Inference to the Best Explanation, Cleaned Up and Made Respectable*

Jonah N. Schupbach
Philosophy, University of Utah
jonah.n.schupbach@utah.edu

Abstract

Despite decades of focused philosophical investigation, Inference to the Best Explanation (IBE) still lacks a precise articulation and compelling defense. The primary reason for this is that it is not at all clear what it means for a hypothesis to be the best available explanation of the evidence. This paper first seeks to rectify this problem by developing a formal explication of the explanatory virtue of power. A resulting account of IBE is then evaluated as a form of uncertain inference. Overall, this paper offers a precise account and novel defense of one important version of IBE.

Keywords: Inference to the Best Explanation, explanatory reasoning, Bayesianism, explanatory power, power, inductive cogency.

Inference to the Best Explanation (IBE) is a form of uncertain inference in which one reasons to a hypothesis based upon the premise that it provides a better potential explanation of some given evidence than any other available, competing hypothesis. When inferring the best explanation, one regards the explanatoriness of a hypothesis as good reason to favor that hypothesis. In this way, IBE links the explanatory value of a hypothesis to its epistemic value.

Philosophers and psychologists alike emphasize the widespread use and intuitive appeal of IBE in human reasoning (Harman, 1965; Lipton, 2004; Keil, 2006; Lombrozo, 2006; Douven, 2011; Douven and Schupbach, 2015b). In everyday affairs, people often reason to hypotheses based on their explanatory value; I might, for example, infer that my train has not yet come through the station because this hypothesis better explains the large number of people standing on the platform than any other plausible, competing hypothesis. And the applicability of IBE stretches far beyond the mundane. Scientists often infer to the best explanation; geologists may infer the occurrence of an earthquake millions of years ago because this event would, more than any other plausible hypothesis, explain various deformations in layers of bedrock. Court cases and forensic studies are decided to various degrees using IBE. This is true also of diagnostic procedures, whether performed by

*I owe special thanks to David Danks, John Earman, David Glass, Edouard Machery, Kevin McCain, Lydia McGrew, Ryan Muldoon, John Norton, Ted Poston, Jan Sprenger, and Rev. Michael van Opstall for helpful comments pertaining to this project. I am doubly grateful to Jan Sprenger, who additionally served as my co-author on the appendix.

clinicians or auto mechanics. Philosophers themselves often rely on IBE when debating some of the most venerable topics in the history of philosophy.[1] In all of these cases across domains, people favor hypotheses on account of their ability to explain evidence.

Despite its ubiquity and apparent cogency, IBE has a stormy history. It is difficult to think of another form of inference that has been, at once, so heartily defended by its champions and disparaged by its critics. Harman (1965, p. 88) boldly claims that IBE is the "basic form of nondeductive inference," having normative and conceptual priority over other forms of uncertain inference. Fumerton (1980) argues for the opposite claim that IBE is no more than an incomplete description of simpler forms of induction, having no independent epistemic merit. Van Fraassen (1989, pp. 142-43) famously offers the "bad lot" objection against IBE: IBE *assumes without argument* that the true hypothesis is likely to be one of the hypotheses under consideration. The upshot is that it can hardly be said to give us a reliable vehicle for inferring to conclusions that are more probably true.[2]

Of all the objections put to IBE, however, there is one that is most fundamental. The worry, expressed by the proponents and opponents of IBE alike, is that despite decades of serious philosophical investigation, the specific nature of IBE is still up for grabs. In the words of one of IBE's foremost supporters (Lipton, 2004, p. 2), "[IBE] is more a slogan than an articulated philosophy." This worry is of primary importance because it needs to be addressed before IBE's more specific vices and virtues may be explored; who is to say whether Harman, Fumerton, van Fraassen, and others are correct in their evaluations of IBE so long as this inference form has no clear articulation?

This paper, first of all, attempts to rectify this situation by specifying more precisely the nature of IBE. The most significant roadblock currently standing in the way of a clear account of IBE is our lack of understanding regarding the concept(s) of explanatoriness. The key premise of any instance of IBE claims a difference in explanatoriness between available potential explanations. Yet, the notion of explanatoriness is ambiguous. Section 1 accordingly distinguishes one particular version of IBE by first explicating precisely one prevalent sense of explanatoriness.

This paper is not merely interested in the clear *articulation* of IBE however, but also in its *evaluation*. To this end, Section 2.1 argues that the specific version of IBE introduced in Section 1 is cogent, meaning that its premise always lends epistemic support to its conclusion. Section 2.2 goes further and defends, through a series of computer simulations, IBE as a respectably reliable mode of inductive inference (at least when compared to the somewhat less contentious case of Bayesian inference).

1 IBE, Cleaned Up

The key premise of any particular inference to the best explanation refers to a difference in explanatoriness—or explanatory goodness—between considered hypotheses. But

[1] To take a small but informative sample: In the philosophy of religion, several well-known arguments for and against the existence of God are instances of IBE (e.g., Swinburne 2004, p. 20). Some epistemologists claim that IBE provides us with our best response to various forms of skepticism (e.g., Vogel 1990). In the philosophy of science, arguments to the existence of unobservables as well as arguments for scientific progress generally have the form of IBE (e.g., Putnam 1975 and Psillos 1999). And the same can be said of debunking arguments in ethics, and arguments for realist positions in metaethics and metaphysics—witness Lewis's (1986) central argument to possible worlds realism.

[2] See (Schupbach, 2014) for a recent response to the bad lot objection. Douven and Schupbach (2015a) additionally offer a brief response to van Fraassen's claim that IBE is a poor form of inference insofar as it commits the probabilistic, epistemic agent to diachronically incoherent updates.

explanatoriness is famously evaluated along different dimensions, corresponding to the various acclaimed explanatory virtues. Potential explanations may be prized for their great simplicity, unification, generality, power, or some combination of these (or other) virtues. One immediate consequence of this, often overlooked by IBE's commentators, is that the nature of an inference to the best explanation will depend upon the notion of explanatoriness at work therein. As a general category, IBE is polymorphous. There are at least as many distinct forms of IBE as there are distinct senses in which a hypothesis may be judged more explanatory than others; Inference to the Most Unifying Potential Explanation, for example, differs (prima facie, quite substantially) from Inference to the Simplest Potential Explanation.[3]

This basic point gives rise to a concern for generalist accounts and evaluations of IBE (i.e., much of the extant work on IBE). Such accounts gloss over potentially crucial differences between versions of IBE, confounding any attempt to evaluate seriously any specific version—the normative upshot of Inference to the Most Unifying Potential Explanation plausibly differs from that of Inference to the Simplest Potential Explanation. Any careful articulation and evaluation of IBE must rather build upon a precise account of the notion of explanatoriness determining what it takes for a potential explanation to be best. In the remainder of this paper, I heed this advice by focusing my sights on one particular acclaimed explanatory virtue and the corresponding version of IBE.[4]

1.1 Explanatoriness as Power

Our aim is to distinguish a particular version of IBE by first explicating an important notion of explanatory goodness. The result will be more interesting to the extent that the notion of explanatory goodness we focus on is one that reasoners indeed have in mind on some of the occasions in which they infer best explanations. With that in mind, we take a cue from C. S. Peirce's (1935, 5.189) description of explanatory inference (or "abduction"):

> Long before I first classed abduction as an inference it was recognized by logicians that the operation of adopting an explanatory hypothesis—which is just what abduction is—was subject to certain conditions. Namely, the hypothesis cannot be admitted, even as a hypothesis, unless it be supposed that it would account for the facts or some of them. The form of inference, therefore, is this:
>
> > The surprising fact, C, is observed;
> > But if A were true, C would be a matter of course;
> > Hence, there is reason to suspect that A is true.

According to Peirce, an inference in which one adopts an explanatory hypothesis begins when a "surprising fact" calls out for explanation. A hypothesis is put forth then, which must render the surprising fact a "matter of course." The key idea here is that a hypothesis explains some surprising fact well if it is able to render that fact unsurprising (i.e.,

[3] If one is a pluralist about the nature of explanation itself, then varieties of IBE may further multiply, with Inference to the Best Causal-Mechanical Explanation, for example, differing from Inference to the Best Covering Law Explanation, and so on. Whether these are differences that make a difference to the logic of IBE not already captured by the distinct notions of explanatoriness is an important question. Regardless, my focus in this paper will be on one particular brand of IBE distinguished by a single explanatory virtue at work in its central premise.

[4] None of this is meant to suggest that all precisely articulated notions of explanatoriness—and corresponding species of IBE—will only refer to one explanatory virtue. Plausibly, many instances of IBE involve a notion of explanatoriness that effectively strikes a balance between several distinct virtues. Any informative evaluation of this brand of IBE must build upon a precise account of what these virtues are and how they are balanced.

expected). Let us call Peirce's notion of explanatoriness, having to do with a hypothesis's ability to make evidence unsurprising, "power".

Reasoners often assess how explanatory a hypothesis is with respect to some evidence by gauging its power over that evidence (Schupbach and Sprenger, 2011, p. 108). Indeed, this particular notion of explanatoriness is so prevalent in instances of IBE that Peirce just seems to *identify* power with the general notion of explanatoriness in the above passage. While Peirce is surely wrong to suggest that we *always* adopt explanatory hypotheses on the basis of their power over explananda,[5] it does seem that this virtue adequately describes the notion of explanatory goodness at work in many applications of IBE. Accordingly, we focus in the rest of this paper on applications of IBE in which explanatoriness is evaluated purely as power.

To develop a precise explication of power, we start with Peirce's idea that an explanatory hypothesis has power over some surprising explanandum if it is able to render that explanandum unsurprising. This thought naturally lends itself to a subtler condition for an explication of power: a hypothesis has power over a proposition *to the extent that* it makes that proposition less surprising—or more expected—than it otherwise was. So, a geologist will favor a prehistoric earthquake as a powerful explanation of certain observed deformations in layers of bedrock to the extent that deformations of that particular character, in that particular layer of bedrock, and so on would be less surprising given the occurrence of such an earthquake. This condition is not a mere restatement of Peirce's idea. For one thing, given this condition, a hypothesis may provide a powerful explanation of a surprising proposition and still not render it a matter of course in any sense; i.e., a hypothesis may make a proposition much less surprising while still not making it unsurprising. Additionally, this subtler condition does not suggest that a proposition must be surprising in order to be explained; a hypothesis may make a proposition much less surprising (or more expected) even if the latter is not so surprising to begin with.

This condition may be used to motivate further conditions for an account of power. First, just as (positive) power comes with a decrease in surprise, one might say that a hypothesis has "negative power" over some proposition to the extent that it makes that proposition *more* surprising. I would judge the hypothesis that my train has already come through the station to be a terrible explanation of the large number of people standing on the platform; this is because I know that the majority of people in the station at this time of day are there to catch this particular train; my train typically leaves behind an empty platform. This hypothesis thus has negative power; if adopted, the crowd that I observe before me is even *more* surprising than it already was.

Given the above, a hypothesis lacks all (positive or negative) power whatever relative to some given explanandum if the latter is neither more nor less surprising in light of that hypothesis. The perceived motions of the planet Uranus are less surprising in light of the hypothesized existence of Neptune, but they are neither more nor less surprising given that my train has not yet passed through the station. The latter hypothesis is simply impotent with respect to that explanandum.

Insofar as a hypothesis has power over a proposition to the extent that it renders the latter unsurprising, one might additionally conclude that a hypothesis provides a *maximally* powerful explanation of some proposition just when it would lead one to expect that proposition to be true with certainty; this occurs when the hypothesis implies the

[5]After all, sometimes we infer best explanations based on their having virtues best describable as monadic properties (as opposed to relational properties between these hypotheses and evidence), simplicity being the most obvious example.

truth of that proposition. On the other hand, a *minimally* powerful explanation of some known proposition is one that renders the latter maximally surprising, and this occurs when the hypothesis implies that the proposition in question is false.

Finally, the less surprising a proposition's truth is in light of a hypothesis, the more surprising is its falsity. Given the above, this means that the more power a hypothesis has over a proposition, the less it has over the negation of that proposition. To summarize then, the intuitive starting point provided by Peirce can naturally be extended so that it provides the following compelling conditions for an explication of power:

Condition 1: A hypothesis has *(positive) power* over a proposition to the extent that it decreases the degree to which that proposition is surprising (i.e., increases the degree to which we expect that proposition to be true).

Condition 2: A hypothesis has *negative power* over a proposition to the extent that it increases the degree to which that proposition is surprising.

Condition 3: A hypothesis has *no power* over (i.e., is *impotent* with respect to) a proposition if and only if the latter is neither more nor less surprising in light of that hypothesis.

Condition 4: A hypothesis has *maximal power* over a proposition if and only if it leads us to expect with certainty that the proposition is true.

Condition 5: A hypothesis has *minimal power* over a proposition if and only if it leads us to expect with certainty that the proposition is false.

Condition 6: The more power a hypothesis has relative to a proposition, the less it has relative to the negation of that proposition.

1.2 The Measure of Power $\mathcal{E}$

The task of this section of the paper will be to apply the above considerations in order to arrive at a precise explication of power. If one makes use of the probability calculus to clarify and interpret these conditions, then only one measure of power with a certain desirable mathematical structure satisfies a subset of **Conditions 1-6**. Hence, the intuitions pertaining to power presented in the previous section already suffice to pin down a formal account of this concept. This account then clarifies, in the precise language of the probability theory, what it takes for a hypothesis to provide the best available explanation, when explanatoriness is evaluated purely in terms of power.

The key interpretive move of this section is to formalize a decrease in surprise (increase in expectedness) as an increase in probability. This move may seem dubious depending upon one's interpretation of probability. Given a physical interpretation (e.g., a relative frequency or propensity interpretation), it would indeed be difficult to saddle such a psychological concept as surprise with a probabilistic account. However, when probabilities are themselves given a more psychological interpretation (whether in terms of simple degrees of belief or the more normative degrees of *rational* belief), this move makes sense. In this case, probabilities map neatly onto degrees of (rational) expectedness. Accordingly, given the inverse relation between surprise and expectedness (the more surprising a proposition, the less one expects it to be true), surprise is straightforwardly related to

probabilities: the observation that h decreases the degree to which e is surprising corresponds with the judgment that h increases the degree to which e is expected, expressed probabilistically by the inequality $Pr(e) < Pr(e|h)$.[6]

As part of its "desirable mathematical structure" (which we specify exactly with two purely formal conditions of adequacy in the appendix), we require that the degree of power that hypothesis h has over evidence e, $\mathcal{E}(e,h)$, be real-valued on the closed interval $[-1,1]$. In explanatory contexts, $\mathcal{E}(e,h) = 1$ ($\mathcal{E}$'s maximal value) is the value at which h is interpreted as a maximally powerful potential explanation of e. $\mathcal{E}(e,h) = -1$ indicates the minimal degree of power for h relative to e, where h is interpreted as providing a maximally powerful potential explanation for e being *false*. $\mathcal{E}(e,h) = 0$ is the "neutral point" at which h lacks any power relative to e and its negation.

What are the corresponding formal conditions under which $\mathcal{E}$ takes these values? Here is where **Conditions 1-6** become relevant. As noted, $\mathcal{E}(e,h)$ should take the value 0 precisely when h lacks any power relative to e (and $\neg e$). **Condition 3** specifies that this occurs if and only if e (and $\neg e$) is neither more nor less surprising in light of h. Given the inverse relation between surprise and probability, this condition is explicated as h and e being statistically irrelevant to one another: $Pr(e|h) = Pr(e)$, or equivalently (remembering that Pr is a regular probability measure and that e and h are contingent propositions), $Pr(h \wedge e) = Pr(h) \times Pr(e)$.

CA1 (Neutrality**):** $\mathcal{E}(e,h) = 0$ if and only if $Pr(h \wedge e) = Pr(h) \times Pr(e)$.

Normality also demands that $\mathcal{E}(e,h)$ takes a maximum value of 1 if and only if h is maximally powerful with respect to e. **Condition 4** clarifies that such will be the case precisely when h leads us to expect with certainty that e is true. Such a notion is straightforwardly formalized with the equality $Pr(e|h) = 1$.

CA2 (Maximality**):** $\mathcal{E}(e,h) = 1$ if and only if $Pr(e|h) = 1$.

Condition 6 above requires that as the power of h relative to e increases, that of h relative to $\neg e$ decreases. When explanatoriness is assessed as power, this amounts to the idea that the more h explains the truth of e, the less it explains its falsity. Maximality and Neutrality provide us with further rationale for this condition. Maximality tells us that $\mathcal{E}(e,h)$ should be maximal only if $Pr(e|h) = 1$. Importantly, in such a case, $Pr(\neg e|h) = 0$, and this value intuitively corresponds to the point at which we should expect $\mathcal{E}(\neg e,h)$ to be minimal (see **Condition 5** above). In other words, given Maximality, we see that $\mathcal{E}(e,h)$ takes its maximal value 1 precisely when $\mathcal{E}(\neg e,h)$ takes its minimal value -1 and vice versa. Also, we know from Neutrality that $\mathcal{E}(e,h)$ and $\mathcal{E}(\neg e,h)$ should always equal zero at the same point given that $Pr(h \wedge e) = Pr(h) \times Pr(e)$ if and only if $Pr(h \wedge \neg e) = Pr(h) \times Pr(\neg e)$. These considerations lead to the following requirement:

CA3 (Symmetry**):** $\mathcal{E}(e,h) = -\mathcal{E}(\neg e,h)$.

The final condition of adequacy appeals to a scenario in which degree of power is unaffected. If a hypothesis h_2 is impotent with respect to another hypothesis h_1, to some proposition e, and to any logical combination of h_1 and e, then **Condition 3** tells us that it does nothing to increase or decrease the degree to which these are surprising. In such

[6]The background knowledge term k always belongs to the right of the solidus "|" in Bayesian formalizations (e.g., $Pr(e|k) < Pr(e|h \wedge k)$). Nonetheless, here and in the remainder of this paper, I leave k implicit in all formalizations for ease of exposition.

a case, conjoining h_2 to h_1 will do nothing to increase or decrease the degree to which e is surprising in light of h_1. Given Neutrality, we can state this in other words: if h_2 has no *power* whatever relative to e, h_1, or any logical combination of e and h_1, then its presence will not affect the overall power of h_1 relative to e. This gives us the following condition:

CA4 (Irrelevant Conjunction**):** If $Pr(e \wedge h_2) = Pr(e) \times Pr(h_2)$ and $Pr(h_1 \wedge h_2) = Pr(h_1) \times Pr(h_2)$ and $Pr(e \wedge h_1 \wedge h_2) = Pr(e \wedge h_1) \times Pr(h_2)$, then $\mathcal{E}(e, h_1 \wedge h_2) = \mathcal{E}(e, h_1)$.

These four adequacy conditions conjointly determine a unique measure of power as stated in the following theorem (proof in the appendix).[7]

Theorem 1. *The only measure with a desirable mathematical structure that satisfies* **CA1-CA4** *is*

$$\mathcal{E}(e, h) = \frac{Pr(h|e) - Pr(h|\neg e)}{Pr(h|e) + Pr(h|\neg e)}.$$

Note that this measure also satisfies the conditions from Section 1.1 that were not needed in order to prove **Theorem 1**. **Conditions 1** and **2** require that power increases (decreases) as the degree to which e is surprising decreases (increases) in light of h. Put more formally, these conditions require that $\mathcal{E}(e, h) > 0$ to the extent that $Pr(e) < Pr(e|h)$. These conditions are satisfied by $\mathcal{E}$ given that $\mathcal{E}(e, h) > 0$ to the extent that $Pr(h|e) > Pr(h|\neg e)$, which in turn is true just to the extent that $Pr(e|h) > Pr(e)$.[8] **Condition 5** requires that power is minimal if and only if e is certainly false in light of h. This fact also follows necessarily from $\mathcal{E}$ given that $\mathcal{E}(e, h) = -1$ if and only if $Pr(e|h) = 0$.[9] Thus, these conditions determine for us an intuitively well-grounded, unique measure of power.[10]

With $\mathcal{E}$ in hand, we may formally articulate an important version of IBE. In cases where the premise that h provides the best available potential explanation of the evidence e can be restated as the claim that this hypothesis has more power over e than any competing hypothesis, we have that $\mathcal{E}(e, h) > \mathcal{E}(e, h_i)$ for any and all of h's explanatory competitors h_i. The corresponding full version of IBE, which we can denote IBE$_p$ ("p" designating the notion of explanatoriness as power), has the following form:

[7] Measure $\mathcal{E}$ is structurally equivalent to Kemeny and Oppenheim's (1952) measure of "factual support,"

$$F(h, e) = \frac{Pr(e|h) - Pr(e|\neg h)}{Pr(e|h) + Pr(e|\neg h)},$$

which itself is ordinally equivalent to the log-likelihood measure of incremental confirmation $L(h, e) = \log[Pr(e|h)/Pr(e|\neg h)]$ (Good, 1983; Fitelson, 1999). The key difference between $\mathcal{E}$ and these measures is in their interpretation and application; $\mathcal{E}(e, h)$ is $F(h, e)$ with h and e interchanged. This difference is significant, as the conditions of adequacy used to motivate the measures differ. It is easy to verify that F and L at least fail to satisfy **CA2** and **CA3**, making them unsuitable for measuring power—though both are among the most plausible measures of incremental confirmation. This is appropriate, since these conditions properly constrain measures of power, but they make little sense as constraints on measures of incremental confirmation.

[8] This is easy to see in light of the fact that

$$\frac{Pr(h|e)}{Pr(h|\neg e)} = \frac{Pr(e|h)}{Pr(e)} \times \frac{1 - Pr(e)}{1 - Pr(e|h)}.$$

[9] $\mathcal{E}(e, h) = -1$ just in case $\mathcal{E}(e, h) = -Pr(h|\neg e)/Pr(h|\neg e)$. But this equality holds only if $Pr(h) \neq 0$ and $Pr(h|e) = 0$ which implies that $Pr(e|h) = 0$.

[10] Alternative uniqueness theorems providing different axiomatic foundations for $\mathcal{E}$ may be found in (Schupbach and Sprenger, 2011) and (Cohen, 2015). That $\mathcal{E}$ can be defended via several distinct uniqueness theorems helps alleviate the worry that our result is driven by a faulty condition of adequacy. Schupbach and Sprenger (2011) also provide further support for $\mathcal{E}$ via several theorems, which show that $\mathcal{E}$ matches clear intuitions about power. $\mathcal{E}$ gains yet another line of support as an accurate measure of (explanatory) power in (Schupbach, 2011), where I show experimentally that $\mathcal{E}$ is a good predictor of actual human judgments of explanatoriness.

$$\text{(IBE}_p\text{)} \qquad \begin{array}{l} e \\ \mathcal{E}(e,h) > \mathcal{E}(e,h_i), \text{ for any } h_i \text{ competing with } h \\ \hline \therefore \quad h \end{array}$$

The question of whether or not this species of IBE is a cogent inference form is now more tractable. We investigate this question in the next section.

2 IBE, Made Respectable

The nature of IBE changes depending on the precise sense of explanatoriness at work in its central premise. And the evaluation of IBE naturally follows suit. Any informative evaluation of IBE will attend to a precisely explicated species of IBE. Correspondingly, any attempt to evaluate (defend or criticize) IBE in general without first precisely articulating the version of IBE will be at least as confused as the general category of explanatoriness itself. Once different versions of IBE are disentangled, it may well turn out that some of these are epistemically defensible and others not. This will depend most obviously on whether the notion of explanatoriness at work in a particular version of IBE carries any genuine epistemic force.

This section evaluates IBE$_p$, the version of IBE instantiated when explanatoriness is evaluated as power. The strategy is as follows: Section 2.1 first defends IBE$_p$ as cogent, arguing that there is a clear sense in which its premises always support its conclusion. Section 2.1 also suggests that IBE$_p$ is useful as an informal heuristic allowing us to approximate sound probabilistic reasoning. Section 2.2 thus asks just how reliable this inference form is when compared to Bayesian inference. It turns out that IBE$_p$ stacks up quite well. Indeed, under certain (arguably common) conditions, IBE$_p$ provides a *more* reliable mode of inference than that based on sound probabilistic reasoning.

2.1 Some Implications of Power

As a first step toward evaluating IBE$_p$, it is enlightening to spell out the probabilistic implications of a single hypothesis h having positive power over evidence e, $\mathcal{E}(e,h) > 0$. Filling in the details of $\mathcal{E}$, this judgment can be shown to have the following probabilistic consequences (where '$\Leftrightarrow$' symbolizes interderivability):[11]

$$\begin{aligned} &\frac{Pr(h|e) - Pr(h|\neg e)}{Pr(h|e) + Pr(h|\neg e)} > 0 \\ &\quad \Leftrightarrow \; Pr(h|e) > Pr(h|\neg e) \\ &\quad \Leftrightarrow \; \frac{Pr(e|h)}{Pr(e)} > \frac{Pr(\neg e|h)}{Pr(\neg e)} \\ &\quad \Leftrightarrow \; Pr(e|h) - Pr(e|h)Pr(e) > Pr(e) - Pr(e|h)Pr(e) \\ &\quad \Leftrightarrow \; Pr(e|h) > Pr(e) \\ &\quad \Leftrightarrow \; Pr(e|h) > Pr(e|\neg h) \qquad \text{(L)} \\ &\quad \Leftrightarrow \; Pr(h|e) > Pr(h) \qquad \text{(C)} \end{aligned}$$

[11] Recall that Pr is a regular probability measure and that e and h are contingent propositions.

(L) and (C) are especially significant; these results tell us that positive power can be probabilistically represented using either a likelihood comparison or the notion of incremental confirmation respectively. We will have more to say, in the rest of this section, about the likelihood comparisons indicated by certain explanatory judgments. (C) reveals that, to the extent that a hypothesis is able to provide a powerful explanation of the evidence in question, that evidence confirms (raises the probability of) that hypothesis. This suggests a particular sense in which the judgment that a hypothesis is positively explanatory of the evidence does constitute a reason to favor that hypothesis.

IBE$_p$'s central premise does not claim, however, that h has positive power over e. Instead, it makes the comparative claim that h offers a more powerful potential explanation of e than does any competing hypothesis h_i, $\mathcal{E}(e,h) > \mathcal{E}(e,h_i)$. Filling in the probabilistic details of $\mathcal{E}$, this explanatory judgment is explicated as follows:

$$\begin{aligned}
&\frac{Pr(h|e) - Pr(h|\neg e)}{Pr(h|e) + Pr(h|\neg e)} > \frac{Pr(h_i|e) - Pr(h_i|\neg e)}{Pr(h_i|e) + Pr(h_i|\neg e)} \\
&\quad \Leftrightarrow \ \frac{Pr(h|e)}{Pr(h|\neg e)} > \frac{Pr(h_i|e)}{Pr(h_i|\neg e)} \\
&\quad \Leftrightarrow \ \frac{Pr(e|h)Pr(\neg e)}{Pr(\neg e|h)Pr(e)} > \frac{Pr(e|h_i)Pr(\neg e)}{Pr(\neg e|h_i)Pr(e)} \\
&\quad \Leftrightarrow \ Pr(e|h) - Pr(e|h)Pr(e|h_i) > Pr(e|h_i) - Pr(e|h)Pr(e|h_i) \\
&\quad \Leftrightarrow \ Pr(e|h) > Pr(e|h_i)
\end{aligned}$$

$\mathcal{E}$ thus reveals that, in multiple-hypothesis settings, the hypothesis that offers the most powerful potential explanation of some proposition will be the one that makes that proposition the most likely. In Bayesian terms, the hypothesis judged to provide the best explanation will have the greatest corresponding *likelihood* of any explanatory hypothesis considered. This result clarifies the nature of the reason that favors the most explanatory hypothesis over those that are explanatorily inferior. A hypothesis's likelihood ($Pr(e|h)$) is positively related to its overall probability in light of the evidence ($Pr(h|e)$) as can be seen via Bayes's Theorem:

$$Pr(h|e) = \frac{Pr(h) \times Pr(e|h)}{Pr(e)}$$

Holding all else constant, the greater a hypothesis's corresponding likelihood, the greater its probability given e.

Furthermore, when comparing various hypotheses with respect to the same evidence e (as in instances of IBE), $Pr(e)$ is the same regardless of which hypothesis one has in mind. Accordingly, we can say that if h offers the most powerful of the available potential explanations of e, then it is also the most probable hypothesis given e so long as it is at least as plausible as its competitors apart from considerations of e—i.e., so long as $Pr(h) \geq Pr(h_i)$, for all rival hypotheses h_i. Of course, the most explanatory hypothesis may be less plausible apart from considerations of e as compared to other hypotheses; in this case, it is possible for h to provide the best explanation and *not* be the most probable available hypothesis overall. Nonetheless, it is also true that the power of h over e may be greater than that of rival hypotheses to such an extent that it overcomes the fact that $Pr(h)$ is comparatively low and makes it the case that h *is* the most probable competing hypothesis.

In general then, the judgment that a hypothesis provides the most powerful explanation of the evidence provides us with a good reason to favor that hypothesis. This is because comparative judgments of power bear witness to relative degrees of statistical relevance between e and considered hypotheses. The hypothesis with the greatest power over e corresponds to that which is the most statistically relevant to e, implying that this hypothesis has the greatest corresponding likelihood. A hypothesis's likelihood is positively related to its overall probability in light of the evidence. The judgment that a hypothesis provides the best available explanation of the evidence *does* therefore constitute reason to favor that hypothesis over its explanatory competitors, because this judgment reflects probabilistic information that has a positive bearing on h's overall probability in light of e. In this sense, IBE$_p$ is manifestly a cogent form of nondeductive inference.

At this point, it is important to bear in mind what a general defense of a nondeductive inference *form* can and cannot provide. Precisely in virtue of its nondeductive nature, such a form cannot fairly be criticized for not always guiding us from true premises to a true conclusion. Instead, the most that we can *generally* require of such an inference form is that, whenever we instantiate it, we do end up with premises that—in some way, to some extent—positively support the conclusion. The above claim that IBE$_p$ is cogent thus amounts to the claim that any inference to the most powerful explanation's premises will provide positive support for the corresponding conclusion.[12]

2.2 What Computers can Teach us about IBE

The picture that arises out of the above defense of IBE$_p$'s cogency is that considerations of power have epistemic value on account of the role they play in reflecting important probabilistic information. When a person recognizes that a hypothesis has the most power over the evidence, that person has taken account of a fact with probabilistic ramifications in favor of that hypothesis. In this way, IBE$_p$ enables us to account for relevant probabilistic information when reasoning without necessarily having explicit awareness of the individual probabilities involved or even any working knowledge of probability theory. The foregoing investigation into the epistemic implications of power thus sheds new light on Peter Lipton's oft-repeated dictum that "explanatory loveliness is a guide to judgments of likeliness" (2004, p. 121).

IBE$_p$ describes a cogent inference form because the power of a hypothesis is a genuine epistemic virtue; all else being equal between competing hypotheses, the most powerful hypothesis will also be the most probable. But all else is seldom equal in real life. Consequently, in contexts where people typically make inferences to the most powerful explanations, it might be that, despite its cogency, this inference form is not very *useful*; though considerations of power reflect important probabilistic information in such contexts, IBE$_p$ may commonly misguide us because of the probabilistic information that these considerations ignore (viz., prior probabilities). Just how useful IBE$_p$ is depends *inter alia*

[12]Note that this is a far cry from claiming that the conclusion of any particular inference of this form is *justified*. Whether an inference *form* is cogent is determined at a general level—based upon whether there is a logical sense in which the sort of premises required by that form provide positive evidence for the sort of conclusion described. Whether a particular conclusion of an inference is *justified*, on the other hand, is not generally decidable. There must be at least some reason in favor of the conclusion of any particular instance of an inference form, if that form is cogent. However, other epistemic considerations may bear upon this conclusion in such a way that it is overall unjustified. Whether or not the conclusion of a particular such inference is justified is determined by the full epistemic details of one's context; whether or not IBE$_p$ is a cogent form of inference is not determined by such contextually specific factors.

on its potential for guiding us to true hypotheses despite selectively attending only to some of the relevant probabilistic information.

In this section, I use computer simulations—based closely upon those devised and reported by Glass (2010)—to model and compare the performance of IBE_p versus probabilistic reasoning for the sorts of everyday contexts in which people are inclined to infer most powerful explanations. The general methodology that these simulations employ is summarized in the following steps:

1. For each of a specified number n of competing (mutually exclusive) explanatory hypotheses, assign values of the prior probabilities ($Pr(h_i)$) and likelihoods ($Pr(e|h_i)$). Priors and likelihoods are drawn randomly from a normal and uniform distribution respectively (see discussion below for more details).

2. Using weights corresponding to the respective values of $Pr(h_i)$, randomly select the "true" hypothesis h_j from $h_1, h_2, ..., h_n$. Each h_i has a $Pr(h_i)$ chance of being selected.

3. Using the value of $Pr(e|h_j)$ (the likelihood associated with the true hypothesis), check whether e "occurs." If e occurs, continue with steps 4-6; otherwise, end this iteration.

