

THE PERILS OF PARSIMONY*

It is widely thought in philosophy and elsewhere that parsimony (or simplicity) is a theoretical virtue in that:

PAR: If T_1 is more parsimonious than T_2 , then T_1 is preferable to T_2 other things being equal.

This admits of many distinct precisifications. I shall focus on the following:

PAR₁: If (a) T_1 is more parsimonious than T_2 and (b) T_1 is strictly equal to T_2 in respects r_1, \dots, r_n taken individually, then $\Pr(T_1 | E \& B) > \Pr(T_2 | E \& B)$.

Here and throughout B is the background information on hand, E is a piece of evidence, and T_1 and T_2 are rival theories each of which potentially explains E .

PAR₁ itself admits of many distinct precisifications. This is because there are many distinct ways of understanding parsimony, and because, further, there are many distinct ways of specifying exactly what respects are included in r_1, \dots, r_n . However, I want to set aside for now both the question of how parsimony is to be understood, and the question of exactly what respects are to be included in r_1, \dots, r_n .¹

I aim to show that *PAR₁* is false. If I succeed in this, then it follows that the same is true of any logically stronger parsimony thesis. For example:

PAR₂: If (a) T_1 is more parsimonious than T_2 and (b) T_1 is *at least roughly* equal to T_2 in r_1, \dots, r_n taken together, then $\Pr(T_1 | E \& B) > \Pr(T_2 | E \& B)$.

PAR₃: If (a) T_1 is more parsimonious than T_2 and (b) T_1 is strictly equal to T_2 in r_1, \dots, r_n taken individually, then *it is rational to believe T_1 and disbelieve T_2 given E and B* .

PAR₂ is like *PAR₁* except that its antecedent is logically weaker than *PAR₁*'s antecedent. *PAR₃*, in turn, is like *PAR₁* except that its consequent is logically stronger than *PAR₁*'s consequent. I am assuming here that it is rational to believe T_1 and disbelieve T_2 given E and B only if $\Pr(T_1 | E \& B) > \Pr(T_2 | E \& B)$. This strikes me as a very weak assumption. I see absolutely no plausibility in the idea that there can be cases where $\Pr(T_1 | E \& B) \leq \Pr(T_2 | E$

* Thanks to Wesley Cray, William Melanson, Michael Roche, Tomoji Shogenji, Elliott Sober, two anonymous referees for this JOURNAL, and the editors for this JOURNAL for helpful comments/discussion.

¹ I shall return to the former question in Section I and the latter question in Section II.

& B) but it is rational to believe T_1 and disbelieve T_2 given E and B .² So, though my focus shall be on PAR_1 , my main point about PAR_1 —that it is false—carries over to parsimony theses other than PAR_1 .

It also carries over to Inference to the Best Explanation understood as:

IBE: If, given B , T_1 is a better potential explanation of E than is T_2 , then it is rational to believe T_1 and disbelieve T_2 given E and B .

I mean for this to be understood so that whether, given B , T_1 is a better potential explanation of E than is T_2 hinges on how T_1 and T_2 score in terms of the so-called “explanatory virtues” (or “theoretical virtues”), and so that T_1 and T_2 are the only available potential explanations of E in the running.³ It follows that *IBE*’s antecedent implies that, given B , T_1 is the best

² Some theorists distinguish between belief and acceptance, and hold that a subject can accept a given hypothesis without believing it. See, for example, Kevin Elliott and David Willmes, “Cognitive Attitudes and Values in Science,” *Philosophy of Science*, LXXX (2013): 807–17. I leave it open that they are right. I also leave it open that there can be cases where $\text{Pr}(T_1 | E \ \& \ B) \leq \text{Pr}(T_2 | E \ \& \ B)$ but it is rational to accept T_1 (but not believe it) and reject T_2 (though not disbelieve it) given E and B .

³ Different theorists give different lists of explanatory virtues. See, for example, James Beebe, “The Abductivist Reply to Skepticism,” *Philosophy and Phenomenological Research*, LXXIX (2009): 605–36; Igor Douven, “Abduction,” in Edward Zalta, ed., *The Stanford Encyclopedia of Philosophy* (Summer 2017), URL = <http://plato.stanford.edu/archives/sum2017/entries/abduction/>, section 2; Gilbert Harman, “The Inference to the Best Explanation,” *Philosophical Review*, LXXIV (1965): 88-95; Peter Kosso, *Reading the Book of Nature: An Introduction to the Philosophy of Science* (Cambridge: Cambridge University Press, 1992), chapter 2; Thomas Kuhn, *The Essential Tension: Selected Essays in Scientific Tradition and Change* (Chicago: University of Chicago Press, 1977), chapter 13; Peter Lipton, *Inference to the Best Explanation*, 2nd ed. (London: Routledge, 2004), chapters 7 and 8; William Lycan, “Explanation and Epistemology,” in Paul Moser, ed., *The Oxford Handbook of Epistemology* (Oxford: Oxford University Press, 2002), pp. 408–33, section 3; Kevin McCain, *Evidentialism and Epistemic Justification* (New York: Routledge, 2014), chapter 6; Ernan McMullin, “The Virtues of Good Theories,” in Stathis Psillos and Martin Curd, eds., *The Routledge Companion to Philosophy of Science* (London: Routledge, 2008), pp. 498–508; Ted Poston, *Reason and Explanation: A Defense of Explanatory Coherentism* (New York: Palgrave Macmillan, 2014), chapters 2 and 4; Stathis Psillos, “Abduction: Between Conceptual Richness and Computational Complexity,” in Peter Flach and Antonis Kakas, eds., *Abduction and Induction: Essays on Their Relation and Integration* (Dordrecht: Kluwer, 2000), pp. 59–74;

available potential explanation of E .⁴ PAR_1 is entailed by PAR_3 , and the latter, in turn, is entailed by IBE. For, any case where PAR_3 's antecedent holds is a case where, given B , T_1 is a better potential explanation of E than is T_2 .⁵ Hence if, as I aim to show, PAR_1 is false, then so too is IBE.

The core idea behind IBE can be put as follows:

*IBE**: Inferences (at least some of them) should be guided at least in part by explanatory considerations.

This is weaker than IBE. It could be that inferences should be guided at least in part by explanatory considerations, but the way in which this should happen is not the way specified by IBE. Nothing in what I aim to show is meant to undermine *IBE**.⁶

It is important to note that many writings on parsimony involve no appeal, explicit or implicit, to PAR_1 (or PAR_2 or PAR_3). This is true, for example, of various writings on parsimony and *divergence from the truth*,⁷ various writings on parsimony and *favoring in the sense of likelihoodism*,⁸ various writings on parsimony and *human cognition*,⁹ various

Stathis Psillos, "Simply the Best: A Case for Abduction," in Antonis Kakas and Fariba Sadri, eds., *Computational Logic: Logic Programming and Beyond* (Berlin: Springer-Verlag, 2002), pp. 605–25; W. V. Quine and J. S. Ullian, *The Web of Belief*, 2nd ed. (New York: Random House, 1978), chapter 6; Richard Swinburne, *Simplicity as Evidence of Truth* (Milwaukee: Marquette University Press, 1997), section III; Paul Thagard, "The Best Explanation: Criteria for Theory Choice," *Journal of Philosophy*, LXXV (1978): 76–92; Timothy Williamson, "Abductive Philosophy," *Philosophical Forum*, XLVII (2016): 263–80.

⁴ It might be that IBE's antecedent should be modified so that it is explicit that, given B , T_1 is a *much* better potential explanation of E than is T_2 , or so that it is explicit that, given B , T_1 is a *satisfactory* (or good enough) potential explanation of E . For discussion and references, see Douven, "Abduction," *op. cit.*, section 2. But this is unimportant for my purposes.

⁵ I am assuming here that whether, given B , T_1 is a better potential explanation of E than is T_2 is fully determined by how T_1 and T_2 score in terms of parsimony and r_1, \dots, r_n .

⁶ The same is true of a variant of IBE developed and defended in Frank Cabrera's "Can There be a Bayesian Explanationism? On the Prospects of a Productive Partnership," *Synthese*, CXCIV (2017): 1245–72, section 5.

⁷ See, for example, Tomoji Shogenji, *Formal Epistemology and Cartesian Skepticism: In Defense of Belief in the Natural World* (New York: Routledge, 2018), chapter 6.

⁸ See, for example, Elliott Sober, *Ockham's Razors: A User's Manual* (Cambridge: Cambridge University Press, 2015).

writings on parsimony and *predictive accuracy*,¹⁰ and various writings on parsimony and *truth-finding efficiency*.¹¹ Nothing in what I aim to show is meant to tell against anything in any such writings.¹²

The remainder of the paper is organized as follows. In Section I, I note some different ways of understanding parsimony. Each is an instance of the more general idea that parsimony is a matter of the number of “things” on/in a theory. In Section II, I argue that PAR_1 is false. In Section III, I discuss the upshot of this negative result. In Section IV, I conclude.

I. PARSIMONY AS A MATTER OF THE NUMBER OF THINGS ON/IN A THEORY

There are many distinct ways of understanding parsimony. Here are but two examples:

P_1 : Parsimony is a matter of the number of entity tokens on a theory.

P_2 : Parsimony is a matter of the number of concept tokens in a theory.

Other examples are like P_1 but where “entity tokens” is replaced by “entity types,” “fundamental entity tokens,” or “fundamental entity types.” Still other examples are like P_2 but where “concept tokens” is replaced by “concept types,” “primitive concept tokens,” or “primitive concept types.”^{13,14}

⁹ See, for example, Nick Chater and Paul Vitányi, “Simplicity: A Unifying Principle in Cognitive Science?,” *Trends in Cognitive Sciences*, VII (2003): 19–22.

¹⁰ See, for example, Malcolm Forster and Elliott Sober, “How to Tell When Simpler, More Unified, or Less Ad Hoc Theories will Provide More Accurate Predictions,” *British Journal for the Philosophy of Science*, XLV (1994): 1–35.

¹¹ See, for example, Kevin Kelly, “A New Solution to the Puzzle of Simplicity,” *Philosophy of Science*, LXXIV (2007): 561–73.

¹² The extant literature on parsimony is vast to say the least. For a helpful overview, and for references, see Alan Baker, “Simplicity,” in Edward Zalta, ed., *Stanford Encyclopedia of Philosophy* (Winter 2016 ed.), URL = <https://plato.stanford.edu/archives/win2016/entries/simplicity/>.

¹³ Since P_1 and P_2 concern *tokens* as opposed to *types*, each is a kind of “quantitative” parsimony as opposed to “qualitative” parsimony. But given that P_1 concerns *entities* whereas P_2 concerns *concepts*, the former is a kind of “ontological” parsimony whereas the latter is a kind of “conceptual” or “ideological” parsimony.

¹⁴ Is there a non-arbitrary way of specifying the entity/concept types on/in a given theory? There are some difficult issues here. For discussion, see Alan Baker, “Occam’s

It is important to note the difference between the expressions “*on* a theory” and “*in* a theory.” If P_1 is assumed, then the issue is the number of entity tokens *posited by a theory* (or, strictly speaking, by a proponent of a theory). If, in contrast, P_2 is assumed, then the issue is the number of concept tokens *in which a theory is formulated* (which might well be different than the number of concept tokens posited by a theory).

I shall remain neutral on how exactly parsimony is to be understood, and shall assume just that:

(*): Parsimony is a matter of the number of things on/in a theory.

This covers P_1 , P_2 , the variants noted above, and much more.¹⁵

Any way of understanding parsimony on which (*) holds can be used to precisify PAR_1 . For example, if P_1 is assumed, then whether T_1 is more parsimonious than T_2 hinges on whether there are fewer entity tokens on T_1 than on T_2 .¹⁶

Razor in Science: A Case Study from Biogeography,” *Biology and Philosophy*, XXII (2007): 193–215, at p. 96; Baker, “Simplicity,” *op. cit.*, section 2; Sam Cowling, “Ideological Parsimony,” *Synthese*, CXC (2013): 3889–908, at p. 3899; Jonathan Schaffer, “What Not to Multiply Without Necessity,” *Australasian Journal of Philosophy*, XCIII (2015): 644–64, at p. 646.

¹⁵ It also covers, for example, the idea that parsimony is a matter of the number of fundamental entity types on a theory *in a particular area of inquiry* (for example, biology).

¹⁶ There is a growing debate on the relative merits of P_1 , P_2 , and variants thereof in the context of theses such as PAR_1 . See, for example, Alan Baker, “Quantitative Parsimony and Explanatory Power,” *British Journal for the Philosophy of Science*, LIV (2003): 245–59; Baker, “Occam’s Razor in Science,” *op. cit.*; Sam Baron and Jonathan Tallant, “Do Not Revise Ockham’s Razor Without Necessity,” *Philosophy and Phenomenological Research*, XCVI (2018): 596–619; Ross Cameron, “How to Have a Radically Minimal Ontology,” *Philosophical Studies*, CLI (2010): 249–64; Richard Caves, “Emergence for Nihilists,” *Pacific Philosophical Quarterly* (forthcoming); Cowling, “Ideological Parsimony,” *op. cit.*; Sam Cowling, *Abstract Entities* (London: Routledge, 2017); Lina Jansson and Jonathan Tallant, “Quantitative Parsimony: Probably for the Better,” *British Journal for the Philosophy of Science*, LXVIII (2017): 781–803; Uriah Kriegel, “The Epistemological Challenge of Revisionary Metaphysics,” *Philosophers’ Imprint*, XIII (2013): 1–30, section 3.1; David Lewis, *Counterfactuals* (Malden: Blackwell, 1973), at p. 87; Daniel Nolan, “Quantitative Parsimony,” *British Journal for the Philosophy of Science*, XLVIII (1997): 329–43; Schaffer, “What Not to Multiply Without Necessity,” *op. cit.*; Theodore Sider, *Writing the Book of the World* (Oxford: Oxford University Press, 2011); Theodore Sider, “Against Parthood,” in Karen Bennett and Dean Zimmerman, eds., *Oxford Studies in*

Is parsimony sensitive to what information is included in B or how E is understood? It might seem that given (*), the answer is negative. Take P_1 for example. It might seem that since the number of entity tokens on T_1 is internal to T_1 and hinges not at all on what information is included in B or how E is understood, and since, similarly, the number of entity tokens on T_2 is internal to T_2 and hinges not at all on what information is included in B or how E is understood, it follows that whether T_1 is more parsimonious than T_2 in the sense of P_1 is *not* sensitive to what information is included in B or how E is understood.

This is a tempting line of thought. But there is a potential problem. Suppose that there is exactly one entity token e_1 on T_1 . Suppose that there are exactly two entity tokens e_2 and e_3 on T_2 . Consider two situations. Situation X is such that B includes the information that e_2 and e_3 exist but does not include the information that e_1 exists. This means that e_1 is “new” relative to B whereas e_2 and e_3 are “old” relative to B . Situation Y , in contrast, is such that B does not include the information that e_1 exists, does not include the information that e_2 exists, and does not include the information that e_3 exists. This means that e_1 , e_2 , and e_3 are all new relative to B . It is not implausible *prima facie* that T_1 is *less* parsimonious than T_2 in Situation X but *more* parsimonious than T_2 in Situation Y . If so, then it is *not* the case that whether T_1 is more parsimonious than T_2 is not sensitive to what information is included in B or how E is understood.

There is another way to look at things however. Perhaps T_1 is more parsimonious than T_2 both in Situation X and in Situation Y (because in each situation there is one entity token on T_1 and two entity tokens on T_2), and perhaps, consistent with this, $T_1 \& B$ is less parsimonious than $T_2 \& B$ in Situation X (because there is one more entity token on the former conjunction than on the latter) but more parsimonious than $T_2 \& B$ in Situation Y (because there is one less entity token on the former conjunction than on the latter).

I want to remain neutral on all this. The important point for my purposes is this:

(**): Take any way of understanding parsimony P_i on which (*) holds. Let “ P_i -things” be the things to be counted when it comes to parsimony on P_i . If (a) there are fewer P_i -things on/in T_1 than on/in T_2 and (b) the things in question are all new relative to B , then T_1 is more parsimonious than T_2 regardless of what information, consistent with (a) and (b), is included in B , and regardless of how E is understood.

Metaphysics, Vol. 8 (Oxford: Oxford University Press, 2013), 237–93; Jeroen Smid, “‘Identity’ as a Mereological Term,” *Synthese*, CXCIV (2017): 2367–85; Jonathan Tallant, “Quantitative Parsimony and the Metaphysics of Time: Motivating Presentism,” *Philosophy and Phenomenological Research*, LXXXVII (2013): 688–705; William Vanderburgh, “Quantitative Parsimony, Explanatory Power and Dark Matter,” *Journal for General Philosophy of Science*, XLV (2014): 317–27.

Consider P_1 and Situation Y for illustration. Given that (a) there is one entity token on T_1 but two on T_2 , and given that (b) the entity tokens in question are all new relative to B , it follows by (**) that T_1 is more parsimonious than T_2 regardless of what information, consistent with (a) and (b), is included in B , and regardless of how E is understood. If, for example, B included the negation of an observational consequence of T_1 , or if E were the negation of an observational consequence of T_1 , then though it would follow that T_1 is false, it would remain the case that T_1 is more parsimonious than T_2 .

I turn now to my critique of PAR_1 .

II. PAR_1 CRITIQUED

This section is divided into five subsections. In Section II.1, I introduce some terminology. In Section II.2, I raise a yes/no question about how PAR_1 is to be understood. In Section II.3, I argue that if the answer is yes, then PAR_1 is false. In Section II.4, I argue that if the answer is no, then, still, PAR_1 is false. In Section II.5, I conclude that PAR_1 is false.

II.1. Terminology. It will help to introduce some terminology. First, $\Pr(T_1 | B)$ and $\Pr(T_2 | B)$ are T_1 's and T_2 's "prior probabilities." These are their probabilities given B (the background information on hand) and thus not taking into account E . If B is tautological, then $\Pr(T_1 | B)$ and $\Pr(T_2 | B)$ are first (or a priori) prior probabilities. But it should not be assumed that B is tautological. The typical case, in fact, is where B is not tautological. Second, $\Pr(E | T_1 \& B)$ and $\Pr(E | T_2 \& B)$ are T_1 's and T_2 's "likelihoods" (with respect to E). These are the probabilities that they confer on E given B . Third, $\Pr(T_1 | E \& B)$ and $\Pr(T_2 | E \& B)$ are T_1 's and T_2 's "posterior probabilities" (with respect to E). These are their probabilities given E and B . They should not be confused with T_1 's and T_2 's likelihoods (though the terminology, which is standard, can be misleading on that front).

Suppose, to illustrate, that B is the proposition that a card is randomly drawn from a standard (and well-shuffled) deck of cards, and let R be the proposition that the card drawn is a Red, H be the proposition that the card drawn is a Heart, and J be the proposition that the card drawn is the Jack of Hearts. Suppose that R is the evidence, and that H and J are the hypotheses. Then:

$$\Pr(H | B) = 1/4 > 1/52 = \Pr(J | B)$$

$$\Pr(R | H \& B) = 1 = \Pr(R | J \& B)$$

$$\Pr(H | R \& B) = 1/2 > 1/26 = \Pr(J | R \& B)$$

Here H 's prior probability of $1/4$ is greater than J 's prior probability of $1/52$, H 's likelihood of 1 is equal to J 's likelihood of 1, and H 's posterior probability of $1/2$ is greater than J 's posterior probability of $1/26$.

I noted above that the typical case is where B is not tautological. The card case just given is a case in point, since B includes information about the makeup of the deck and how the card is drawn.

II.2. A Yes/No Question about How to Understand PAR₁. Is PAR₁ to be understood so that prior probability and likelihood are included in r_1, \dots, r_n ? This is the yes/no question alluded to above.

There is a natural reading of PAR₁ on which the answer is yes. First, suppose that each of r_1, \dots, r_n is an explanatory virtue distinct from parsimony. Second, suppose that prior probability and likelihood are explanatory virtues distinct from parsimony. Then it follows that prior probability and likelihood are included in r_1, \dots, r_n .¹⁷

This reading of PAR₁ is suggested in part by numerous passages in the literature in which it is claimed in effect (though in different terminology) that prior probability and likelihood are explanatory virtues. Here, first, are some passages in which it is claimed in effect (given the surrounding context) that prior probability is an explanatory virtue:

- (A) ... *fit with background* is also an inferential factor that has an explanatory aspect. One reason this is so is because *background beliefs* may include beliefs about what sorts of accounts are genuinely explanatory. For example, at given stages of science no appeal to action at a distance or to an irreducibly chance mechanism could count as an adequate explanation, whatever its empirical adequacy. The role of *background belief* in *determining the quality of an explanation* shows how explanatory virtue is 'contextual', since the

¹⁷ There are probabilistic measures of "explanatory power" on which T_1 is equal to T_2 in explanatory power with respect to E given B if $\Pr(E | T_1 \& B) = \Pr(E | T_2 \& B)$ and $\Pr(T_1 | B) \neq \Pr(T_2 | B)$. See Vincenzo Crupi and Katya Tentori, "A Second Look at the Logic of Explanatory Power (with Two Novel Representation Theorems)," *Philosophy of Science*, LXXIX (2012): 365–85; Jonah Schupbach, "Comparing Probabilistic Measures of Explanatory Power," *Philosophy of Science*, LXXVIII (2011): 813–29; Jonah Schupbach and Jan Sprenger, "The Logic of Explanatory Power," *Philosophy of Science*, LXXVIII (2011): 105–27. If explanatory virtue is a matter of explanatory power thus understood, then prior probability is not an explanatory virtue. But there is a wider sense of "explanatory virtue" on which an explanatory virtue is simply a "theoretical virtue." This is the sense at issue here (in the reading of PAR₁ under consideration) and throughout.

