

Inference to the Best Explanation and the Screening-Off Challenge

William Roche and Elliott Sober

RESUMEN

Defendemos en Roche y Sober (2013) que la explicatividad es evidencialmente irrelevante, esto es, que $\Pr(H \mid O \& EXPL) = \Pr(H \mid O)$, donde H es una hipótesis, O una observación y $EXPL$ es la proposición de que si H y O fueran verdaderas, entonces H explicaría O . Esta es una “tesis de neutralización” [“screening off” thesis, de ahí el nombre “SOT”]. En el presente artículo clarificamos esta tesis, replicamos a las críticas presentadas por Lange (2017), consideramos algunas formulaciones alternativas de la “Inferencia a la mejor explicación”, defendemos dos versiones más fuertes de la tesis, que denominamos “SOT*” y “SOT**”, y consideramos cómo estas inciden en la afirmación de que la virtud teórica de la unificación es evidencialmente relevante.

PALABRAS CLAVE: *bayesianismo; relevancia evidencial; explicatividad; inferencia a la mejor explicación; Lange; neutralización; unificación.*

ABSTRACT

We argue in Roche and Sober (2013) that explanatoriness is evidentially irrelevant in that $\Pr(H \mid O \& EXPL) = \Pr(H \mid O)$, where H is a hypothesis, O is an observation, and $EXPL$ is the proposition that if H and O were true, then H would explain O . This is a “screening-off” thesis (hence the name “SOT”). Here we clarify SOT, reply to criticisms advanced by Lange (2017), consider alternative formulations of Inference to the Best Explanation, defend two strengthened screening-off theses called “SOT*” and “SOT**”, and consider how they bear on the claim that unification is evidentially relevant.

KEYWORDS: *Bayesianism; Evidential Relevance; Explanatoriness; Inference to the Best Explanation; Lange; Screening-off; Unification.*

I. INTRODUCTION

We argue in Roche and Sober (2013) that explanatoriness is evidentially irrelevant in that $\Pr(H \mid O \& EXPL) = \Pr(H \mid O)$, where, here and throughout, H is a hypothesis, O is an observation, and $EXPL$ is the proposition that if H and O were true, then H would explain O . This is a

“screening-off” thesis (hence the name “SOT”) to the effect that O screens-off $EXPL$ from H in that given O , $EXPL$ has no impact on the probability of H . Suppose, for example, that you examine a large random sample of people older than age 50, and that on this basis, you arrive at the following population frequency estimate:

$$(1) \text{Freq}(\text{heavy smoking before age 50} \mid \text{lung cancer after age 50}) = \alpha$$

Suppose that from your perspective, Joe is a random member of the population (and was not in your sample). Let H be the hypothesis that Joe was a heavy smoker before age 50, O be the observation that Joe got lung cancer after age 50, and $EXPL$ be the proposition that if H and O were true, then H would explain O . From your perspective, Joe is a random member of the population, so $\Pr(H \mid O) = \alpha$. Given this, and given that $EXPL$ should have no impact on your estimate of the frequency in (1), $\Pr(H \mid O \& EXPL) = \alpha$. Hence $\Pr(H \mid O \& EXPL) = \Pr(H \mid O)$.

SOT and our defense of it have not been met with universal approval, to say the least. McCain and Poston (2014) were the first to chime in with objections. Climenhaga (2017) was next, and then Lange (2017) took his turn. We have responded to McCain and Poston [see Roche and Sober (2014)] and to Climenhaga [see Roche and Sober (2017b)].¹ Here we respond to Lange, but that’s not all. There are numerous non-equivalent versions of Inference to the Best Explanation (IBE) in logical space. We will argue that SOT refutes some of them, but not others. We then will defend two variants of SOT called “SOT*” and “SOT**” and argue that they refute many of the remaining versions, specifically, versions of IBE that say that whether H is the best available potential explanation of O hinges on how H scores in terms of unification.

II. RESPONSE TO LANGE

Lange holds that there are many realistic cases where $EXPL$ isn’t screened-off from H by O , because $\Pr(H \mid O \& EXPL) > \Pr(H \mid O)$.² He gives two examples that he claims are of this sort. We discuss one of them in Section 2.1 and address the other in Section 2.2. We argue in each case that Lange fails to show, or even to make it plausible, that $\Pr(H \mid O \& EXPL) > \Pr(H \mid O)$. In Section 2.3, we clarify SOT, and provide a more general response to objections like Lange’s.

II.1 Lange's Robbery Example

Lange's first alleged example of a realistic case where $\Pr(H \mid O \& EXPL) > \Pr(H \mid O)$ involves a robbery in which a jewel is stolen from a safe. After introducing the example, Lange argues that $\Pr(H \mid O)$ is greater than $\Pr(H)$ but less than maximal (i.e., less than unity):

... suppose that H is that Jones is the person who stole the jewel from the safe, O is that the single strand of hair found inside the safe was blond, and the background information tells us that there was exactly one robber and one strand of hair found inside the safe, that Jones has blond hair, and that such a hair has a serious (though not overwhelming) likelihood to have been left by the robber during the robbery (though there are other ways in which the hair could have gotten into the safe). The background information also tells us that Jones is a serious suspect, unlike many other people with blond hair – although Jones is one among several serious suspects with blond hair and there is also a fair likelihood that the robber is not listed among our serious suspects. Background also tells us that if the hair were Jones's, then Jones would probably be the robber (since he would have left it during the robbery); Jones would have had no occasion to access the safe except to rob it.

Accordingly, since the hair that was found is the same colour as Jones's hair, O lends some support to H – $\Pr(H \mid O) > \Pr(H)$ – though this support is less than maximal, considering that the hair may not have come from the robber and that, even if it did come from the robber, the robber need not be Jones since many other people (including some other serious suspects) have blond hair [Lange (2017), p. 305].

Lange then adds $EXPL$ to the mix,³ and argues that $\Pr(H \mid O \& EXPL) > \Pr(H \mid O)$:

But to O in the evidence let's now add $EXPL$: that if Jones were the robber and the single strand of hair found inside the safe were blond, then that Jones is the robber would explain why the strand of hair found in the safe is blond. The explanation would be that Jones left the hair in the course of the robbery. Without $EXPL$, the evidence's power to confirm H is rendered less than maximal by (among other things) serious doubt that the hair comes from the robber. But that doubt is removed by $EXPL$ (at least in the event that Jones is the robber). Of course, the evidence's power to confirm H remains somewhat less than maximal because of other factors, such as doubt about whether the hair comes from Jones. But $EXPL$ removes one consideration that mitigated the degree to which O pointed to Jones (namely, the possibility that even if Jones were the rob-

ber, such a hair would not belong to Jones because it was not left by the robber). Consequently, H is better confirmed by $O \& EXPL$ than by O alone: $\Pr(H \mid O \& EXPL) > \Pr(H \mid O)$ [Lange (2017), pp. 305-306].

Lange's rationale, then, is two-fold. First, there's the claim that because of various doubts, $\Pr(H \mid O)$ is less than maximal. