On the Logical Structure of Best Explanations[*]

Jonah N. Schupbach

Abstract

Standard articulations of Inference to the Best Explanation (IBE) imply the uniqueness claim that exactly one explanation should be inferred in response to an explanandum. This claim has been challenged as being both too strong (sometimes agnosticism between candidate explanatory hypotheses seems the rational conclusion) and too weak (in cases where multiple hypotheses might sensibly be conjointly inferred). I propose a novel interpretation of IBE that retains the uniqueness claim while also allowing for agnostic and conjunctive conclusions. I then argue that a particular probabilistic explication of explanatory goodness helpfully guides us in navigating such options when using IBE.

1 Uniqueness and IBE's Critics

Inference to the Best Explanation (IBE) is a form of uncertain inference that favors *the* single best potential explanation of some given explanandum. As such, standard articulations of IBE imply the following uniqueness claim:

Uniqueness. In any case that a reasoner is in possession of a set of potential explanations of some given explanandum, IBE advises that agent to infer *at least* and *at most* one of these explanations.

[*] I owe special thanks to David H. Glass, Jonathan Livengood, and Lauren Ross for helpful conversations and feedback regarding this project. I am also very grateful for feedback I received when presenting this material at the Düsseldorf Center for Logic and Philosophy of Science and as part of the University of Minnesota's Studies in Science and Technology colloquium series. Work on this paper was supported by the *Charles H. Monson Mid-Career Award* administered by the University of Utah's Philosophy Department and through the support of a grant (#61115) from the John Templeton Foundation. The opinions expressed in this publication are those of the author and do not necessarily reflect the views of the John Templeton Foundation.

Uniqueness secures the intuitive usefulness of IBE, which seems to have a near ubiquitous presence in everyday and scientific reasoning (Harman, 1965; Lipton, 2004; Keil, 2006; Lombrozo, 2006; Douven and Schupbach, 2015; Douven, 2016; Schupbach, 2017). However, this claim ostensibly makes IBE vulnerable to multiple lines of attack, **Uniqueness** featuring in the most well-known criticisms of IBE.

For one thing, **Uniqueness** opens IBE up to the criticism that it is unable in some cases to avoid legislating inferences to manifestly poor explanations. **Uniqueness** requires reasoners to infer at least one explanation, problematically including cases in which the best explanation is not good enough. This concern underlies van Fraassen's (1989, pp. 143-43) "bad lot" objection to IBE. Were it not for **Uniqueness**'s requirement that one best explanation always be inferred, there would be no obvious concern with cases in which we have only a bad lot of hypotheses from which to choose. More recently, McCain and Poston (2019, p. 1) present a closely related "disjunction objection" to IBE (see also Fumerton 1995, p. 209):

> [A] particular hypothesis [may be] the best explanation of a given set of evidence even though the disjunction of its rivals is more likely to be true. [... But] it is not rational to believe that H is true when H is more likely to be false than true.

In cases like this, where the most explanatory hypothesis is in some sense not good enough, it is unreasonable to go ahead nonetheless and infer that hypothesis. The rationally appropriate response, rather, would seem to be agnosticism between at least some of the explanatory options.

Scientific instances of such situations and the appropriate agnostic response are easy to find, for example, in cases of causal heterogeneity and multiple realizability (Ross, 2020, 2022). To focus on one example in particular, here and throughout this paper, Parkinson's disease (PD) is produced by a complex of causal pathways across different patients. Nandipati and Litvan (2016) review a variety of studies linking cases of PD to environmental exposures to pathologic agents found in pesticides and industrial compounds. Lesage and Brice (2009) survey a host of genetic studies, highlighting "more than 13 loci and 9 genes that have been identified" as having some role in PD's etiology. Such studies reveal monogenic forms of PD as well as forms attributable to various gene-environment interactions. As Ross (2022, p. 10) summarizes, PD "can be produced by single gene variants, single environmental factors, and combinations of genetic and environmental factors." Without knowing a patient's genetics or environmental background, it would be heedless for a clinician to attribute a case of PD to any one specific causal pathway; a disjunctive conclusion between possible pathways would be more prudent. Apparently, **Uniqueness** compels the clinician in this case incautiously to infer too much.

But **Uniqueness** also makes IBE vulnerable to the criticism that, in other cases, it forces reasoners to infer too little. Specifically, it guides reasoners to infer at most one explanation, even in cases where a multiplicity of compatible explanations are conjointly warranted (Salmon, 2001, p. 67). The reasonable inference here would rather be to the conjunction of explanatory hypotheses.

Again, examples of such situations and the appropriate conjunctive response abound, for example, in cases of ”multicausality” (Ross, 2022). Causal pathways to PD may cite particular combinations of gene variants and/or environmental conditions working together. Lesage and Brice (2009, p. R52) note that the large majority ($> 90\%$) of cases of PD are not monogenic. Rather, most cases seem to “result from complex interactions among genes and between genes and environmental factors.” If the evidence for a particular patient is such that a multicausal pathway involving exposure to pesticides combined with a genetic variant as a susceptibility factor is most explanatory, then the appropriate inference would be to the conjunction of these factors. To the extent that IBE forces the clinician instead to pick at most one factor to the exclusion of others, it recommends an unreasonably stringent conclusion.