4. Check which of the n hypotheses has the greatest power; i.e., find h_k where $\mathcal{E}(e, h_k) > \mathcal{E}(e, h_i)$ for all $i \neq k$.

5. Check which of the n hypotheses is the most probable in light of e; i.e., find h_l where $Pr(h_l|e) > Pr(h_i|e)$ for all $i \neq l$.

6. If $h_k = h_j$, count this as a case where the most explanatory hypothesis matches the true hypothesis; if $h_l = h_j$, count this as a case where the most probable hypothesis matches the true hypothesis.

Steps 1-6 constitute one iteration of the simulation. After a large number of repeated iterations, the simulation provides estimates of how often the hypothesis with the greatest power (relative to e) corresponds to the true hypothesis and how often the hypothesis with the greatest probability (conditional on e) corresponds to the true hypothesis. In either case, this is calculated as the number of times that one gets such a match divided by the number of instances in which e occurs.

The goal is for this procedure to model *real-world* contexts in which people are inclined to infer most powerful explanations, and thereby to give us an estimate of IBE_p's average, actual accuracy in such contexts. Whether one is able to accomplish this end (and precisely which real-world contexts are modeled) is contingent upon several assumptions built into the simulation. Two important decisions in particular constrain the model's proper application: (1) whether one includes a "catch-all" hypothesis, and (2) how exactly one assigns prior probabilities (values of $Pr(h_i)$) to the hypotheses.

Regarding (1), in general, if explanatory hypotheses h_1 through h_n are not only assumed to be mutually exclusive but also jointly exhaustive, then one's model will represent a situation in which it is known that one of these competing hypotheses must be true. In such a case, there is no need to include a "catch-all" hypothesis to represent all unimagined hypotheses. To take a simple example, one might be interested in inferring whether a particular coin is fair or biased by examining how well these respective hypotheses explain a series of observed coin flips. Given that the coin must either be fair or biased, there is no room to include a third, catch-all hypothesis.

However, there are many contexts in which it is not known with certainty that the true hypothesis is one of those considered. In order to represent this scenario, a model must include a catch-all hypothesis. Within the above simulation procedure, a catch-all hypothesis can be chosen as the true hypothesis h_j in step 2, but it cannot be chosen as the most explanatory (h_k in step 4) or probable (h_l in step 5) of the available competing hypotheses for the simple reason that it is not considered by—and therefore not available to—the reasoner.

Decisions pertaining to (2) are the more difficult. How should one go about assigning prior probabilities to the explanatory hypotheses in these simulations if the goal is to model contexts in which people are inclined to infer most powerful explanations? Such probabilities must always sum to one,[13] but is there more to say than this? At least the following seems clear: the set of hypotheses reasoners are willing to entertain in such contexts will be determined in part by how plausible those hypotheses are to begin with. When faced with evidence in need of explanation, a person may be able to conjure up any number of alternative, explanatory hypotheses having various degrees of power over that evidence. But the fact that a given hypothesis is conjurable and powerful is not enough to place that hypothesis within the ranks of those that a reasoner is willing to infer. No matter how well I think that an ancient extraterrestrial visitation, for example, would explain the patterned deformations that I observe in layers of bedrock, I will not consider this hypothesis when inferring the best explanation; this is because, to my mind, that hypothesis is so implausible to begin with that it's not worth consideration. By contrast, insofar as someone believes that the extraterrestrial hypothesis *is* plausible, that person will find it appropriate to consider for potential inference.

This is particularly true when reasoners are inclined to rest all inferential weight on considerations of power. In such cases, considerations of prior plausibility are neglected. But people are not inclined to neglect such considerations when they weigh heavily for or against considered hypotheses. That is, it is plausible to think that people only allow power alone to do the inferential heavy lifting in cases where there is no substantial difference in prior plausibility that also weighs in favor of one of the hypotheses.

The upshot is that the hypotheses considered when people infer most powerful explanations will all typically be comparably plausible (though they might all have low probability—e.g., if there are a sufficiently large number of mutually exclusive hypotheses to consider). For the sake of modeling the usual IBE_p context then, the prior probabilities of the considered hypotheses are chosen in such a way that they tend to be closer in value to one another. This is only enforced for the *considered* hypotheses though; when a catch-all hypothesis is included in a simulated context, the prior probability of this catch-all hypothesis is allowed to stray from the values of the prior probabilities corresponding to the considered hypotheses.[14]

This basic simulation design was run for two distinct scenarios corresponding to the choice of whether or not to include a catch-all hypothesis. Within each of these two scenarios, a specific simulation was run for a particular number n of competing explanatory hypotheses (n ranging from 2 to 10). Any individual simulation included $1,000,000$

[13]This is true in either case regarding decisions about (1). If no catch-all hypothesis is required, then h_1 through h_n are mutually exclusive and jointly exhaustive, their prior probabilities thus necessarily summing to one. If a catch-all is required, then h_1 through h_n *plus the catch-all hypothesis* are mutually exclusive and jointly exhaustive, with prior probabilities thus summing to one.

[14]This is achieved by sampling prior probabilities randomly from a normal distribution ($\mu = .5$, $\sigma = .15$), choosing the prior probability of the catch-all randomly from a uniform distribution between 0 and 1, and then renormalizing so that the probabilities sum to 1.

repetitions to secure accuracy.

Results are shown in Figures 1 and 2. For a given number of hypotheses, these figures display the percentage of cases in which the most powerful hypothesis is true as compared to the percentage of cases in which the most probable hypothesis is true. For reference, the percentage accuracies of a random ("chance") guess from the lot of available hypotheses is also displayed. Figure 1 shows these results for contexts that do not include a catch-all hypothesis while Figure 2 shows the results corresponding to contexts that do.

Both figures reveal that percentage accuracies decrease as the number of hypotheses increases. This validates the intuitive idea that as the number of competing hypotheses increases, so does the number of ways in which one's inferred conclusion could go wrong. Hence, accuracy decreases when there are more hypotheses to which one can infer. Note, however, that IBE$_p$ and probabilistic reasoning are both unsurprisingly much more accurate in contexts with no catch-all hypothesis. This fact allows us to clarify one sense in which increasing the number of considered hypotheses could actually *increase* the respective accuracies of these inference rules. Each new hypothesis added to the lot of those considered decreases the probability of (i.e., the need for) a catch-all hypothesis; each such addition brings us a step closer to the special case where our considered hypotheses *partition* the space of possibilities, leaving the catch-all with zero probability. And as one moves closer to a context in which there is no space left for a catch-all in this way, the result may be an overall increase in accuracy. Thus, comparing Figures 1 and 2, the addition of an explanatory hypothesis that exhausts the remaining possibility space (so that there is no longer any need for a catch-all hypothesis) slightly *improves* the average accuracy of IBE$_p$ and probabilistic reasoning in all cases.

As can be seen from Figures 1 and 2, IBE$_p$ approximates probabilistic reasoning very well indeed, the average accuracy of the former being consistently only slightly less than that of the latter. More specifically, both in contexts that do and those that do not include a catch-all hypothesis, IBE$_p$'s accuracy is consistently, on average, only about 3% below that of probabilistic reasoning. To compare IBE$_p$'s reliability to that of probabilistic reasoning more directly, we can calculate its *relative* percentage accuracy (i.e., the percentage accuracy of IBE$_p$ divided by that of probabilistic reasoning). These results are displayed in Table 1. Again, the results suggest that IBE$_p$'s reliability is not much worse than that of probabilistic reasoning. Whether or not a context includes a catch-all, IBE$_p$ identifies the true hypothesis about 90% as often as probabilistic reasoning—averaging over the simulated contexts.

Thus far, our results suggest that IBE$_p$'s epistemic import is parasitic upon Bayesianism's. IBE$_p$ is cogent insofar as it gives us an informal handle on some, but not all, of the probabilistic information needed for Bayesian inference, and it is useful because it is nearly as reliable as the latter (and much more reliable than chance). Practically speaking, we might point out that IBE$_p$ seems eminently more useful to human reasoners than Bayesian inference insofar as it serves reasoners who are, for whatever reason, not able to apply probabilistic reasoning directly; still, if this is right, IBE$_p$ may be thought merely heuristically useful as a poor man's Bayesianism.

