy appeal to unaccounted-for interstellar dust, but these are still under investigation and cannot clearly be classified as "bad data." See Freedman (2021) for discussion.
[32] Similar considerations may apply to Lakatos's (1970) notion of "research programs," with the process of switching from one to another guided, in part, by parsimony.

https://doi.org/10.1017/psa.2026.10210 Published online by Cambridge University Press

simpler hypothesis mathematically, leading to Galle's eventual discovery of Neptune.[33] Here, parsimony guided not theory acceptance, but Galle's decision about where to point the telescope.

That such uses of ideal parsimony have often proved fruitful in science might at first glance suggest that it tracks truth. But this appearance is misleading: ideal parsimony's success can be alternatively explained by *pragmatic* factors. I will outline three ways in which investigating parsimonious theories facilitates zetetic efficiency: by minimizing unnecessary reversals in inquiry, being easier to disconfirm, and being more computationally tractable.

4.2 Minimizing Unnecessary Reversals

Kelly argues that favoring simple frameworks "keep[s] us on the straightest possible path to truth."[34] In his framework, we should follow ideal parsimony because it reduces the number of errors, reversals, and delays an inquirer incurs before arriving at the true theory, regardless of whether the truth itself is simple. To see how evaluating simple theories is an efficient means towards truth, let's consider an example in which the truth is complicated. Suppose the true explanation of an orbital anomaly involved three new planets. If scientists leapt to a complex hypothesis that posited five new planets, then as evidence accumulated, they would suffer multiple reversals: first rejecting one imagined planet, then another, perhaps relocating a third, before eventually converging on the correct three. By contrast, if they began with the simplest hypothesis—one new planet—they would start off with fewer commitments. Each new piece of evidence would then force only a single change of course, from one to two planets, and from two to three, rather than repeated backtracking through more elaborate guesses. No matter how complex the truth turns out to be, always preferring the simplest hypothesis results in more efficient inquiry.

A little more formally, Kelly models inquiry as a sequence of conjectures and retractions over an expanding evidential base. Each new observation narrows the range of hypotheses compatible with experience. The "Ockham strategy" is to always choose the simplest hypothesis consistent with the evidence and to retain it until new evidence renders it no longer so. Kelly measures an inquirer's total *loss* as the number of false answers given before convergence, together with the number and timing of reversals. A method is *efficient* if

[33] See Uri (2021) for further historical detail.
[34] Kelly (2007, 563).

https://doi.org/10.1017/psa.2026.10210 Published online by Cambridge University Press

no alternative could guarantee fewer errors or retractions across every possible evidential history. Kelly formally shows that Ockham strategies—and only Ockham strategies—meet this standard. In other words, always preferring and maintaining the simplest available hypothesis minimizes wasted motion on the path to truth: every change of mind is forced by new information rather than by premature complexity. Even when the world itself is complicated, ideal parsimony operates not as an epistemic wager on truth, but as a principle of zetetic efficiency.

As Kelly himself recognizes, his constraint on efficiency requires that no alternative could *guarantee* fewer errors or retractions across *every* possible evidential history. Such a stringent standard of efficiency is far removed from scientific practice.[35] In science, methods are judged not by their optimality across all conceivable worlds, but by their reliability and fruitfulness within the evidential contexts we actually encounter. Thus, Kelly's argument idealizes away the local tradeoffs, background assumptions, and domain-specific heuristics that make parsimony useful in practice. Still, it plausibly illustrates a more general dynamic that often operates in real inquiry, even if no formal learning-theoretic proof applies to realistic cases: when a complex theory turns out to be wrong, it is rarely obvious how to revise it productively. Simpler frameworks, by contrast, leave clearer paths for modification and discovery.

Kelly's defense of parsimony is merely pragmatic, because even if obeying ideal parsimony gets us to the truth the fastest, it does not follow that we have warrant to believe simpler theories in general. Consider that inquiry begins with a relative dearth of evidence. We should expect that the simplest theory compatible with such sparse evidence will be false, because theories based on sparse evidence are usually false. Still, investigating that theory very well might be the most efficient means of ultimately finding the truth. Thus, the pragmatic and epistemic dimension frequently come apart, severely curtailing ideal parsimony's potential applications in philosophy. Even if it minimizes mistaken philosophical commitments on the way to the truth, it doesn't give us much reason to *believe* simpler theories.

4.3 Ease of Disconfirmation

Because complex paradigms let us tinker liberally until our theories fit the data, it is *very* hard to disconfirm them. Just as overfitted models accommodate lots of *existing* data

[35] Thanks to Marc Lange for pressing me on this point.

https://doi.org/10.1017/psa.2026.10210 Published online by Cambridge University Press

points, complex paradigms (and the theories they support) can accommodate lots of *possible* data points. Imagine that you have travelled back to the 5th century and are attempting to convince Emperor Theodosius II that Ptolemaic astronomy is mistaken. You point out that the apparent motion of Mars sometimes fails to match Ptolemy's tables. But the court astronomers assure the emperor that he needn't worry, because Ptolemy's paradigm leaves room for them to add another epicycle or adjust an equant—tuning the model so that it once again fits the observations. Undeterred by your knowledge of history, you continue to point out discrepancies, and the royal astronomers continue providing *ad hoc* fixes. Eventually, the emperor decides to cut off your head as punishment for your backwards ideas about "parsimony."

Now imagine that you are instead speaking to a cousin who grew up among fundamentalists and believes in geocentrism, but has never heard of epicycles. You point out that, contrary to her naïve geocentrism, Venus is sometimes visible as a crescent and sometimes as a nearly full disk. Whereas this would be no problem for Theodosius II's astronomers, your cousin cannot make sense of it: she has no intricate machinery of epicycles and equants to hide behind. The simplest alternative—heliocentrism—is the only one left standing.

Of course, these vignettes are mostly allegorical. But they illustrate the general point that very complex paradigms often undergird complex theories that are nearly impossible to disconfirm. By contrast, when a simple theory is false, disconfirming it is (usually) far less trouble.[36] Insofar as scientists are interested in disconfirmation, they should be interested in simplicity. Of course, *contra* Popper, we shouldn't think disconfirmation or falsification is the whole story of scientific progress. But it remains a crucial feature of inquiry that aims to learn from its mistakes, as science does. Ideal parsimony matters not because simpler theories are more likely, but because it makes error visible: it limits the ways a theory can go wrong and thereby keeps the path to correction open.

Ease of disconfirmation is merely pragmatic: it depends on facts about us and the tools at our disposal. Apart from taking ideal parsimony on faith, we have no antecedent reason to suppose that truth correlates with these things.

[36] This is not to say that simple claims cannot be difficult to disconfirm. For instance, consider the difficulty in disconfirming "there is exactly one pink elephant." Still, making a theory *more* complex will generally make it harder to disconfirm. (Thanks to Jill North for this example.)

https://doi.org/10.1017/psa.2026.10210 Published online by Cambridge University Press

4.4 Computational Tractability

Complex theories and models are incredibly difficult to work with. In astrophysics, for example, the equations governing galaxy formation and stellar evolution are so intricate that their exact solutions remain computationally intractable, even for supercomputers. Researchers therefore rely on heavily simplified models.[37] In this way, parsimony again serves inquiry—not because the universe itself is necessarily simple, but because only simplified models can be rendered tractable enough to yield useful predictions.

4.5 Ideal Parsimony in Philosophy

Pragmatic explanations of ideal parsimony's fruitfulness undermine its claim to track truth, since its success can be explained without appeal to truth-tracking. This is devastating for philosophers who rely on the analogy to science for parsimony's epistemic force: physicalists, for example, should find little comfort in a pragmatic justification of ideal parsimony.

Moreover, it's not clear that ideal parsimony performs any of the pragmatic functions in philosophy that it does in science. Consider reversals: as new, disconfirming evidence accumulates, scientists discard or modify their initial hypotheses until they converge on the truth. In this setting, simplicity minimizes the number of such backtracks. But philosophers have no stream of empirical feedback that overtly disconfirms even our simplest theories.[38] In principle, perhaps it's easier to identify theoretical problems in simpler theories, thereby promoting efficiency. However, the persistence of disagreement in just about every area of philosophy strongly suggests that this possibility is not realized in practice.[39] Indeed, it's not clear that we've made much philosophical progress at all.[40] If parsimony is keeping our philosophical inquiries efficient, I shudder to imagine what inefficient philosophical inquiry would look like. And while there may be some exceptions, empirical disconfirmation and computational tractability are largely irrelevant in philosophy. Thus, none of the pragmatic justifications for parsimony in science apply significantly in philosophy.