In response to these challenges, IBE’s defenders have largely given up on **Uniqueness**. The bad lot objection and related concerns have led philosophers to recast IBE as only guiding explanatory inference in cases where the best explanation is also “sufficiently good” (Musgrave 1988, pp. 238-39; Lipton 2004, p. 93; McCain and Poston 2019, p. 5). Concerns like Salmon’s have similarly led philosophers to limit IBE’s instances to cases in which the available explanatory hypotheses compete in some sense (Lipton, 2001a, p. 104). If the hypotheses do not compete with one another, then IBE does not apply and thus does not force us to choose between them. As I argue elsewhere (Schupbach, 2019), such hedges impose absurdly strong limitations on IBE’s intuitively wide applicability.

The remainder of this paper argues that philosophers have been too quick to weaken IBE. I defend a traditional account of IBE, **Uniqueness** and all, against the above challenges. At the heart of my response is the claim that these criticisms rely upon a questionable interpretation of “best explanation”. An alternative reading of this phrase allows IBE’s proponents both to endorse **Uniqueness** and sidestep the criticisms.

2 The "Best Explanation"

Virtually all commentators on IBE—defenders and critics alike—presume a common interpretation of "best explanation". This reading associates any best explanation with *the most explanatory individual hypothesis*. More generally, potential explanations in any instance of IBE correspond one-to-one with individual hypotheses. The best such explanation then naturally amounts to the *individual hypothesis* which best explains the relevant explanandum. Thus, in his pioneering article on the subject, Harman (1965, p. 89) describes IBE as follows: "In making this inference one infers, from the premise that a given hypothesis would provide a 'better' explanation for the evidence than would any other hypothesis, to the conclusion that the given hypothesis is true." Lipton (2001b, pp. 56-57) similarly presumes this interpretation of "best explanation" when defending IBE. More recently, Lange (2022, p. 85) describes IBE as an inference form in which "we argue that one hypothesis derives some plausibility over its rivals from the fact that the explanations it would give (if it were true) are better than those its rivals would give (if they were true)." This sample is small but entirely characteristic of the literature.

While this interpretation is all but universally assumed, it is never called out for defense. However, the interpretation is indeed questionable. Arguably, it is also the actual source of trouble when it comes to the above challenges to **Uniqueness**. The standard interpretation makes any instance of IBE problematically sensitive to a given individuation of hypotheses, an odd consequence with no apparent parallel in real-world examples of explanatory inference. The inferences we may draw become artificially limited by the contingent way in which we have carved up the space of hypotheses. Any such individuation becomes an inferential barrier, blocking our path to conclusions that may describe better explanations.

Returning to the example of PD, any particular lot of available explanatory hypotheses can either be too finely- or coarsely-grained for a proper diagnostic explanation. Coarsely-individuated hypotheses (e.g., pesticide exposure, industrial compound exposure, genetic variations, etc.) will often not be sufficiently informative to have any substantial explanatory value. If the evidence is best explained by a combination of genetic and environmental causes, then the clinician would naturally want to start conjoining some of these coarsely-grained alternatives to construct more informative explanations. Any finely-grained hypothesis (e.g., long-term, repeated exposure to organochlorines combined with short-term exposure to rotenone and the G2385R mutation in the LRRK2 gene, etc.) may end up being too specific. Given less detailed evidence, the most explanatory diagnosis may be merely to posit that there was, for example, some past exposure to pesticides combined with a genetic susceptibility factor. A less committed explanation such as this is the logical disjunction of more finely-grained alternatives.

These observations provide us with direction on how we might attempt to defend IBE without having to give up **Uniqueness**. Namely, we might reconsider our interpretation of "best explanation". There is at least one other natural interpretation available. On this alternative, individual hypotheses may offer best explanations in some cases, but this need not be the case. More generally, "best explanation" can refer to whatever *epistemic stance* or *conclusion* is most explanatory (Schupbach, 2019, pp. 158-59).

On such an interpretation, explanatory inference is not strangely beholden to the way in which the space of hypotheses happens to be carved up. But relative to such a space, explanatory stances available for inference include, in principle, any and all Boolean combinations of the hypotheses. Individual hypotheses may be combined in any way that results in better explanations. If multiple hypotheses conjointly offer a more explanatory stance than any of these individually—as in cases of multicausality—then IBE will guide us to infer such conjunctive explanations when "best explanation" is understood in the proposed sense. If an agnostic shrug between individual hypotheses is all the explanation one can properly infer—as in some cases of causal heterogeneity or multiple realizability—then that is exactly the conclusion that IBE, so interpreted, will recommend.

3 Determining the Structure of Best Explanations

This reinterpretation provides IBE with a potential end-around well-known criticisms; however, it also gives rise to a pressing, new demand. According to this reading, individual hypotheses may be conjoined, disjoined, or otherwise logically combined to formulate best explanations. Accordingly, reasoners must not only compare the explanatory goodness of the various individual hypotheses on the table, but also must assess whether any explanatory improvements come by way of logically strengthening or weakening such hypotheses. Both possibilities may seem problematic. Strictly stronger explanations are inevitably less probable than correspondingly weaker alternatives. On the upside, they are also more informative. By contrast, strictly weaker explanatory stances are less informative but inevitably more probable than correspondingly stronger stances. The new challenge facing supporters of this account of IBE is to offer some clarity on how to negotiate these opposing goals.

This is a familiar demand, effectively being an explanatory version of

Shun error! ←→ Believe truth!

h_1	h_2	$h_1 \vee \neg h_1$	$h_1 \vee h_2$	h_1	$h_1 \& h_2$	$h_1 \& \neg h_1$
T	T	T	T	T	T	F
T	F	T	T	T	F	F
F	T	T	T	F	F	F
F	F	T	F	F	F	F

Figure 1: Some explanatory stances one might infer apropos h_1 and h_2 in order of increasing logical strength.

the challenge to find a balance between James's (1896, pp. 338-39) "dual goals of cognition:" the *pursuit of truth* and *avoidance of error*.[1] The "chase for truth" is achieved perfectly by an agent who accounts for the explanandum by inferring the grand conjunction of all claims. The duty to "shun error" is instead accomplished perfectly by an agent who "explains" the evidence by inferring the grand disjunction of all claims. The problem with the perfect truth-chaser, however, is that this agent is also the perfect error-chaser. The problem with the perfect error-avoider, by contrast, is that this agent is also the perfect information-avoider, resolute on accepting a maximally uninformative stance.

Is there a principled way to strike a balance between these considerations in explanatory inference? James asserts that attitudes relating these duties "are in any case only expressions of our passional life" (p. 339). However, in the context of IBE, a less subjective balance seems more promising. It is worth considering whether there is a determinable tipping point at which a potential explanation is exactly as logically strong and informative with respect to the explanandum as it should be—any stronger and its inevitable loss of likeliness is not worth any remaining potential gains in informativeness.

The truth-table shown in Figure 1 lists some of the explanatory stances available for inference in the case that only two hypotheses, h_1 and h_2, are given. We seek an account of the notion of *explanatory goodness* that guides us when choosing between such possibilities. Such an account should clarify for us when it is explanatorily better, for example, to infer the more informative $h_1 \& h_2$ instead of h_1, or alternatively when to infer the more probable $h_1 \vee h_2$ over the more informative h_1.

In recent work, Glass and Schupbach (2023a,b) develop and defend the

[1] Philosophers of science might also recognize in this project the challenge to find the proper balance between a scientific theory's evidential fit and its informational simplicity (see Good 1968; Sober 2015, p. 12; and Schupbach 2022, sect. 3.4).

following measure of the degree of explanatory goodness that an explanatory hypothesis h has with respect to an explanandum e:[2]

$$\mathcal{E}(e,h) = \log\left(\frac{Pr(e|h)Pr(h)^{1/2}}{Pr(e)}\right).$$

Rather than rehearse the full case for this measure, the remainder of this section briefly compares it to alternatives, highlighting $\mathcal{E}$'s particular suitability as an explication of our target notion of explanatory goodness.

Formal epistemologists offer accounts of explanatory "power" that provide *prima facie* plausible alternative explications of our target concept (see footnote 2). These take the form of "relevance measures" in the sense of (Fitelson, 1999, p. S363). Relevance measures gauge the degree of statistical relevance between any h and e; any such measure $r(e,h)$ consequently implies the following:

$$r(e,h_1) > r(e,h_2) \text{ iff } Pr(e|h_1) > Pr(e|h_2).$$

This simple implication of relevance measures provides a strong argument against their application for our purposes. In effect, if any relevance measure is used to explicate the notion of explanatory goodness at work in IBE (as we've interpreted it), then IBE will virtually always guide us to infer logically stronger explanations. Let h_1 provide an appealing potential explanation with substantial positive relevance to e: $Pr(e|h_1) \gg Pr(e)$. Now consider any additional h_2 at all; so long as it isn't contrary to h_1, it can be irrelevant to or even as negatively associated with h_1 as you like. If e is even slightly more likely given $h_1\&h_2$ than given h_1 alone, $Pr(e|h_1\&h_2) > Pr(e|h_1)$, this account tells us to favor the logically stronger (and possibly exceedingly improbable) explanation. In general, whenever the likelihood of e can be bumped up by strictly strengthening one's explanatory stance, this logically stronger position will win out in terms of the above inequality. This makes logically stronger explanations far too easy to come by. In terms of James's dual goals, we end up approaching the position of the perfect truth-chaser, accepting far too many claims into our explanatory conclusion and thereby all but guaranteeing that this overly-specified conclusion is false.

There is a price we pay when we favor a logically stronger, informationally more complex conclusion; there are strictly more ways such a stance could be wrong. The problem with accepting any relevance measure as our explication of explanatory goodness is that it essentially ignores this

[2] This work considers several measures of explanatory power (e.g., those defended by Popper 1959; McGrew 2003; Schupbach and Sprenger 2011; and Crupi and Tentori 2012), arguing that $\mathcal{E}$ provides a more appropriate explication for our purposes than these.

price. Any benefit in accounting for e (no matter how slight) is worth any price we pay by complicating our explanatory stance (no matter how great). This account thus fails for our purposes. Evidently, what is needed is an account of explanatory goodness that tempers considerations of explanatory relevance with penalties for informational complexity.

Plausibly, Bayes's Theorem does exactly this. The ratio $r_{GM}(e,h) = Pr(e|h)/Pr(e)$ is a relevance measure proposed by Good (1968) and McGrew (2003) for gauging h's ability to account explanatorily for e; as such, r_{GM} fails to penalize for complexity. Because increasing informational complexity (increasing logical strength) corresponds to a decreasing probability, an explanatory conclusion's prior probability provides a straightforward penalization factor for complexity. For example, for logically independent h_1 and h_2, $h_1 \& h_2$ is strictly more informationally complex than either individual hypothesis, and so it should be penalized relative to these weaker options. This can be achieved by weighting a hypothesis's explanatory goodness by it's probability—since $Pr(h_1 \& h_2) \leq Pr(h_1[h_2])$. Bayes's Theorem does precisely this:

$$Pr(h|e) = \frac{Pr(e|h)}{Pr(e)} \times Pr(h) = r_{GM}(e,h) \times Pr(h).$$

However, using $Pr(h|e)$ to balance explanatory relevance against informational complexity essentially leads to the opposite problem as that facing relevance measures. If posterior probabilities are used to explicate explanatory goodness, then IBE always guides us to infer logically *weaker* stances. Let h_1 provide an appealing potential explanation with some substantial positive relevance to e: $Pr(e|h_1) \gg Pr(e)$. Now consider any additional h_2 at all. So long as h_2 is not logically dependent on h_1 or e, then we get the result $Pr(h_1 \vee h_2|e) > Pr(h_1|e)$, and thus this account tells us to favor the logically weaker (and possibly maximally uninformative) explanation. Logically weaker positions inevitably win out in terms of the above inequality and so are favored by IBE if we use posterior probabilities to gauge explanatory goodness. This makes logically weaker explanations sure winners, precluding us from ever inferring informative explanations. In terms of the Jamesian tradeoff, we end up approaching the position of the perfect error-avoider, ever shrugging at explananda and inferring informationally-vacuous "explanations" in all cases.

Relevance measures ignore informational complexity altogether, while posterior probabilities place extreme weight on complexity such that logically stronger explanations are banned from the start. The notion of explanatory goodness at work in IBE when we adjudicate between explanatory options at different levels of logical strength must accordingly strike a balance between these options. Measure $\mathcal{E}$ plausibly provides exactly such a balance:

$$\mathcal{E}(e,h) = \log\left(\frac{Pr(e|h)Pr(h)^{1/2}}{Pr(e)}\right).$$

Indeed, Good (1968, p. 130) originally developed and defended this measure because it "gives equal weights" to explanatory relevance and informational simplicity (the "avoidance of clutter").

Note that $\mathcal{E}$ would amount simply to the logarithm of the posterior probability, but for the fact that the factor penalizing for complexity, $Pr(h)$, is given less weight—being raised to the power 1/2 instead of 1. This is appropriate since, for IBE purposes, posterior probability enforces a problematically extreme such penalty. We can also think of $\mathcal{E}$ as being equivalent to the relevance measure r_{GM} but for the fact that $Pr(h)$ is given non-zero weight (if the exponent were 0, of course, $\mathcal{E}$ would amount to the log-normalized version of r_{GM}). This too is appropriate, since r_{GM} and all other relevance measures fail for our purposes because they do not penalize for complexity.[3]

4 IBE's Critics Revisited

With our interpretation of "best explanation" and $\mathcal{E}$ in hand, this section reconsiders the objections to IBE summarized earlier. Salmon's objection is that IBE's **Uniqueness** claim prohibits reasoners from inferring multiple explanatory hypotheses in cases of explanatory multiplicity or multicausality. In response, philosophers give up **Uniqueness**, greatly limiting IBE's domain of applicability to cases in which individual hypotheses compete. Our novel interpretation of "best explanation" instead allows us to respond by showing that IBE in its traditional form (**Uniqueness** and all) *can* recommend inference to conjunctions of hypotheses in the salient cases. Once we interpret IBE as allowing inferences to stances that may have the logical structure of any Boolean combination of the individual hypotheses on offer, $\mathcal{E}$ reveals that it is indeed possible for logically stronger, less probable stances to be explanatorily superior to weaker alternatives.[4]

Consider a case in which h_1 and h_2 each offer potential explanations of e. The stronger, conjunctive stance $h_1\&h_2$ is explanatorily superior to both individual hypotheses under the following condition (Glass and Schup-

[3] Of course, there are any number of other middling weightings one might try aside from setting this exponent to 1/2. Justifying this particular value is part of the full case for $\mathcal{E}$ that has been taken up elsewhere (Glass, 2022; Schupbach, 2022).