same hypothesis may provide a lovely explanation in one theoretical milieu but not be explanatory in another ...¹⁸

- (B) H will be preferred to H' if H *fits better with what we already believe*. If this sounds dogmatic or pigheaded, notice again that, inescapably, we never even consider competing hypotheses that would strike us as grossly implausible; the detective would never so much as entertain the hypothesis that the crime was committed by invisible Venusian invaders, nor the mechanic that your car trouble is caused by an infusion of black bile or evil fairy dust. Nor should we consider such hypotheses, even if we could enumerate them all; someone who insisted on doing so would be rightly accused of wasting everyone's time. All inquiry is conducted against a *background of existing beliefs*, and we have no choice but to rely on some of them while modifying or abandoning others—else how could any such revisions be motivated?¹⁹

Here, second, are some passages in which it is claimed in effect (given the surrounding context) that likelihood is an explanatory virtue:

- (C) On the a posteriori side [of the divide between a priori criteria and a posteriori criteria for determining the best potential explanation] there is, first, the criterion of *yielding the data*, that is, *leading us to expect the events to be explained*—either with deductive certainty or with inductive probability. The more data and the more probable some hypothesis renders their occurrence, the more likely it is that—that is, the more probable it is that—that hypothesis is true.²⁰
- (D) In the best case, T explains E by entailing E. More typically, T must be conjoined with auxiliary hypotheses to entail E. Those auxiliary hypotheses must be evaluated in turn; they may be independently plausible, or the abductive assessment may have to be of their conjunction with T. Obviously the auxiliary hypotheses should not entail E by themselves, otherwise T would be redundant. In still other cases, the connection is irremediably probabilistic: E is probable conditional on T, perhaps in conjunction with auxiliary hypotheses, which as before must themselves be evaluated, and

¹⁸ Lipton, *Inference to the Best Explanation*, *op. cit.*, pp. 122–23, emphasis added.

¹⁹ Lycan, “Explanation and Epistemology,” *op. cit.*, p. 416, emphasis added.

²⁰ Swinburne, *Simplicity as Evidence of Truth*, *op. cit.*, p. 18, emphasis added.

should not render T redundant. At a bare minimum, T must be consistent with E. In brief, *the closer T comes to entailing E, the better* (ceteris paribus).²¹

It might be that the data in question in passage (C) go beyond E and include data given in B. But that would be okay for my purposes. It would still be the case that at least part of the claim is that whether, given B, T_1 is a better potential explanation of E than is T_2 is determined at least in part by T_1 's and T_2 's likelihoods.

However, I do not want to insist on this way of understanding PAR₁. I do not even want to insist on understanding it so that prior probability and likelihood are included in r_1, \dots, r_n . I aim to show that PAR₁ is false regardless of whether it is understood so that prior probability and likelihood are included in r_1, \dots, r_n .

II.3. PAR₁ Understood so that Prior Probability and Likelihood are Included in r_1, \dots, r_n . It is a theorem of the probability calculus that:

$$(1) \quad \frac{\Pr(T_1 | E \& B)}{\Pr(T_2 | E \& B)} = \frac{\Pr(T_1 | B)}{\Pr(T_2 | B)} \times \frac{\Pr(E | T_1 \& B)}{\Pr(E | T_2 \& B)}$$

This means that the ratio of T_1 's posterior probability to T_2 's posterior probability equals the product of (a) the ratio of T_1 's prior probability to T_2 's prior probability and (b) the ratio of T_1 's likelihood to T_2 's likelihood.²²

Suppose that PAR₁ is understood so that prior probability and likelihood are included in r_1, \dots, r_n . Take some case where PAR₁'s antecedent is true. It follows that T_1 is equal to T_2 both in prior probability and in likelihood:

$$(2) \quad \Pr(T_1 | B) = \Pr(T_2 | B)$$

$$(3) \quad \Pr(E | T_1 \& B) = \Pr(E | T_2 \& B)$$

But then by (1) it follows that, contra PAR₁'s consequent, T_1 is equal to T_2 in posterior probability:

$$(4) \quad \Pr(T_1 | E \& B) = \Pr(T_2 | E \& B)$$

This means that PAR₁ is false if it is understood so that prior probability and likelihood are included in r_1, \dots, r_n .

²¹ Williamson, "Abductive Philosophy," *op. cit.*, p. 266, emphasis added.

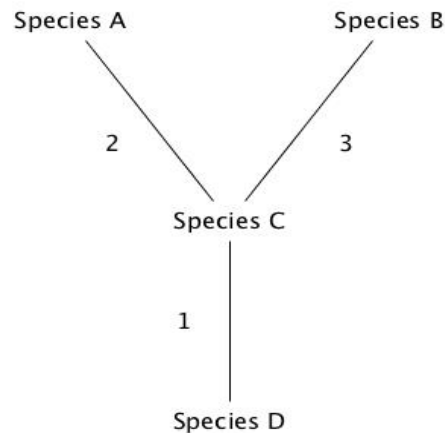
²² (1) is a simple consequence of Bayes's theorem.

This is true whether or not PAR_1 is understood so that each of r_1, \dots, r_n is an explanatory virtue distinct from parsimony, and whether or not it is understood so that prior probability and likelihood are explanatory virtues distinct from parsimony. If, for whatever reason, PAR_1 is understood so that prior probability and likelihood are included in r_1, \dots, r_n , then PAR_1 is false.

It is not just that if PAR_1 is understood so that prior probability and likelihood are included in r_1, \dots, r_n , then PAR_1 is open to counterexample. If PAR_1 is thus understood, then *all* cases where its antecedent is true are cases where its consequent is false.

II.4. PAR_1 Understood so that it is Not the Case that Prior Probability and Likelihood are Included in r_1, \dots, r_n . Suppose that PAR_1 is understood so that it is not the case that prior probability and likelihood are included in r_1, \dots, r_n . Suppose, that is, that it is understood so that (a) prior probability is not included in r_1, \dots, r_n or (b) likelihood is not included in r_1, \dots, r_n (where “or” here is inclusive). What then?

It will help to consider an example.²³ Suppose that you know of Species A and Species B, but you are ignorant as to whether the members of either species have wings. Suppose further that you know that Species A and Species B are related as depicted in the graph below:



Species C is the “parent” of Species A and Species B (which are the “children” of Species C). Species D, in turn, is the parent of Species C (which is the child of Species D). You consider two theories. T_1 is the theory that there is exactly *one* branch on which there was an evolutionary change (from a parent with no wings to a child with wings, or from a parent with wings to a child with no wings). T_2 is the theory that there are exactly *two* branches on

²³ This example is adapted from Elliott Sober’s “Let’s Razor Ockham’s Razor,” in Dudley Knowles, ed., *Explanation and Its Limits* (Cambridge: Cambridge University Press, 1990), section 3.

which there was an evolutionary change. There are fewer branches on which there was an evolutionary change on T_1 than on T_2 . Given this, and supposing that the branches in question (understood as branches on which there was an evolutionary change) are all new relative to B , it follows by (**) that T_1 is more parsimonious in the sense of P_1 than T_2 .²⁴

Let C_i be the proposition that there was an evolutionary change on branch i (for any $i = 1, 2, 3$). There are three ways for T_1 to be true, and three ways for T_2 to be true. The former are:

$$w_1 \quad C_1 \ \& \ \sim C_2 \ \& \ \sim C_3$$

$$w_2 \quad \sim C_1 \ \& \ C_2 \ \& \ \sim C_3$$

$$w_3 \quad \sim C_1 \ \& \ \sim C_2 \ \& \ C_3$$

The latter are:

$$w_4 \quad C_1 \ \& \ C_2 \ \& \ \sim C_3$$

$$w_5 \quad C_1 \ \& \ \sim C_2 \ \& \ C_3$$

$$w_6 \quad \sim C_1 \ \& \ C_2 \ \& \ C_3$$

Suppose that B includes the information that:

- (5) The probability of an evolutionary change on a given branch is 0.5 regardless of whether there were evolutionary changes on other branches, and regardless of whether the members of the parent species (if there is a parent species) had wings.