Since the foregoing remarks are largely schematic, it's worth offering an additional reason to doubt the philosopher's use of ideal parsimony. For the sake of argument, let's

[37] For further discussion of these models, see Federrath & Offner (2025).
[38] While the philosopher *does* have an endless supply of cases and intuitions, the history of philosophy suggests that this sort of feedback is far less dispositive than the empirical feedback we find in science.
[39] See the PhilPapers survey for evidence of this disagreement (Bourget & Chalmers, 2023).
[40] For discussions of philosophical progress, see Chalmers (2015); Dellsén et al. (2024).

https://doi.org/10.1017/psa.2026.10210 Published online by Cambridge University Press

grant that the use of ideal parsimony in science is at least *somewhat* truth conducive, such that we should assign higher credences to more parsimonious scientific theories. The problem is that because philosophers cannot generally hope for relevant empirical evidence, they must lean *heavily* on ideal parsimony: to secure warrant for their theories, ideal parsimony must be *extremely* truth conducive. But the history of science doesn't bear this out. As Bueno and Shalkowski note, many relatively simple theories have turned out to be false (e.g., Newtonian mechanics), while more complicated ones have turned out to be true (e.g., non-relativistic quantum mechanics).[41] This is typical. So, even if ideal parsimony does grant a scientific theory an epistemic edge, that edge isn't nearly sufficient for warrant. [42] Philosophers who are hoping for warrant, then, will not find reasons for optimism from science. If ideal parsimony is to retain a legitimate role in philosophy, some other justification must be found.

5 Conclusion

In this paper, I have argued that the scientific use of parsimony cannot vindicate its application in philosophy. While non-ideal parsimony enjoys strong inductive support in science as a safeguard against overfitting, that support does not transfer to philosophy, where the relevant data are not significantly noisy. Philosophers instead need ideal parsimony, which finds merely pragmatic or weak epistemic support from science.

Yet abandoning ideal parsimony would amount to abandoning much of philosophical theorizing itself. Philosophers should therefore seek independent grounds on which to motivate their appeals to parsimony. Securing such grounds would not only provide a more stable epistemic basis for philosophical theory choice, but would also supply science with a stronger non-pragmatic rationale for ideal parsimony. Still, there is cause for pessimism: the very appeal of the analogy to science lay in the difficulty of identifying an independent justification for ideal parsimony.

Recent developments in metaphysics suggest a way forward. Schaffer, Bennett, and others have argued that (ideal) parsimony only ranges over the *metaphysically fundamental*.[43] Unlike derivative facts, fundamental facts don't have metaphysical explanations—i.e., they are not fully grounded. We already have reason to believe that fundamental facts are theoretically noxious. For example, if my child asked, "why is the sky blue?" they would not

[41] See Bueno & Shalkowski (2020).
[42] Perhaps philosophers can lean on the many other proposed theoretical virtues (for a taxonomy, see Keas, 2018) to fill the vacuum. But parsimony is clearly the most central and commonly applied of them, suggesting that this strategy will ultimately fall short.
[43] See Schaffer (2015); Bennett (2017); Korman (2015).

https://doi.org/10.1017/psa.2026.10210 Published online by Cambridge University Press

be satisfied to hear "that's just the way it is," because even children know not to settle for fundamental facts. Restricting parsimony to the fundamental promises a principled way of motivating its use in philosophy without appealing to the scientific analogy. The primary challenge for this strategy is to render it compatible with the many uses of parsimony in sciences that concern non-fundamental subject matter, such as in biology.[44]

To deal with this problem, we might understand applications of parsimony in the special sciences as efforts to minimize fundamentality *relative to a domain of inquiry*. If my child asks why tigers have stripes, it will not satisfy them to hear, "the only answer lies in fundamental physics," because we have every reason to expect that the domain of evolutionary biology is reasonably insulated from fundamental physics. Parsimony in the special sciences, then, may be understood as a commitment to the explanatory autonomy of those domains combined with an aversion to positing fundamental facts within them. Just as fundamentality relative to the entire autonomous domain of facts is epistemically suspect, so too is fundamentality within the autonomous domains of the special sciences. If this is right, then the appeal to parsimony across both science and philosophy reflects a common impulse: to seek explanations that invoke as little fundamentality as possible at whatever level of inquiry is under consideration.[45]

Of course, this is only the barest sketch, and many details remain to be spelled out. However, these remarks do suggest that the fundamentality-based picture of parsimony should not be immediately abandoned in the face of its use in the special sciences.