[4] Glass and Schupbach (2023a,b) provide more in-depth explorations into the logic and formal epistemology of such "conjunctive explanations."

bach, 2023a,b):[5]

$$\log\left[\frac{Pr(e|h_1\&h_2)}{Pr(e|h_1)}\right] > \log\left[\frac{1}{Pr(h_2|e\&h_1)}\right] \tag{1}$$

The left hand side of (1) is $r_{GM}(e, h_2|h_1)$; in words, this is the explanatory relevance that h_2 has with respect to e in light of already accepting h_1. The right hand side explicates the degree to which h_2 is penalized for adding more information to our explanation in light of already accepting h_1 and e. Condition (1) thus clarifies that stronger explanations are to be preferred whenever the explanatory relevance to e that would be added by inferring h_2 in addition to h_1 outweighs the price in informational complexity incurred by this move. The point at which this just ceases to be true is the tipping point at which the conclusion is exactly as informative as it should be, since any further logical strengthening of the inferred explanation would not be worth the price in increased complexity (decreased probability).

This is the situation, for example, in diagnoses of PD for which the evidence is sufficiently rich to be satisfactorily accounted for only with a multicausal explanation. Such a conclusion would inevitably be less probable than a less detailed, more agnostic explanation. But that lower probability is worth the gains in explanatory relevance and informativeness in accounting for the evidence.

By contrast to Salmon's objection, the bad lot and disjunction objections both argue that **Uniqueness** forces reasoners to infer too much. In cases where none of our individual hypotheses are explanatorily good (or good enough), the complaint is that **Uniqueness** still rashly compels us to infer one of these. IBE's defenders offer the ad hoc response that IBE only applies when there is no such concern—i.e., when the best *is* good enough, when the lot *is not* entirely bad, etc. An alternative response made possible by our reinterpretation of "best explanation" instead shows that IBE

[5] Proof: Assume without loss of generality that $\mathcal{E}(e, h_1) > \mathcal{E}(e, h_2)$. Then $\mathcal{E}(e, h_1\&h_2) > \mathcal{E}(e, h_1)$

$$\Leftrightarrow Pr(e|h_1)Pr(h_1)^{1/2} < Pr(e|h_1\&h_2)Pr(h_1\&h_2)^{1/2}$$

$$\Leftrightarrow \left(\frac{Pr(h_1)}{Pr(h_1\&h_2)}\right)^{1/2} < \frac{Pr(e|h_1\&h_2)}{Pr(e|h_1)}$$

$$\Leftrightarrow \frac{Pr(h_1)}{Pr(h_1\&h_2)} < \frac{Pr(e|h_1\&h_2)Pr(e|h_1\&h_2)}{Pr(e|h_1)Pr(e|h_1)}$$

$$\Leftrightarrow \frac{Pr(h_1)Pr(e|h_1)}{Pr(h_1\&h_2)Pr(e|h_1\&h_2)} < \frac{Pr(e|h_1\&h_2)}{Pr(e|h_1)}$$

$$\Leftrightarrow \frac{Pr(e\&h_1)}{Pr(e\&h_1\&h_2)} < \frac{Pr(e|h_1\&h_2)}{Pr(e|h_1)}$$

$$\Leftrightarrow \log\left[\frac{1}{Pr(h_2|e\&h_1}\right] < \log\left[\frac{Pr(e|h_1\&h_2)}{Pr(e|h_1)}\right] \quad \square$$

in its traditional form (**Uniqueness** and all) does not require us to infer explanatorily poor individual hypotheses.

Consider the main case put forward by McCain and Poston (2019, p. 2) in presenting their objection:

> Let h_1 be the hypothesis that a fair coin has been chosen, i.e., $Pr(heads|h_1) = 1/2$; h_2 is the hypothesis that a coin with a strong bias for heads is chosen, e.g., $Pr(heads|h_2) = 3/4$; and h_3 is the hypothesis that a coin with a strong bias against heads is chosen, e.g., $Pr(heads|h_3) = 1/4$. There are only three coins to be chosen, and each has the same probability of being chosen—$Pr(h_1) = Pr(h_2) = Pr(h_3) = 1/3$. The results of the flip of each coin are independent of previous flips. A coin is selected at random and flipped four times. The results are e: $\langle H, T, T, H\rangle$.

The most explanatory of the individual hypotheses *apropos e* seems clearly to be h_1; however, $Pr(h_1|e) \approx .47 < .5$. That is, considering inferences only to individual hypotheses, h_1 provides the intuitively best explanation of e while nonetheless being more likely false than true in light of e. It is this fact that McCain and Poston insist makes h_1 not explanatorily good enough for inference. They go on to endorse an ad hoc response, simply requiring that IBE only applies in cases where the best explanation is more probable than not.