However, the above simulations incorporate an unrealistic, simplifying assumption that gives Bayesianism a substantial advantage. Specifically, these assume that an agent's prior probabilities perfectly match the objective chances of the various hypotheses being true. Thus, h_i's chance of being selected as the true hypothesis in any iteration of the simulation is determined straightforwardly by the value of $Pr(h_i)$. Relaxing this assumption by allowing agents to have inaccurate priors accordingly results in Bayesian

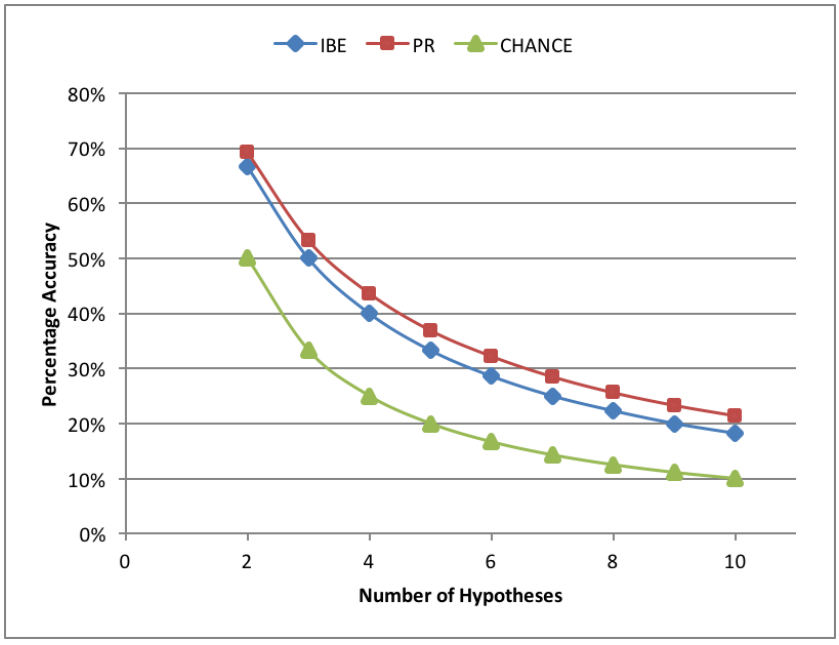

Figure 1: Percentage accuracies in contexts with no catch-all.

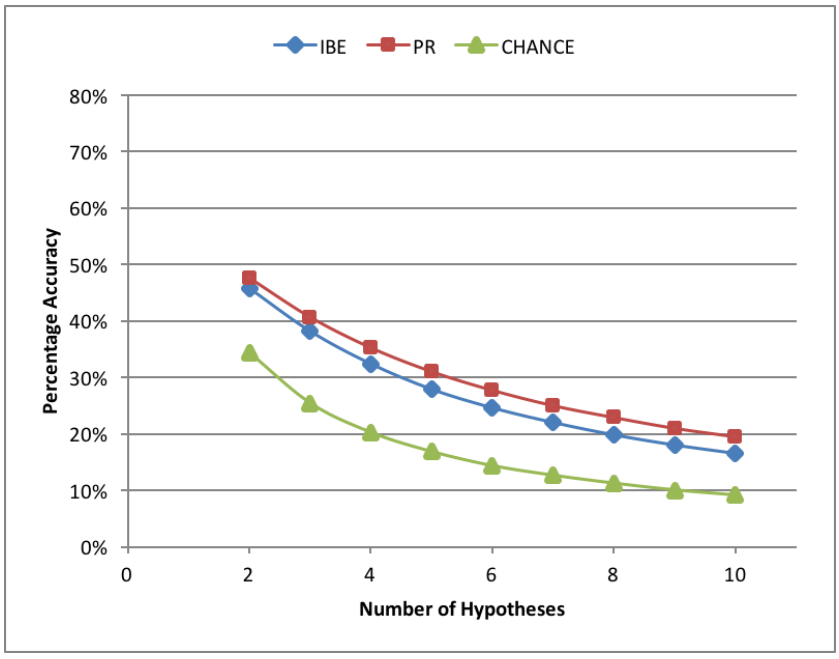

Figure 2: Percentage accuracies in contexts that include a catch-all.

n	No Catch-all	Catch-all
2	.9639	.9642
3	.9398	.9409
4	.9174	.9205
5	.9024	.9000
6	.8882	.8881
7	.8772	.8800
8	.8711	.8646
9	.8584	.8571
10	.8505	.8462

Table 1: Relative percentage accuracies of IBE_p (percentage accuracy of IBE_p / percentage accuracy of probabilistic reasoning).

reasoning having a worse reliability. By contrast, IBE_p neglects priors and ultimately puts all inferential weight on likelihood comparisons. And so, relaxing this assumption has no effect on IBE_p's reliability. The predicted upshot is that, as an agent's priors are allowed, on the average, to diverge from objective chances, IBE_p may become more reliable than probabilistic reasoning.

This is easily verified by complicating the above simulations as follows. Steps 1-3 remain the same, although the "prior probabilities" referred to in those steps are now interpreted as the objective chances that the various hypotheses are true. After these initial steps, each prior is calculated by adding to the corresponding chance the value of a normally-distributed random variable with mean 0 and specified standard deviation (and then renormalizing to ensure that they sum to 1). This standard deviation explicates the average error of the agent's prior probabilities. While the true hypothesis (and whether e occurs) is determined on the basis of the objective chances, the remaining steps calculate greatest power and posterior probability using the (erroneous) prior probabilities.

The above predictions are verified in the results of all variations—average accuracies for the specific case where $n = 2$, for standard deviations varying from .05 to .50, are shown in Figures 3 and 4. Both in contexts that do and those that do not include a catch-all hypothesis, the average reliability of probabilistic reasoning dips below that of IBE_p already with rather modest allowances for error in the priors—though it never dips below that of chance.

3 Conclusions

Past work on the nature and value of IBE largely treats this inference form as one unified category. However, once one remembers that explanatory goodness is evaluated on distinct dimensions that can (and often do) vary from case to case, this generalist perspective looks dubious and misleading. Different versions of IBE can be distinguished by the notions of explanatoriness at work in their respective central premises. And these are differences that plausibly matter a great deal to IBE's normative evaluation. Depending on how explanatory goodness is evaluated, IBE may or may not describe a respectable form of uncertain inference.

In Section 1 of this paper, I put forward a Bayesian explication of one specific sense of explanatory goodness, and I articulated precisely the corresponding version of IBE. Then,

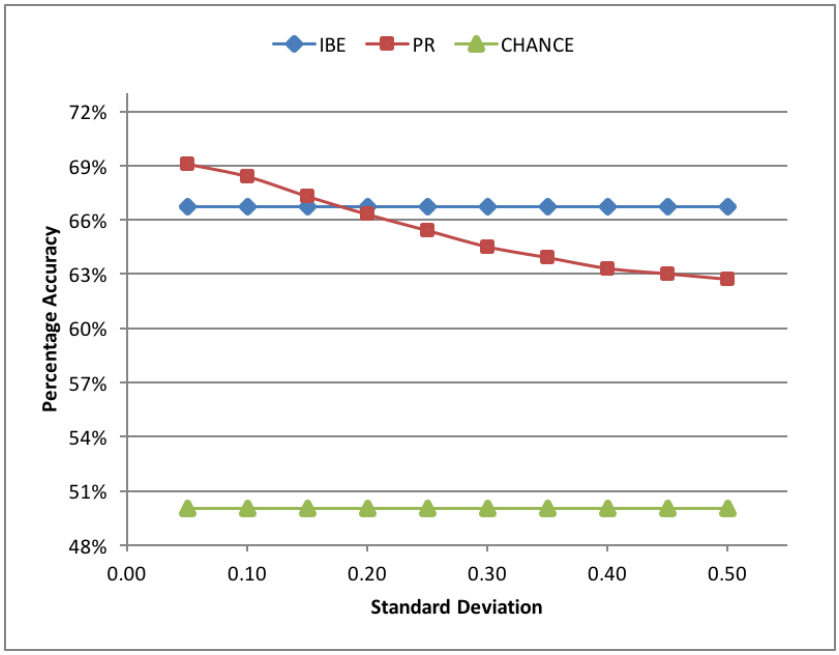

Figure 3: Percentage accuracies in contexts with no catch-all.