Acknowledgments

Thanks to Marc Lange and Jill North for helpful comments on earlier drafts of this paper. ChatGPT was used to assist with the generation of the figures.

Funding Statement

None to declare.

Declarations

None to declare.

[44] For discussion, see Baron & Tallant (2018, 603-608); Fiddaman & Rodriguez-Pereyra (2018, 351-354).
[45] For a development of this argument, see McIntyre (2026).

https://doi.org/10.1017/psa.2026.10210 Published online by Cambridge University Press

References

Baron, Sam, and Jonathan Tallant. 2018. "Do Not Revise Ockham's Razor Without Necessity." *Philosophy and Phenomenological Research* 96(3): 596–619. https://doi.org/10.1111/phpr.12337.

Bengson, John, Terence Cuneo, and Russ Shafer-Landau. 2022. *Philosophical methodology: From data to theory*. Oxford University Press. https://doi.org/10.1093/oso/9780192862464.001.0001.

Bennett, Karen. 2017. *Making things up*. Oxford University Press. https://doi.org/10.1093/oso/9780199682683.001.0001.

Bourget, David, and David J. Chalmers. 2023. "Philosophers on philosophy: The 2020 PhilPapers survey." *Philosophers' Imprint* 23(11). https://doi.org/10.3998/phimp.2109.

Bradley, Darren. 2018. "Philosophers should prefer simpler theories." *Philosophical Studies* 175(12): 3049–3067. https://doi.org/10.1007/s11098-017-0994-2.

Brenner, Andrew. 2017. "Simplicity as a criterion of theory choice in metaphysics." *Philosophical Studies* 174(11): 2687–2707. https://doi.org/10.1007/s11098-016-0805-1.

Brenner, Andrew. 2023. "Theoretical virtues and the methodological analogy between science and metaphysics." *Synthese* 201(2): 54. https://doi.org/10.1007/s11229-023-04047-z.

Bryant, Amanda. 2020. "Keep the chickens cooped: The epistemic inadequacy of free range metaphysics." *Synthese* 197(5): 1867–1887. https://doi.org/10.1007/s11229-017-1398-8.

Bueno, Otavio, and Scott A. Shalkowski. 2020. "Troubles with Theoretical Virtues: Resisting Theoretical Utility Arguments in Metaphysics." *Philosophy and Phenomenological Research* 101(2): 456–469. https://doi.org/10.1111/phpr.12597.

Chalmers, David J. 2015. "Why isn't there more progress in philosophy?" *Philosophy* 90(1): 3–31. https://doi.org/10.1017/S0031819114000436.

Clark, Andy. 2013. "Whatever Next? Predictive Brains, Situated Agents, and the Future of Cognitive Science." *Behavioral and Brain Sciences* 36(3): 181–204. https://doi.org/10.1017/S0140525X12000477.

https://doi.org/10.1017/psa.2026.10210 Published online by Cambridge University Press

Dellsén, Finnur, Tina Firing, Insa Lawler, and James Norton. 2024. “What is philosophical progress?” *Philosophy and Phenomenological Research* 109(2): 663–693. https://doi.org/10.1111/phpr.13067.

Federrath, Christoph, and Stella Offner. 2025. “Computational advances and challenges in simulations of turbulence and star formation” (Preprint). arXiv. https://doi.org/10.48550/arXiv.2510.12203.

Fiddaman, Mark, and Gonzalo Rodriguez-Pereyra. 2018. “The razor and the laser.” *Analytic Philosophy* 59(3): 341-358. https://doi.org/10.1111/phib.12128.