What becomes of this same case if we adopt our interpretation of "best explanation" along with $\mathcal{E}$ as our explication of explanatory goodness? Doing so multiplies the explanatory conclusions available for inference, our lot of candidate explanations now including the individual hypotheses h_1, h_2, and h_3 *along with their various Boolean combinations*. Importantly, these new alternatives include agnostic stances like $h_2 \vee h_3$ and $h_1 \vee h_2 \vee h_3$. Taking into account these possible stances leads to a different inference than in McCain and Poston's discussion. First, $\mathcal{E}$ assesses the most explanatory hypothesis h_1 as having negative explanatory value:

$$\begin{aligned}\mathcal{E}(e, h_1) &= \log\left(\frac{Pr(e|h_1)Pr(h_1)^{1/2}}{Pr(e)}\right)\\ &= \log\left(\frac{0.5^4 \cdot 1/3^{1/2}}{1/3 \cdot 0.25^2 \cdot 0.75^2 + 1/3 \cdot 0.5^4 + 1/3 \cdot 0.75^2 \cdot 0.25^2}\right) = -.089\end{aligned}$$

Nonetheless, h_1 unsurprisingly performs far better than the other hypotheses: $\mathcal{E}(e, h_2) = \mathcal{E}(e, h_3) = -.339$. Interestingly, h_1 scores better even than the stance that remains agnostic between the other two individual hypotheses: $\mathcal{E}(e, h_2 \vee h_3) = -.188$. However, there remains one (and only one) alternative stance that has *non-negative* explanatory value: $\mathcal{E}(e, h_1 \vee$

$h_2 \vee h_3) = 0$. The "best explanation" then according to $\mathcal{E}$ is the maximally agnostic $h_1 \vee h_2 \vee h_3$, acknowledging that the evidence is not sufficiently rich to warrant any informative explanatory conclusion whatever.

Far from being more likely false than true, this stance has unit probability, $Pr(h_1 \vee h_2 \vee h_3|e) = 1$. However, $\mathcal{E}$ also rightly reveals that this maximally uninformative conclusion is the best of bad explanatory options, having no positive explanatory value. In fact, in all cases, a fully agnostic stance between alternatives *necessarily* has zero explanatory value over any explanandum according to $\mathcal{E}$.[6] This is completely appropriate, lending formal backing to the common observation that tautologies cannot explain anything (neither an explanandum nor its negation).

Since $Pr(h_1 \vee h_2 \vee h_3|e) > .5$, this case no longer serves McCain and Poston's purposes as a counterexample. But the fact that this stance can at once be so probable and explanatorily impotent starts to suggest a deeper issue with their objection. Specifically, this result reveals an important disconnect between a stance's explanatory value and its probability; $h_1 \vee h_2 \vee h_3$ couldn't score better in terms of probability, but it couldn't be more worthless in terms of explanatory value. This disconnect is also suggested in our response to Salmon's objection, in which we observed that probability can rightly be sacrificed in explanatory inference for the sake of greater explanatory relevance. The upshot is that there is no *simple* connection between explanatory value (the determining factor in explanatory inference) and probability. If this is right, then contrary to the account of IBE that McCain and Poston end up defending, we cannot simply identify "sufficient explanatory goodness" with probability exceeding .5—or any other probabilistic threshold.

Through the lens of our alternative interpretation of IBE, McCain and Poston's coin case—far from constituting a counterexample—helpfully clarifies a rational response to cases involving "bad lots" of explanatory hypotheses. When none of the individual hypotheses in our lot of potential explanations have any positive explanatory value with respect to the explanandum, agnostic stances between such hypotheses become explanatorily appealing. In the worst case, when we are truly working with a purely bad lot of such hypotheses, we can still do no worse than acknowledge that we are in the weakest of explanatory positions by concluding the disjunction of all of these. $\mathcal{E}$ provides the guide here, either directing us to take the fully agnostic, explanatorily vacuous stance or to more informative options that provide us with positive explanatory value when such exist.

[6] Let T be the probabilistically tautologous disjunction of all hypotheses in the possibility space such that $Pr(\phi_j) = 1$. Then $\mathcal{E}(e, \phi_j) = \log\left(\frac{Pr(e|\phi_j)Pr(\phi_j)^{1/2}}{Pr(e)}\right) = \log\left(\frac{Pr(e)}{Pr(e)}\right) = 0$.

Real-life examples like the diagnosis of a particular patient with PD will sometimes fit this description. Relative to the evidence of such a case e, a disjunction of causal hypotheses (e.g., $h_1 \vee h_2$) is explanatorily better than the individual etiologies on offer (h_1 and h_2) under the following condition:[7]

$$\log\left[\frac{1}{Pr(h_1|(h_1 \vee h_2)\&e)}\right] > \log\left[\frac{Pr(e|h_1\&(h_1 \vee h_2))}{Pr(e|h_1 \vee h_2)}\right] \tag{2}$$

The left hand side of (2) explicates the degree to which a commitment specifically to h_1 would be penalized for adding more information to our explanation in light of already accepting the disjunctive stance $h_1 \vee h_2$ along with e. The right hand side is $r_{GM}(e, h_1|h_1 \vee h_2)$—i.e., the explanatory relevance that committing to h_1 would gain us with respect to e compared to merely inferring $h_1 \vee h_2$. This condition thus makes precise the sensible idea that we should opt for more agnostic, weaker explanations so long as the gains in explanatory relevance we could acquire by strengthening our conclusions would not be worth the cost incurred by inferring such an inevitably less probable, more specific conclusion.

Other interesting results may be observed by extending McCain and Poston's example. Let's say they flip their coin four more times, with the result being the new evidence set e': $\langle H, T, T, H, H, T, H, T\rangle$. In this case (as in *any* case), the maximally agnostic $h_1 \vee h_2 \vee h_3$ continues to have zero explanatory value. However, some of the alternative stances now have positive explanatory goodness. Most notably, the best explanation in this case is provided by h_1 with $\mathcal{E}(e', h_1) = 0.025$. By contrast, h_2, h_3, and the agnostic stance $h_2 \vee h_3$ all appropriately score worse relative to e', with $\mathcal{E}(e, h_2) = \mathcal{E}(e, h_3) = -.474$ and $\mathcal{E}(e', h_2 \vee h_3) = -.323$. Importantly, this extended case also provides no counterexample à la McCain and Poston, as h_1 is more likely true than false relative to e': $Pr(h_1|e') = .612$.

The natural follow-up question to this last point is whether our approach provides us with a general way around McCain and Poston's alleged counterexamples. Is the "best explanation" in our sense (i.e., the stance with maximal value of $\mathcal{E}$) always more probably true than false? In

[7] Proof: Assume that $\mathcal{E}(e, h_1) > \mathcal{E}(e, h_2)$. Then $\mathcal{E}(e, h_1 \vee h_2) > \mathcal{E}(e, h_1)$

$$\Leftrightarrow Pr(e|h_1 \vee h_2)Pr(h_1 \vee h_2)^{1/2} > Pr(e|h_1)Pr(h_1)^{1/2}$$

$$\Leftrightarrow \frac{Pr(h_1 \vee h_2)Pr(e|h_1 \vee h_2)}{Pr(e|h_1)Pr(h_1)} > \frac{Pr(e|h_1)}{Pr(e|h_1 \vee h_2)}$$

$$\Leftrightarrow \frac{Pr((h_1 \vee h_2)\&e)}{Pr(h_1\&e)} > \frac{Pr(e|h_1)}{Pr(e|h_1 \vee h_2)}$$

$$\Leftrightarrow \frac{Pr((h_1 \vee h_2)\&e)}{Pr(h_1\&(h_1 \vee h_2)\&e)} > \frac{Pr(e|h_1\&(h_1 \vee h_2))}{Pr(e|h_1 \vee h_2)}$$

$$\Leftrightarrow \log\left[\frac{1}{Pr(h_1|(h_1 \vee h_2)\&e)}\right] > \log\left[\frac{Pr(e|h_1\&(h_1 \vee h_2))}{Pr(e|h_1 \vee h_2)}\right] \quad \square$$

fact, no; there exist cases for which the stance with maximal $\mathcal{E}$ nonetheless has probability $< .5$ conditional on the explanandum.[8]

However, this fact is both unsurprising and unconcerning in light of the work set out in this paper. We have already suggested why above, when diagnosing the "deeper issue" with McCain and Poston's objection. Recall the disconnect between a stance's explanatory value and its probability; there is no straightforward relation between these such that, for example, the most explanatory stance must be more probable than not. To endorse such a connection is ultimately to overvalue "shunning error" at the expense of "believing truth" in the Jamesian balance. A blanket requirement that best explanations must rise above a certain probabilistic threshold can be motivated by attending *exclusively* to the goal of "shunning error". Reasoners who are also appropriately concerned with "believing truth" are willing to sacrifice probabilistic assurance that their stance is not false for the sake of endorsing more informative explanatory positions. Instead of attempting to force a simple link between explanatory value and probability, the present account respects both goals of cognition, arguing that $\mathcal{E}$ provides us with a principled means of negotiating the Jamesian tradeoff in contexts of explanatory inference.

References

Crupi, V. and Tentori, K. (2012). A second look at the logic of explanatory power (with two novel representation theorems). *Philosophy of Science*, 79(3):365–385.

Douven, I. (2016). Abduction. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, winter 2016 edition.

Douven, I. and Schupbach, J. N. (2015). The role of explanatory considerations in updating. *Cognition*, 142:299–311.

[8] Such examples exist already in the simplest cases, where the available explanatory stances relate only to a single individual hypothesis: h, $\neg h$, $h \vee \neg h$, and $h \& \neg h$. For example, if $Pr(h) = .2$, $Pr(e) = .4$, and $Pr(e|h) = .995$ (in which case $Pr(e|\neg h) = .25125$), then the following two items are simultaneously true:

(a) $\mathcal{E}(e,h)$ is maximal: $\mathcal{E}(e,h) = .046 > -.25 = \mathcal{E}(e,\neg h)$ and $\mathcal{E}(e,h) = .046 > 0 = \mathcal{E}(e, h \vee \neg h)$.