It follows that T_1 is equal to T_2 in prior probability:

$$\begin{aligned} (6) \quad \Pr(T_1 | B) &= \Pr(w_1 | B) + \Pr(w_2 | B) + \Pr(w_3 | B) \\ &= (3) \times (0.5)(0.5)(0.5) \\ &= 0.375 \end{aligned}$$

²⁴ I mean for the expression “entity tokens” in P_1 to be understood broadly so as to include object tokens, event tokens, fact tokens, process tokens, and property tokens.

$$\begin{aligned}
(7) \quad \Pr(T_2 | B) &= \Pr(w_4 | B) + \Pr(w_5 | B) + \Pr(w_6 | B) \\
&= (3) \times (0.5)(0.5)(0.5) \\
&= 0.375
\end{aligned}$$

Now let E be the proposition that the members of Species A and the members of Species B have wings, and suppose that B also includes the information that:

$$(8) \quad \text{The members of Species D did not have wings.}$$

It follows that T_1 is equal to T_2 in likelihood:

$$\begin{aligned}
(9) \quad \Pr(E | T_1 \& B) &= \Pr(w_1 | T_1 \& B) \\
&= (0.5)(0.5)(0.5) \\
&= 0.125
\end{aligned}$$

$$\begin{aligned}
(10) \quad \Pr(E | T_2 \& B) &= \Pr(w_6 | T_2 \& B) \\
&= (0.5)(0.5)(0.5) \\
&= 0.125
\end{aligned}$$

This is a case, then, where (i) T_1 is more parsimonious in the sense of P_1 than T_2 and (ii) T_1 is equal to T_2 both in prior probability and in likelihood.

It is important to understand *why* there can be cases like this. The key is (***) from Section I (which I repeat below for convenience):

(**): Take any way of understanding parsimony P_i on which (*) holds. Let “ P_i -things” be the things to be counted when it comes to parsimony on P_i . If (a) there are fewer P_i -things on/in T_1 than on/in T_2 and (b) the things in question are all new relative to B , then T_1 is more parsimonious than T_2 regardless of what information, consistent with (a) and (b), is included in B , and regardless of how E is understood.

Since (a) there are fewer branches on which there was an evolutionary change on T_1 than on T_2 and (b) the branches in question are all new relative to B , it follows by (***) that T_1 is more parsimonious in the sense of P_1 than T_2 regardless of what information, consistent with (a) and (b), is included in B , and regardless of how E is understood. Given this, and given that (5) and (8) are consistent with (a) and (b), it follows that there can be cases where (a) and (b) hold, (5) and (8) are included in B , and E is understood as the proposition that the members of Species A and the members of Species B have wings. This means, though, that

there can be case where (i) T_1 is more parsimonious in the sense of P_i than T_2 and (ii) T_1 is equal to T_2 both in prior probability and in likelihood.

This generalizes to *any* other way of understanding parsimony P_i on which (*) holds. Suppose that (a) there are fewer P_i -things on/in T_1 than on/in T_2 and (b) the things in question are all new relative to B . It follows by (**) that B can be further specified and E can be understood in such a way that (i) T_1 is more parsimonious in the sense of P_i than T_2 and (ii) T_1 is equal to T_2 both in prior probability and in likelihood.

This is problematic for PAR_1 understood so that it is not the case that prior probability and likelihood are included in r_1, \dots, r_n . For, since, surely, at least some of the cases in question—cases where (i) T_1 is more parsimonious in the sense of P_i than T_2 and (ii) T_1 is equal to T_2 both in prior probability and in likelihood—are cases where T_1 is strictly equal to T_2 in r_1, \dots, r_n taken individually, it follows that there are cases where PAR_1 's antecedent is true but its consequent is false because T_1 is equal to T_2 in posterior probability.

I am assuming here that any proposed specification of r_1, \dots, r_n should meet the condition that there are cases where (i) T_1 is more parsimonious than T_2 and (ii) T_1 is strictly equal to T_2 in r_1, \dots, r_n taken individually. For, if r_1, \dots, r_n were specified so that there are no such cases, then this would trivialize PAR_1 (by making its antecedent inconsistent).²⁵

The result, then, is that PAR_1 is false even if it is understood so that it is not the case that prior probability and likelihood are included in r_1, \dots, r_n .

II.5. Summary. In Section II.2, I raised the question of whether PAR_1 is to be understood so that prior probability and likelihood are included in r_1, \dots, r_n . In Section II.3, I argued that if the answer is yes, then PAR_1 is false, because then its antecedent *requires* that T_1 be equal to T_2 both in prior probability and in likelihood, and thus, by (1), its antecedent requires that

²⁵ It would not be enough to save PAR_1 (on the current way of understanding it) to insist that though there are cases where (i) T_1 is more parsimonious than T_2 and (ii) T_1 is strictly equal to T_2 in r_1, \dots, r_n taken individually, all such cases are cases where T_1 is *not* equal to T_2 in prior probability or likelihood. It would need to be insisted further that all such cases are cases where T_1 is *superior* to T_2 in prior probability or likelihood. For, given (1), if at least some of the cases in question are cases where, say, T_1 is equal to T_2 in prior probability and *inferior* to it in likelihood, then there are cases where PAR_1 's antecedent is true but its consequent is false because T_1 is *inferior* to T_2 in posterior probability. But, given (**), I see absolutely no plausibility in the claim that all cases where (i) T_1 is more parsimonious than T_2 and (ii) T_1 is strictly equal to T_2 in r_1, \dots, r_n taken individually are cases where T_1 is superior to T_2 in prior probability or likelihood. If friends of PAR_1 believe otherwise, then I urge them to explain themselves. What exactly is included in r_1, \dots, r_n in PAR_1 , and how is it that despite (**), all cases where (i) T_1 is more parsimonious than T_2 and (ii) T_1 is strictly equal to T_2 in r_1, \dots, r_n taken individually are cases where T_1 is superior to T_2 in prior probability or likelihood?

T_1 be equal to T_2 in posterior probability. In Section II.4, I argued that if, instead, the answer is no, then, still, PAR_1 is false, since then its antecedent at least *allows* that T_1 be equal to T_2 both in prior probability and in likelihood, and thus, by (1), its antecedent at least allows that T_1 be equal to T_2 in posterior probability. I conclude that PAR_1 is false.²⁶

III. DISCUSSION

This section is divided into three subsections. In Section III.1, I consider how, if not by appeal to PAR_1 , it can be shown that the more parsimonious theory in a given case has the higher posterior probability. In Section III.2, I turn to the idea that central to the aim of science is parsimony understood in terms of the number of brute phenomena on our overall world picture. In Section III.3, I suggest a potential way forward for friends of IBE.

III.1. Higher Posterior Probabilities. PAR_1 is false (and so are parsimony theses such as PAR_2 and PAR_3). What now? How, if not by appeal to PAR_1 , is it to be shown that the more parsimonious theory in a given case has the higher posterior probability?

It follows from (1) in Section II.3 that each of the following holds without exception:

(11) If (a) $\Pr(T_1 | B) > p(T_2 | B)$ and (b) $\Pr(E | T_1 \ \& \ B) \geq p(E | T_2 \ \& \ B)$, then $\Pr(T_1 | E \ \& \ B) > \Pr(T_2 | E \ \& \ B)$.

(12) If (a) $\Pr(T_1 | B) \geq \Pr(T_2 | B)$ and (b) $\Pr(E | T_1 \ \& \ B) > p(E | T_2 \ \& \ B)$, then $\Pr(T_1 | E \ \& \ B) > \Pr(T_2 | E \ \& \ B)$.

Each of these theses provides a way to show that the more parsimonious theory in a given case has the higher posterior probability.²⁷

²⁶ PAR_1 is false regardless of the subject matter. It is not the case, for example, that PAR_1 is false in philosophy but true in science. It is false across the board. For relevant discussion, see Baron and Tallant, “Do Not Revise Ockham’s Razor Without Necessity,” *op. cit.*, section 3.4; Andrew Brenner, “Simplicity as a Criterion of Theory Choice in Metaphysics,” *Philosophical Studies* (forthcoming); Michael Huemer, “When is Parsimony a Virtue?,” *Philosophical Quarterly*, LIX (2009): 216–36; Kriegel, “The Epistemological Challenge of Revisionary Metaphysics,” *op. cit.*; L. A. Paul, “Metaphysics as Modeling: The Handmaiden’s Tale,” *Philosophical Studies*, CLX (2012): 1–29; Sober, *Ockham’s Razors: A User’s Manual*, *op. cit.*, chapter 5; M. B. Willard, “Against Simplicity,” *Philosophical Studies*, CLXVII (2014): 165–81.

²⁷ Jansson and Tallant, “Quantitative Parsimony: Probably for the Better,” *op. cit.*, section 3, give in effect a special case of (11). They give a general condition, where T_1 is more parsimonious than T_2 in the sense of P_1 , under which $\Pr(T_1 | B) > \Pr(T_2 | B)$ and $\Pr(E |$

Return to the case of Species A and Species B from Section II.4. Suppose that (5) is replaced by:

- (13) The probability of an evolutionary change on a given branch is 0.1 regardless of whether there were evolutionary changes on other branches, and regardless of whether the members of the parent species (if there is a parent species) had wings.

It follows that T_1 's prior probability is higher than T_2 's prior probability:

$$\begin{aligned} (14) \quad \Pr(T_1 | B) &= \Pr(w_1 | B) + \Pr(w_2 | B) + \Pr(w_3 | B) \\ &= (3) \times (0.1)(0.9)(0.9) \\ &= 0.243 \end{aligned}$$

$$\begin{aligned} (15) \quad \Pr(T_2 | B) &= \Pr(w_4 | B) + \Pr(w_5 | B) + \Pr(w_6 | B) \\ &= (3) \times (0.1)(0.1)(0.9) \\ &= 0.027 \end{aligned}$$

It also follows that T_1 's likelihood is higher than T_2 's likelihood:

$$\begin{aligned} (16) \quad \Pr(E | T_1 \& B) &= \Pr(w_1 | T_1 \& B) \\ &= (0.1)(0.9)(0.9) \\ &= 0.081 \end{aligned}$$

$$\begin{aligned} (17) \quad \Pr(E | T_2 \& B) &= \Pr(w_6 | T_2 \& B) \\ &= (0.1)(0.1)(0.9) \\ &= 0.009 \end{aligned}$$

Hence both by (11) and by (12) it follows that T_1 's posterior probability is higher than T_2 's posterior probability.

I am not claiming that the only way to show that the more parsimonious theory in a given case has the higher posterior probability is by appeal to (11) or (12). Suppose, for

$T_1 \& B) = 1 = \Pr(E | T_2 \& B)$. It can be put as follows: (a) T_1 entails E , (b) T_2 is a conjunction $T_3 \& T_4$, (c) T_2 entails E but T_3 does not, (d) $\Pr(T_1 | B) \geq \Pr(T_3 | B)$, and (e) $\Pr(T_4 | T_3 \& B) < 1$. They also consider three scientific cases, and argue that their condition holds in each of them.

example, that you grant for the sake of argument that T_1 's prior probability is lower than T_2 's prior probability, and yet you show that T_1 's prior probability is nonetheless *roughly equal* to T_2 in prior probability, and that T_1 's likelihood is *much higher* than T_2 's likelihood:

$$(18) \quad \frac{\Pr(T_1 | B)}{\Pr(T_2 | B)} \approx 1$$

$$(19) \quad \frac{\Pr(E | T_1 \& B)}{\Pr(E | T_2 \& B)} \gg 1$$

Then, by (1), you thereby show that T_1 's posterior probability is higher than T_2 's posterior probability.

There are two main lessons here. First, it is true that oftentimes the more parsimonious theory has the higher posterior probability, but this is not because of parsimony *in itself*. It is because of the background information on hand, and the evidence to be explained. Second, and relatedly, if you want to show that the more parsimonious theory in a given case has the higher posterior probability, then you should look not to PAR_1 , but instead to theses like (11) and (12).

The idea that background information is all-important in the context of parsimony is not new. Elliott Sober stresses and develops it in numerous writings.²⁸ But it has yet to be fully appreciated. It is high time for a change.

III.2. The Aim of Science. Albert Einstein holds that parsimony of a certain kind is central to the aim of science.²⁹ Here he states the view:

The aim of science is, on the one hand, a comprehension, as complete as possible, of the connection between the sense experiences in their totality, and, on the other hand, the accomplishment of this aim by the use of *a minimum of primary concepts and relations*.³⁰

Here he elaborates:

²⁸ See Elliott Sober, *Reconstructing the Past: Parsimony, Evolution, and Inference* (Cambridge, MA: MIT Press, 1988); Sober, "Let's Razor Ockham's Razor," *op. cit.*; Sober, *Ockham's Razors: A User's Manual*, *op. cit.* See also Mike Dacey, "The Varieties of Parsimony in Psychology," *Mind & Language*, xxxi (2016): 414–37; Anya Plutynski, "Parsimony and the Fisher-Wright Debate," *Biology and Philosophy*, xx (2005): 697–713.

²⁹ Albert Einstein, *Ideas and Opinions* (New York: Three Rivers, 1995).

³⁰ Einstein, *Ideas and Opinions*, *op. cit.*, p. 293, emphasis added.