Freedman, Wendy L. 2021. “Measurements of the Hubble constant: Tensions in perspective.” *The Astrophysical Journal* 919(1): 16. https://doi.org/10.3847/1538-4357/ac0e95.

Friston, Karl. 2010. “The Free-Energy Principle: A Unified Brain Theory?” *Nature Reviews Neuroscience* 11(2): 127–138. https://doi.org/10.1038/nrn2787.

Giere, Ronald N. 1988. *Explaining science: A cognitive approach*. University of Chicago Press.

Hitchcock, Christopher, and Elliott Sober. 2004. “Prediction Versus Accommodation and the Risk of Overfitting.” *The British Journal for the Philosophy of Science* 55(1): 1–34. https://doi.org/10.1093/bjps/55.1.1.

Huemer, Michael. 2009. “When is parsimony a virtue?” *The Philosophical Quarterly* 59(235): 216–236. https://doi.org/10.1111/j.1467-9213.2008.569.x.

Keas, Michael N. 2018. “Systematizing the theoretical virtues.” *Synthese* 195(6): 2761–2793. https://doi.org/10.1007/s11229-017-1355-6.

Kelly, Kevin T. 2007. “A New Solution to the Puzzle of Simplicity.” *Philosophy of Science* 74(5): 561–573. https://doi.org/10.1086/525604.

Korman, Daniel. 2015. “Fundamental ontology and the language of the ontology room.” *Noûs* 49(2): 298–321. https://doi.org/10.1111/nous.12027.

Kuhn, Thomas S. 1962. *The structure of scientific revolutions*. University of Chicago Press.

Ladyman, James. (2012). “Science, metaphysics and method.” *Philosophical Studies* 160(1): 31–51. https://doi.org/10.1007/s11098-012-9910-y.

https://doi.org/10.1017/psa.2026.10210 Published online by Cambridge University Press

Lakatos, Imre. 1970. “Falsification and the Methodology of Scientific Research Programmes.” In *Criticism and the Growth of Knowledge: Proceedings of the International Colloquium in the Philosophy of Science, London*, edited by Imre Lakatos and Alan Musgrave, 91–106. Cambridge University Press.

Lewis, David K. 1983. “New work for a theory of universals.” *Australasian Journal of Philosophy* 61(4): 343–377. https://doi.org/10.1080/00048408312341131.

Machery, Edouard. (2017). *Philosophy within its proper bounds.* Oxford University Press. https://doi.org/10.1093/oso/9780198807520.001.0001.

Masterman, Margaret. 1970. “The Nature of a Paradigm.” In *Criticism and the Growth of Knowledge*, edited by Imre Lakatos and Alan Musgrave, 59–89. Cambridge University Press.

McIntyre, James H. 2026. “Brute Fact Parsimony.” *Philosophical Studies*, advance online publication. https://doi.org/10.1007/s11098-026-02487-0.

Melnyk, Andrew. (2003). *A physicalist manifesto: Thoroughly modern materialism.* Cambridge University Press. https://doi.org/10.1017/CBO9780511498817.

Paul, Laurie A. 2012. “Metaphysics as modeling: The handmaiden’s tale.” *Philosophical Studies* 160(1): 1–29. https://doi.org/10.1007/s11098-012-9906-7.

Petrinovich, Lewis, and Patricia O’Neill. 1996. “Influence of wording and framing effects on moral intuitions.” *Ethology and Sociobiology* 17(3): 145–171. https://doi.org/10.1016/0162-3095(96)00041-6.

Powers, Audrey. 2025. “Moral overfitting.” *Philosophical Studies* 182: 2721–2740. https://doi.org/10.1007/s11098-025-02345-5.

Schaffer, Jonathan. 2015. “What not to multiply without necessity.” *Australasian Journal of Philosophy* 93(4): 644-664. https://doi.org/10.1080/00048402.2014.992447.

Shadish, William R., Thomas D. Cook, and Donald T. Campbell. 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. 2nd ed. Houghton Mifflin.

https://doi.org/10.1017/psa.2026.10210 Published online by Cambridge University Press

Shamir, Ohad. 2022. “The Implicit Bias of Benign Overfitting.” In *Proceedings of the 35th Annual Conference on Learning Theory*, edited by Po-Ling Loh and Maxim Raginsky, 448–478. Vol. 178 of Proceedings of Machine Learning Research.