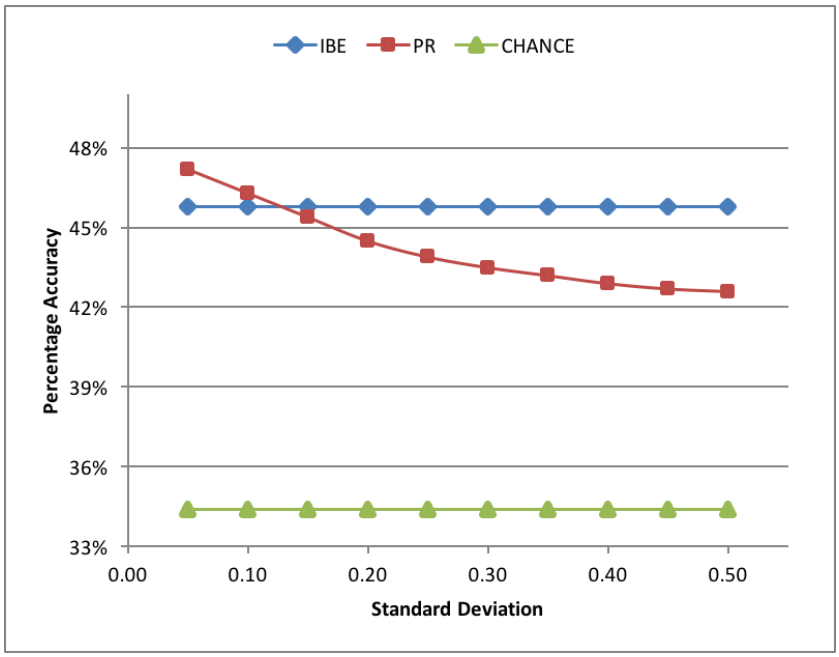

Figure 4: Percentage accuracies in contexts that include a catch-all.

in Section 2, I defended this version of IBE as inductively cogent and respectably reliable (at least when compared to Bayesian reasoning). At the start of his most well-known attack on IBE, van Fraassen (1989, p. 131) writes, "As long as [IBE] is left vague, it seems to fit much rational activity. But when we scrutinize its credentials, we find it seriously wanting." This paper demonstrates, to the contrary, that once we clearly articulate the nature of IBE via an explication of explanatoriness, this inference form can gain a sound new defense.

Appendix A. Uniqueness of $\mathcal{E}$

The mathematical structure that we require of our explicatum is specified in the following two formal conditions of adequacy:

Normality. For any probability space $(\Omega, \mathcal{A}, Pr(\cdot))$—where Pr is a regular probability measure—$\mathcal{E}$ is a measurable function from two contingent propositions $e, h \in \mathcal{A}$ to a real number $\mathcal{E}(e,h) \in [-1,1]$.

Structure. $\mathcal{E}$ is the ratio of two functions of $Pr(e \wedge h)$, $Pr(\neg e \wedge h)$, $Pr(e \wedge \neg h)$ and $Pr(\neg e \wedge \neg h)$, each of which are homogenous in their arguments to degree $k \geq 1$, where k is the smallest integer permitted by Normality and **CA1–CA4**.[15]

These conditions require that $\mathcal{E}(e,h)$ be probabilistic in nature and simple in a well-defined sense.

Theorem 1.[16] *The only measure that satisfies* Normality, Structure, *and* **CA1-CA4** *is*

$$\mathcal{E}(e,h) = \frac{Pr(h|e) - Pr(h|\neg e)}{Pr(h|e) + Pr(h|\neg e)}.$$

Notation. Let $x = Pr(e \wedge h)$, $y = Pr(e \wedge \neg h)$, $z = Pr(\neg e \wedge h)$ and $t = Pr(\neg e \wedge \neg h)$ with $x + y + z + t = 1$. Then, by Structure, $\mathcal{E}(e,h)$ has the form

$$f(x,y,z,t) = \frac{f_n(x,y,z,t)}{f_d(x,y,z,t)},$$

where $f_n(x,y,z,t)$ and $f_d(x,y,z,t)$ are homogeneous in their arguments to the same least degree $k \geq 1$.

[15] A function is homogenous to degree k iff multiplying its arguments all by the same factor c multiplies its value by c^k. The homogeneity requirement ensures that the functional form of $\mathcal{E}$ itself does not determine which of the terms ($Pr(e \wedge h)$, $Pr(\neg e \wedge h)$, $Pr(e \wedge \neg h)$, $Pr(\neg e \wedge \neg h)$) should have more weight. Representing $\mathcal{E}$ as the ratio of two functions serves the purpose of normalization. $Pr(e \wedge h)$, $Pr(\neg e \wedge h)$, $Pr(e \wedge \neg h)$ and $Pr(\neg e \wedge \neg h)$ fully determine the probability distribution over the truth-functional compounds of e and h, so it is appropriate to represent $\mathcal{E}$ as a function of them. Finally, the requirement that $\mathcal{E}$ be the ratio of two functions, each having "the least possible degree $k \geq 1$" reflects a minimal and well-defined simplicity assumption akin to those advocated by Carnap (1950, Chapter 1) and Kemeny and Oppenheim (1952, p. 315). Any reader skeptical of simplicity's place in these conditions of adequacy is referred to (Schupbach and Sprenger, 2011), which contains an alternative uniqueness proof from different conditions of adequacy (not including Structure).

[16] This theorem and its proof are closely related to, and were indeed inspired by, Kemeny and Oppenheim's (1952) discussion and proof of their Theorem 17.

Lemma 1. *There is no f with f_n, f_d of degree 1 that satisfies* Normality, Structure, *and* **CA1-CA4**; *i.e., $k \neq 1$.*

Proof. Let $k = 1$. Then $f_n(x, y, z, t)$ has the form $ax + by + cz + dt$ (a, b, c and d are coefficients). By **CA1**, $f(x, y, z, t) = 0$ (and so $ax + by + cz + dt = 0$) if and only if $x = Pr(h \wedge e) = Pr(h) \times Pr(e) = (x + z)(x + y)$. Now we can show that this biconditional cannot be generally satisfied by locating four different parameter settings of (x, y, z, t) that each satisfy $x = (x + z)(x + y)$ but across which there are no (non-zero) coefficients that satisfy $ax + by + cz + dt = 0$. The following four parameter settings suffice: $(1/2, 1/4, 1/6, 1/12)$, $(1/2, 1/3, 1/10, 1/15)$, $(1/2, 3/8, 1/14, 3/56)$, and $(1/4, 1/4, 1/4, 1/4)$. Since these vectors are linearly independent (i.e., their span has dimension 4), the only way to satisfy $ax + by + cz + dt = 0$ across these cases is if $a = b = c = d = 0$. □

Lemma 2. **CA4** *entails that for any value of $\beta \in (0, 1)$,*

$$f(x, y, z, t) = f(\beta x, y + (1 - \beta)x, \beta z, t + (1 - \beta)z). \tag{1}$$

Proof. This lemma is a consequence of **CA4**, which describes conditions under which degrees of power must be the same. For any $x, y, z, t \in [0, 1]$ such that $x + y + z + t = 1$, allow that there could be an e and h_1 such that $x = Pr(e \wedge h_1), y = Pr(e \wedge \neg h_1), z = Pr(\neg e \wedge h_1)$, and $t = Pr(\neg e \wedge \neg h_1)$. For any β, allow that there may be an h_2 that satisfies the antecedent conditions of **CA4** and such that $Pr(h_2) = \beta$.[17]

With regards to such an e, h_1, and h_2, **CA4** requires that $\mathcal{E}(e, h_1 \wedge h_2) = \mathcal{E}(e, h_1)$. We can show that this is equivalent to (1) by establishing the following:

$$\begin{aligned} \beta x &= Pr(e \wedge (h_1 \wedge h_2)) & \qquad y + (1 - \beta)x &= Pr(e \wedge \neg(h_1 \wedge h_2)) \\ \beta z &= Pr(\neg e \wedge (h_1 \wedge h_2)) & \qquad t + (1 - \beta)z &= Pr(\neg e \wedge \neg(h_1 \wedge h_2)). \end{aligned}$$

These equations are demonstrated straightforwardly, making use of the antecedent conditions of **CA4**. For example, these require that $Pr(e \wedge (h_1 \wedge h_2)) = Pr(h_2)Pr(e \wedge h_1) = \beta x$ (establishing the first equation above). This condition entails that $Pr(e \wedge h_1 \wedge \neg h_2) = Pr(\neg h_2)Pr(e \wedge h_1)$, allowing us to demonstrate the second equation:

$$\begin{aligned} Pr(e \wedge \neg(h_1 \wedge h_2)) &= Pr[(e \wedge \neg h_1) \vee (e \wedge \neg h_2)] \\ &= Pr(e \wedge \neg h_1) + Pr(e \wedge \neg h_2) - Pr(e \wedge \neg h_1 \wedge \neg h_2) \\ &= Pr(e \wedge \neg h_1) + Pr(e \wedge h_1 \wedge \neg h_2) \\ &= Pr(e \wedge \neg h_1) + Pr(\neg h_2)Pr(e \wedge h_1) = y + (1 - \beta)x \end{aligned}$$