(b) $Pr(h|e) = .4975 < .5$.

Joint satisfiability of these general conditions was established and this particular model discovered using Fitelson's (2008) decision procedure PrSAT as implemented in his corresponding *Mathematica* package, available at <http://fitelson.org/PrSAT/>.

Fitelson, B. (1999). The plurality of Bayesian measures of confirmation and the problem of measure sensitivity. *Philosophy of Science*, 66:S362–S378.

Fitelson, B. (2008). A decision procedure for probability calculus with applications. *The Review of Symbolic Logic*, 1(1):111–125.

Fumerton, R. (1995). *Metaepistemology and Skepticism*. Rowman & Littlefield, Lanham, Maryland.

Glass, D. H. (Forthcoming, 2022). Information and explanatory goodness. *Erkenntnis*.

Glass, D. H. and Schupbach, J. N. (2023a). Conjunctive explanation. Unpublished.

Glass, D. H. and Schupbach, J. N. (Forthcoming, 2023b). Conjunctive explanation: Is the explanatory gain worth the cost? In Schupbach, J. N. and Glass, D. H., editors, *Conjunctive Explanations: The Nature, Epistemology, and Psychology of Explanatory Multiplicity*. Routledge, New York.

Good, I. J. (1968). Corroboration, explanation, evolving probability, simplicity and a sharpened razor. *British Journal for the Philosophy of Science*, 19(2):123–143.

Harman, G. H. (1965). The inference to the best explanation. *Philosophical Review*, 74(1):88–95.

James, W. (1896). The will to believe. *The New World: A Quarterly Review of Religion, Ethics and Theology*, V:327–347.

Keil, F. C. (2006). Explanation and understanding. *Annual Review of Psychology*, 57:227–254.

Lange, M. (2022). Putting explanation back into 'inference to the best explanation'. *Noûs*, 56(1):84–109.

Lesage, S. and Brice, A. (2009). Parkinson's disease: From monogenic forms to genetic susceptibility factors. *Human Molecular Genetics*, 18(R1):R48–R59.

Lipton, P. (2001a). Is explanation a guide to inference? a reply to Wesley C. Salmon. In Hon, G. and Rakover, S. S., editors, *Explanation: Theoretical Approaches and Applications*, pages 93–120. Kluwer Academic, Dordrecht.

Lipton, P. (2001b). What Good is An Explanation? In Hon, G. and Rakover, S. S., editors, *Explanation: Theoretical Approaches and Applications*, pages 43–59. Kluwer Academic, Dordrecht.

Lipton, P. (2004). *Inference to the Best Explanation*. Routledge, New York, NY, 2nd edition.

Lombrozo, T. (2006). The structure and function of explanations. *Trends in Cognitive Sciences*, 10(10):464–470.

McCain, K. and Poston, T. (2019). Dispelling the disjunction objection to explanatory inference. *Philosophers' Imprint*, 19(36).

McGrew, T. (2003). Confirmation, heuristics, and explanatory reasoning. *British Journal for the Philosophy of Science*, 54(4):553–567.

Musgrave, A. (1988). The Ultimate Argument for Scientific Realism. In Nola, R., editor, *Relativism and Realism in Science*, pages 229–252. Kluwer Academic, Dordrecht.

Nandipati, S. and Litvan, I. (2016). Environmental exposures and Parkinson's disease. *International Journal of Environmental Research and Public Health*, 13(9):881.

Popper, K. R. (1959). *The Logic of Scientific Discovery*. Hutchinson, London.

Ross, L. N. (2020). Multiple realizability from a causal perspective. *Philosophy of Science*, 87(4):640–662.

Ross, L. N. (Forthcoming, 2022). Explanation in contexts of causal complexity: Lessons from psychiatric genetics. *Minnesota Studies in the Philosophy of Science*, From Biological Practice to Scientific Metaphysics.

Salmon, W. C. (2001). Explanation and confirmation: A Bayesian critique of Inference to the Best Explanation. In Hon, G. and Rakover, S. S., editors, *Explanation: Theoretical Approaches and Applications*, pages 61–91. Kluwer Academic, Dordrecht.

Schupbach, J. N. (2017). Inference to the Best Explanation, cleaned up and made respectable. In McCain, K. and Poston, T., editors, *Best Explanations: New Essays on Inference to the Best Explanation*, pages 39–61. Oxford University Press, Oxford.

Schupbach, J. N. (2019). Conjunctive explanations and Inference to the Best Explanation. *TEOREMA*, 38(3):143–162.

Schupbach, J. N. (2022). *Bayesianism and Scientific Reasoning*. Elements in the Philosophy of Science. Cambridge University Press, Cambridge.

Schupbach, J. N. and Sprenger, J. (2011). The logic of explanatory power. *Philosophy of Science*, 78(1):105–127.

Sober, E. (2015). *Ockham's Razors: A User's Manual*. Cambridge University Press, Cambridge.

van Fraassen, B. C. (1989). *Laws and Symmetry*. Oxford University Press, New York.