Smart, John J. C. 1959. “Sensations and brain processes.” *Philosophical Review* 68(2): 141–156. https://doi.org/10.2307/2182164 .

Uri, John. 2021. “175 years ago: Astronomers discover Neptune, the eighth planet.” *NASA*, September 22. https://www.nasa.gov/history/175-years-ago-astronomers-discover-neptune-the-eighth-planet/.

van Fraassen, Bas C. 1980. *The scientific image*. Oxford University Press.

Weinberg, Jonathan M. 2017. “Knowledge, Noise, and Curve-Fitting: A Methodological Argument for Justified True Belief?” In *Explaining Knowledge: New Essays on the Gettier Problem*, edited by Rodrigo Borges, Claudio de Almeida, and Peter D. Klein, 253–272. Oxford University Press. https://doi.org/10.1093/oso/9780198724551.003.0016.

Weisberg, Michael. 2013. *Simulation and similarity: Using models to understand the world.* Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199933662.001.0001.

Willard, Mary B. 2014. “Against simplicity.” *Philosophical Studies* 167(1): 165–181. https://doi.org/10.1007/s11098-013-0228-1.

Williamson, Timothy. 2024. *Overfitting and heuristics in philosophy*. Oxford University Press.

Wray, K. Brad. (2011). “Kuhn and the discovery of paradigms.” *Philosophy of the Social Sciences* 41(3): 380–397. https://doi.org/10.1177/0048393109359778.

Zanotti, Giacomo. (2021). “Physicalism and the burden of parsimony.” *Synthese* 199: 11,109–11,132. https://doi.org/10.1007/s11229-021-03281-7.

https://doi.org/10.1017/psa.2026.10210 Published online by Cambridge University Press

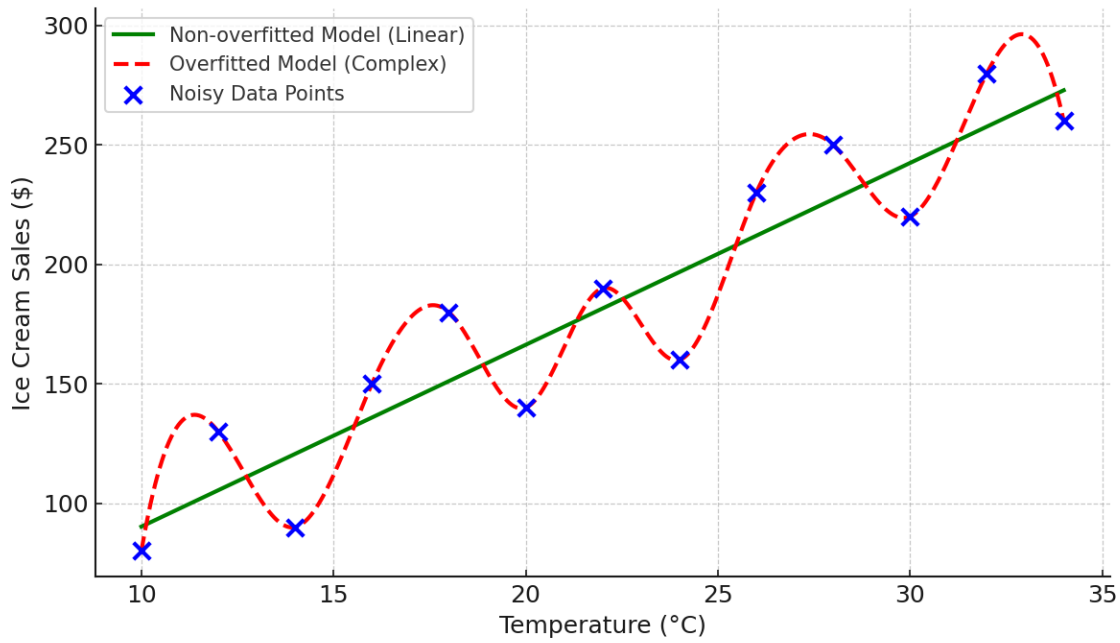


Figure 1. Illustration of Overfitting: Linear vs. Complex Model (noisy data)