Science uses the totality of the primary concepts, i.e., concepts directly connected with sense experiences, and propositions connecting them. In its first stage of development, science does not contain anything else. Our everyday thinking is satisfied on the whole with this level. Such a state of affairs cannot, however, satisfy a spirit which is really scientifically minded; because the totality of concepts and relations obtained in this manner is utterly lacking in logical unity. In order to supplement this deficiency, one invents a system *poorer in concepts and relations*, a system retaining the primary concepts and relations of the “first layer” as logically derived concepts and relations. This new “secondary system” pays for its higher logical unity by having elementary concepts (concepts of the second layer), which are no longer directly connected with complexes of sense experiences. Further striving for logical unity brings us to a tertiary system, *still poorer in concepts and relations*, for the deduction of the concepts and relations of the secondary (and so indirectly of the primary) layer. Thus the story goes on until we have arrived at a system of the greatest conceivable unity, and of *the greatest poverty of concepts of the logical foundations*, which is still compatible with the observations made by our senses.³¹

If Einstein is right about the aim of science, then there are at least two questions to be answered. Is there a rationale behind the aim of science? Can the aim of science be reached without appeal to PAR₁?

I have no definite answer to the first of these questions. But I suspect that many theorists would answer it by appeal to the idea that greater parsimony of the kind at issue makes for greater scientific understanding. This idea is nicely articulated by Michael Friedman.³² He writes:

The kinetic theory effects a significant unification in what we have to accept. Where we once had three independent brute facts—that gases approximately obey the Boyle-Charles law, that they obey Graham’s law, and that they have the specific-heat capacities they do have—we now have only one—that molecules obey the laws of mechanics. Furthermore, the kinetic theory also allows us to integrate the behavior of gases with other phenomena, such as the motions of the planets and of falling bodies near the earth. This is because the laws of mechanics also permit us to derive both the fact that planets obey Kepler’s laws and the fact that falling bodies obey Galileo’s laws. From the fact that all bodies obey the laws of mechanics it follows that the planets behave as they do, falling bodies behave as they do, and gases behave as they do. Once again, we have

³¹ Einstein, *Ideas and Opinions*, *op. cit.*, pp. 293–94, emphasis added.

³² Michael Friedman, “Explanation and Scientific Understanding,” *Journal of Philosophy*, LXXI (1974): 5–19.

reduced a multiplicity of unexplained, independent phenomena to one. I claim that this is the crucial property of scientific theories we are looking for; this is the essence of scientific explanation—*science increases our understanding of the world by reducing the total number of independent phenomena that we have to accept as ultimate or given. A world with fewer independent phenomena is, other things equal, more comprehensible than one with more.*³³

Perhaps if Einstein is right about the aim of science, then the rationale behind it is that scientists want a maximum of scientific understanding, and that we gain in scientific understanding by gaining in parsimony understood in terms of the number of brute phenomena on our overall world picture.³⁴

Friedman goes on to develop a theory of scientific explanation based on the idea that greater parsimony understood in terms of the number of brute phenomena on our overall world picture makes for greater scientific understanding. It is well known, however, that his theory is problematic in its details. Philip Kitcher shows this.³⁵ But the idea on which it is based—the idea that greater parsimony understood in terms of the number of brute phenomena on our overall world picture makes for greater scientific understanding—still has wide appeal.³⁶ So I want to grant it for the sake of argument.

There is still the second question above. Can the aim of science thus understood be reached without appeal to PAR₁?

It would be a mistake to think that the answer is negative. Here is a possibility. The inferred theory at the second layer is less parsimonious than its main rival (which was not inferred), and yet the transition from the first layer to the second constitutes an increase in parsimony because there are less brute phenomena on the theory at the second layer than on the theory at the first. Similarly, the inferred theory at the third layer is less parsimonious

³³ Friedman, “Explanation and Scientific Understanding,” *op. cit.*, pp. 14–15, emphasis added.

³⁴ Andrew Brenner, “Mereological Nihilism and Theoretical Unification,” *Analytic Philosophy*, LVI (2015): 318–37 discusses parsimony thus understood. He calls it “theoretical simplicity.”

³⁵ Philip Kitcher, “Explanation, Conjunction, and Unification,” *Journal of Philosophy*, LXXIII (1976): 207–12. William Roche and Elliott Sober, “Explanation = Unification? A New Criticism of Friedman's Theory and a Reply to an Old One,” *Philosophy of Science*, LXXXIV (2017): 391–413 argue that Friedman’s theory can be modified so that it is immune to Kitcher’s objection. They also argue, though, that Friedman’s theory is problematic nonetheless.

³⁶ See, for example, Stathis Psillos, *Causation and Explanation* (Montreal: Mc-Gill-Queen’s University Press, 2002), pp. 271–72.

than its main rival (which was not inferred), and yet the transition from the second layer to the third constitutes an increase in parsimony because there are less brute phenomena on the theory at the third layer than on the theory at the second. And so on. So, though PAR_1 is false, it might be that Einstein is right about the aim of science and that the aim of science as he describes it can be reached.

There is more. It also might be that it is permissible and perhaps even rational in some sense for scientists to develop and test theories of greater and greater parsimony understood in terms of the number of brute phenomena on our overall world picture. But they could do that without any appeal to PAR_1 .

III.3. IBE. I have assumed a broadly Bayesian epistemology on which (suppressing reference to background information) each of the following holds:

N_S : For any time t , S 's credences at t should be probabilistically coherent (that is, in accord with the probability calculus).

N_D : For any times t_1 and t_2 such that t_1 is before t_2 , if (i) S 's conditional credence at t_1 in T given E is c and (ii) S learns E (and nothing stronger) between t_1 and t_2 , then S 's unconditional credence at t_2 in T should be c .

N_S is a *synchronic* norm (thus the subscript in " N_S "). It implies, for example, that S should never have a higher credence in a conjunction than in its conjuncts taken individually. For, by the probability calculus, for any propositions P and Q , $\Pr(P \ \& \ Q) \leq \Pr(P)$ and $\Pr(P \ \& \ Q) \leq \Pr(Q)$. N_D , in contrast, which is oftentimes called "Strict Conditionalization," is a *diachronic* norm (thus the subscript in " N_D ").³⁷ Return to the card case from Section II.1. Before the card is drawn, your conditional credence in H given R is $1/2$. Suppose that the card is drawn, and that, due to the dealer's sloppiness, you get a glimpse of the card and learn that R (and nothing stronger). Then by N_D you should come to have an unconditional credence in H of $1/2$.³⁸

There is much to like in N_S , N_D , and Bayesian epistemology more generally. It might seem, though, that the proper reaction to my argument against PAR_1 and IBE is not to reject

³⁷ Strict Conditionalization stands in contrast to so-called "Jeffrey Conditionalization." See Richard Jeffrey, *The Logic of Decision* (New York: McGraw-Hill, 1965), chapter 11; Richard Jeffrey, *Subjective Probability: The Real Thing* (Cambridge: Cambridge University Press, 2004), chapter 3.

³⁸ For further discussion of Bayesian epistemology, see William Talbott, "Bayesian epistemology," in Edward Zalta, ed., *The Stanford Encyclopedia of Philosophy* (Winter 2016), URL = <<https://plato.stanford.edu/archives/win2016/entries/epistemology-bayesian/>>.