The other two equations are demonstrated *mutatis mutandis*. □

Proof of Theorem 1 (Uniqueness of $\mathcal{E}$). Lemma 1 shows that there are no f_n, f_d of degree 1 that satisfy our desiderata. Here, I show that there is exactly one ratio of such functions of degree $k = 2$, which completes the proof (given the formal requirements set out in Structure). If $k = 2$, $f(x, y, z, t)$ takes the form

$$\frac{f_n(x, y, z, t)}{f_d(x, y, z, t)} = \frac{ax^2 + bxy + cy^2 + dxz + eyz + gz^2 + ixt + jyt + rzt + st^2}{\bar{a}x^2 + \bar{b}xy + \bar{c}y^2 + \bar{d}xz + \bar{e}yz + \bar{g}z^2 + \bar{i}xt + \bar{j}yt + \bar{r}zt + \bar{s}t^2} \tag{2}$$

[17] In Bayesian terms, this amounts to allowing that an agent could have credences x, y, z, and t in the corresponding conjunctions and β in a proposition that is statistically independent of e, h_1, and $e \wedge h_1$. More generally, it amounts to not restricting the sorts of probability spaces to which $\mathcal{E}$ might apply.

As previously noted, **CA1** tells us that f's numerator has to be zero if and only if $x = (x+y)(x+z)$. Making use of $x+y+z+t = 1$, we conclude that this is the case if and only if:

$$\begin{aligned} x - (x+y)(x+z) = x - x^2 - xy - xz - yz = \\ x(1-x-y-z) - yz = \\ xt - yz = 0 \end{aligned}$$

The obvious way to satisfy **CA1** (i.e., to ensure that $f_n(x,y,z,t) = 0$ iff $xt - yz = 0$) is to set $e = -i$ and all other coefficients (but i) in the numerator to zero. That this is the *only* way to satisfy **CA1** is a straightforward consequence of Hilbert's Nullstellensatz—a fundamental theorem in and to algebraic geometry. In this context, the Nullstellensatz says that, given that the two polynomials $ax^2 + bxy + cy^2 + dxz + eyz + gz^2 + ixt + jyt + rzt + st^2$ and $xt - yz$ have exactly the same zeros, they are constant multiples of each other. Accordingly, f can be reduced to

$$f(x,y,z,t) = \frac{i(xt-yz)}{\bar{a}x^2 + \bar{b}xy + \bar{c}y^2 + \bar{d}xz + \bar{e}yz + \bar{g}z^2 + \bar{i}xt + \bar{j}yt + \bar{r}zt + \bar{s}t^2}$$

Turning now to the denominator, **CA2** requires that $f(x,y,z,t) = 1$ iff $Pr(e|h) = Pr(e \wedge h)/Pr(h) = x/(x+z) = 1$. Thus, if $z = 0$, $f(x,y,z,t) = 1$. Accordingly, for any case in which $y = z = 0$, **CA2** yields $f(x,0,0,t) = 1 = ixt/(\bar{a}x^2 + \bar{i}xt + \bar{s}t^2)$, and by a comparison of coefficients, we get $\bar{a} = \bar{s} = 0$ and $\bar{i} = i$. **CA3** ($\mathcal{E}(e,h) = -\mathcal{E}(\neg e,h)$) is equivalent to

$$f(x,y,z,t) = -f(z,t,x,y). \tag{3}$$

Combining (3) with **CA2**, we have $f(x,0,0,t) = 1 = -f(0,t,x,0) = ixt/(\bar{c}t^2 + \bar{e}xt + \bar{g}x^2)$. Comparing coefficients again, we obtain $\bar{c} = \bar{g} = 0$ and $\bar{e} = i$, reducing f to

$$f(x,y,z,t) = \frac{i(xt-yz)}{\bar{b}xy + \bar{d}xz + i(xt+yz) + \bar{j}yt + \bar{r}zt}$$

Assume now that $\bar{j} \neq 0$. Let $x, z \to 0$. We know by **CA2** that in this case, $f \to 1$. Since the numerator vanishes, the denominator must vanish too, but by $\bar{j} \neq 0$ it stays bounded away from zero, leading to a contradiction ($f \to 0$). Hence $\bar{j} = 0$. In a similar vein, we can argue for $\bar{b} = 0$ by letting $z, t \to 0$ and for $\bar{r} = 0$ by letting $x, y \to 0$—making use of (3) again: $-1 = f(0,0,z,t)$.

Thus, letting $\alpha = \bar{d}/i$, f can be written as

$$\begin{aligned} f(x,y,z,t) &= \frac{i(xt-yz)}{\bar{d}xz + i(xt+yz)} \\ &= \frac{(xt-yz)}{(xt+yz) + \alpha xz}. \end{aligned} \tag{4}$$

To fix the value of α, we make use of **CA4**, which requires $f(x,y,z,t) = f(\beta x, y + (1-\beta)x, \beta z, t + (1-\beta)z)$—see **Lemma 2**, equation (1). Applying this constraint to (4), we obtain

$$\begin{aligned} \frac{xt-yz}{xt+yz+\alpha xz} &= \frac{\beta x(t+z-\beta z) - (y+x-\beta x)\beta z}{\beta x(t+z-\beta z) + (y+x-\beta x)\beta z + \beta^2 xz} \\ &= \frac{xt-yz}{xt+yz+(2-2\beta+\alpha\beta)xz} \end{aligned}$$

For this to be true in general, we have to demand that $\alpha = 2 - 2\beta + \alpha\beta$, which implies that $\alpha = 2$. Hence,

$$f(x, y, z, t) = \frac{xt - yz}{xt + yz + 2xz}.$$

After replacing x, y, z, and t by their corresponding joint probabilities, some algebraic manipulations show that this ratio is equivalent to the following:

$$\mathcal{E}(e, h) = \frac{Pr(h|e) - Pr(h|\neg e)}{Pr(h|e) + Pr(h|\neg e)}$$

which is therefore the unique function satisfying all of the conditions. □

References

Carnap, R. (1950). *Logical Foundations of Probability*. University of Chicago Press, Chicago.

Cohen, M. P. (2015). On Schupbach and Sprenger's measures of explanatory power. *Philosophy of Science*, 82(1):97–109.

Douven, I. (2011). Abduction. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Spring 2011 edition.

Douven, I. and Schupbach, J. N. (2015a). Probabilistic alternatives to Bayesianism: The case of explanationism. *Frontiers in Psychology*, 6(459):1–9.

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

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

Fumerton, R. A. (1980). Induction and reasoning to the best explanation. *Philosophy of Science*, 47:589–600.

Glass, D. H. (2010). Probability and the presumption of atheism. *Yearbook of the Irish Philosophical Society*.

Good, I. J. (1983). *Good Thinking: The Foundations of Probability and Its Applications*. University of Minnesota Press, Minneapolis.

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

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

Kemeny, J. G. and Oppenheim, P. (1952). Degree of factual support. *Philosophy of Science*, 19:307–324.

Lewis, D. (1986). *On the Plurality of Worlds*. Blackwell, Oxford.

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.

Peirce, C. S. (1931-1935). *The Collected Papers of Charles Sanders Peirce*, volume I-VI. Harvard University Press, Cambridge, Mass.

Psillos, S. (1999). *Scientific Realism: How Science Tracks Truth*. Routledge, London.

Putnam, H. (1975). *Mathematics, Matter, and Method*, volume I of *Philosophical Papers*. Cambridge University Press, Cambridge.

Schupbach, J. N. (2011). Comparing probabilistic measures of explanatory power. *Philosophy of Science*, 78(5):813–829.

Schupbach, J. N. (2014). Is the bad lot objection just misguided? *Erkenntnis*, 79(1):55–64.

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

Swinburne, R. (2004). *The Existence of God*. Oxford University Press, Oxford, 2nd edition.

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

Vogel, J. (1990). Cartesian skepticism and Inference to the Best Explanation. *The Journal of Philosophy*, 87(11):658–666.