(created with the assistance of ChatGPT)

Linear Model: $\hat{y} = 5.51x + 86.08$

Complex Model: $\hat{y} \approx -10^{-11}x^9 - 1.5\times10^{-10}x^8 + \cdots + 7.41x + 67.65$

(Both models fit to 13 fictional data points using ordinary least squares with intercepts.)

https://doi.org/10.1017/psa.2026.10210 Published online by Cambridge University Press

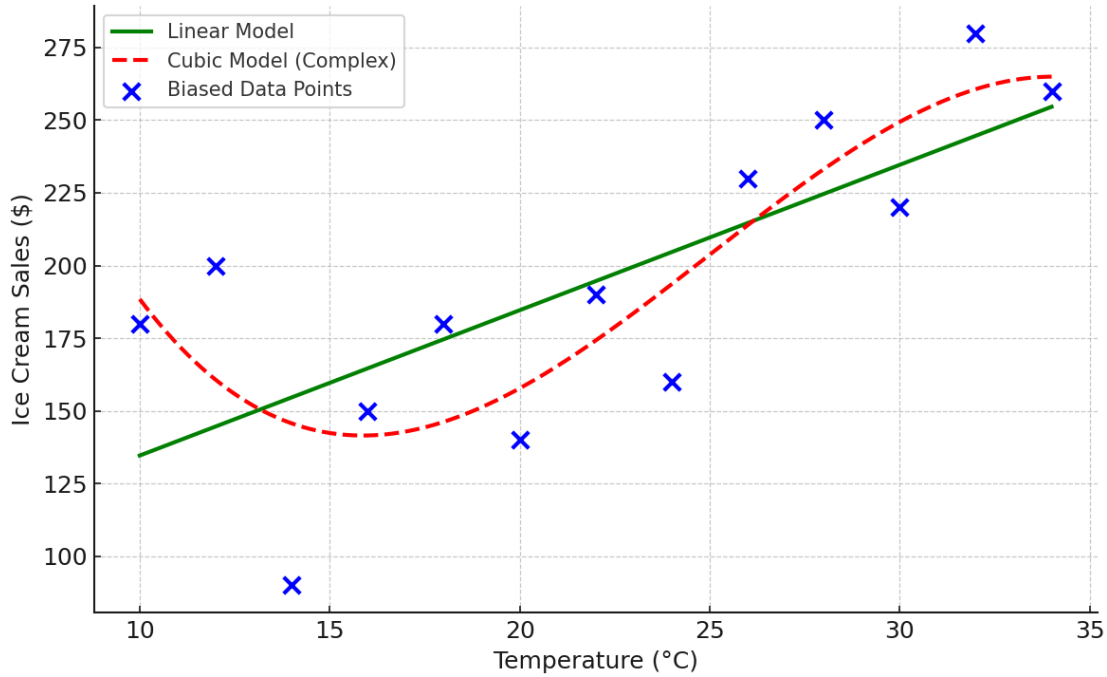


Figure 2. Linear vs. Complex Model (biased data)

(created with the assistance of ChatGPT)

Linear Model: $\hat{y} = 5x + 84.62$

Complex Model: $\hat{y} = -0.0412x^3 + 3.0810x^2 - 66.6879x + 588.372$

(Both models fit to 13 data points using ordinary least squares with intercepts.)

https://doi.org/10.1017/psa.2026.10210 Published online by Cambridge University Press

Table 1. Model Performance Metrics Under Noisy Conditions (lower is better)

Metric	**Linear Model**	**Complex Model**
Parameter Count	2	10
Residual Sum of Squares (RSS)	6810.4	2689.8
Akaike Information Criterion (AIC)	85.4	89.3
Bayesian Information Criterion (BIC)	86.5	95.0

https://doi.org/10.1017/psa.2026.10210 Published online by Cambridge University Press

Table 2. Model Performance Metrics Under Biased Conditions

Metric	**Linear Model**	**Cubic Model**
Parameter Count	2	4
RSS	15923.1	9431.8
AIC	96.4	93.6
BIC	97.6	95.9

https://doi.org/10.1017/psa.2026.10210 Published online by Cambridge University Press