PAR₁ and IBE, but to reject N_S or N_D. Consider, for example, the following alternative to N_D:

*IBE***: For any times t_1 and t_2 such that t_1 is before t_2 , if (i) S 's conditional credence at t_1 in T_1 given E is c , (ii) S 's conditional credence at t_1 in T_2 given E is c^* , (iii) S learns that E (and nothing stronger) between t_1 and t_2 , and (iv) T_1 is a better potential explanation of E than is T_2 , then S 's unconditional credence at t_2 in T_1 should be $c + b$, and S 's unconditional credence at t_2 in T_2 should be $c^* - b^*$.³⁹

Here b is a bonus for being a better potential explanation of E , and b^* is a penalty for being a worse potential explanation of E . *IBE*** is underspecified in several respects. But it will suffice for my purposes. Take some case where PAR₁'s antecedent holds but its consequent does not because T_1 's posterior probability is equal to T_2 's posterior probability. If friends of IBE rejected N_D in favor of *IBE***, then they could claim that S 's unconditional credence in T_1 upon learning E should be *greater* than her unconditional credence in T_2 upon learning E .

I see little plausibility in *IBE***. There is simply too much to be said in favor of N_S, N_D, and Bayesian epistemology more generally, and too much to be said against *IBE***.⁴⁰ This is not the place, however, for a general defense of N_S, N_D, and Bayesian epistemology more generally. I shall instead suggest an alternative way forward for friends of IBE.

I claimed above in the introduction that IBE implies PAR₃. I should have claimed, to be more precise, that *if, as is standard, IBE is understood so that parsimony is an explanatory virtue, then IBE implies PAR₃*. It is important to note, however, that there is no necessity in understanding IBE so that parsimony is an explanatory virtue. It can be understood in terms of any proffered list of explanatory virtues. So the fact that PAR₃ is false tells *not* against IBE *per se*, but against IBE understood so that parsimony is an explanatory virtue.

³⁹ This is adapted from Bas van Fraassen, *Laws and Symmetry* (Oxford: Oxford University Press, 1989), chapter 7.

⁴⁰ For a recent critical discussion of *IBE*** (and related theses), see Nevin Climenhaga, "Inference to the Best Explanation made Incoherent," *Journal of Philosophy*, CXIV (2017): 251–73. I am not claiming that there is nothing to be said in favor of *IBE***. See Igor Douven, "Inference to the Best Explanation made Coherent," *Philosophy of Science*, LXVI (1999): S424–S435; Igor Douven, "Inference to the Best Explanation, Dutch Books, and Inaccuracy Minimisation," *Philosophical Quarterly*, LXIII (2013): 428–44; Igor Douven and Jonah Schupbach, "The Role of Explanatory Considerations in Updating," *Cognition*, CXLII (2015): 299–311; Igor Douven and Sylvia Wenmackers, "Inference to the Best Explanation versus Bayes's Rule in a Social Setting," *British Journal for the Philosophy of Science*, LXVIII (2017): 535–70. The point, to repeat, is just that there's little plausibility in *IBE***.

It would not be enough for friends of IBE to simply excise parsimony from their list of explanatory virtues. They would need to ensure that their list of explanatory virtues is such that there can be no cases where, given B , T_1 is a better potential explanation of E than is T_2 , and yet T_1 is equal to T_2 both in prior probability and in likelihood. For, otherwise IBE would still be open to counterexample. Is there a way for them to do that?

I have a proposal. It is not new, but it is underdeveloped at this point. The rough idea is to embrace N_S , N_D , and Bayesian epistemology more generally, and then find a place in Bayesianism for the explanatory virtues.⁴¹ One way of developing this idea is this. First, designate prior probability and likelihood as “primary” explanatory virtues. Second,

⁴¹ See Cabrera, “Can There be a Bayesian Explanationism? On the Prospects of a Productive Partnership,” *op. cit.*; Nevin Climenhaga, “How Explanation Guides Confirmation,” *Philosophy of Science*, LXXXIV (2017): 359–68; Climenhaga, “Inference to the Best Explanation made Incoherent,” *op. cit.*; Leah Henderson, “Bayesianism and Inference to the Best Explanation,” *British Journal for the Philosophy of Science*, LXV (2014): 687–715; Michael Huemer, “Explanationist Aid for the Theory of Inductive Logic,” *British Journal for the Philosophy of Science*, LX (2009): 345–75; Lipton, *Inference to the Best Explanation*, *op. cit.*, chapter 7; Kevin McCain and Ted Poston, “Why Explanatoriness is Evidentially Relevant,” *Thought*, III (2014): 145–53; Kevin McCain and Ted Poston, “The Evidential Impact of Explanatory Considerations,” in Kevin McCain and Ted Poston, eds., *Best Explanations: New Essays on Inference to the Best Explanation* (Oxford: Oxford University Press, 2018), pp. 121–29; Samir Okasha, “Van Fraassen’s Critique of Inference to the Best Explanation,” *Studies in the History and Philosophy of Science*, XXXI (2000): 691–710; Poston, *Reason and Explanation: A Defense of Explanatory Coherentism*, *op. cit.*; Stathis Psillos, “Inference to the Best Explanation and Bayesianism: Comments on Ilkka Niiniluoto’s ‘Truth-Seeking by Abduction’,” in Friedrich Stadler, ed., *Induction and deduction in the sciences* (Dordrecht: Kluwer, 2004), pp. 83–91; William Roche, “Explanation, Confirmation, and Hempel’s Paradox,” in Kevin McCain and Ted Poston, eds., *Best Explanations: New Essays on Inference to the Best Explanation* (Oxford: Oxford University Press, 2018), pp. 219–241; William Roche and Elliott Sober, “Explanatoriness is Evidentially Irrelevant, or Inference to the Best Explanation meets Bayesian Confirmation Theory,” *Analysis*, LXXIII (2013): 659–68; William Roche and Elliott Sober, “Explanatoriness and Evidence: A Reply to McCain and Poston,” *Thought*, III (2014): 193–99; William Roche and Elliott Sober, “Is Explanatoriness a Guide to Confirmation? A Reply to Climenhaga,” *Journal for General Philosophy of Science*, XLVIII (2017): 581–90; Robert Smithson, “The Principle of Indifference and Inductive Scepticism,” *British Journal for the Philosophy of Science*, LXVIII (2017): 253–72; Jonathan Weisberg, “Locating IBE in the Bayesian Framework,” *Synthese*, CLXVII (2009): 125–43.

designate any additional explanatory virtues as “secondary” explanatory virtues. Third, show that the various secondary explanatory virtues can help in determining prior probability and likelihood (the primary explanatory virtues). If IBE were thus reformulated, then no case where T_1 is equal to T_2 both in prior probability and in likelihood would be a case where, given B , T_1 is a better potential explanation of E than is T_2 .

IV. CONCLUSION

If parsimony (understood in terms of the number of things on/in a theory) is a theoretical virtue, then this is *not* because PAR_1 is true. For, regardless of whether PAR_1 is to be understood so that prior probability and likelihood are included in r_1, \dots, r_n , PAR_1 is false. It follows that the same is true of PAR_2 , PAR_3 , and any other parsimony thesis that is logically stronger than PAR_1 . It further follows that the same is true of IBE understood so that parsimony is an explanatory virtue. This is the negative side of the story. However, there is also a positive side. First, there are still ways to show that the more parsimonious theory in a given case has the higher posterior probability. Second, it might be that Einstein is right that central to the aim of science is parsimony understood in terms of the number of brute phenomena on our overall world picture, and it might be that the aim of science as he describes it can be reached. Third, it might be that there are ways of understanding IBE on which it is not open to counterexample.

WILLIAM ROCHE

Texas Christian University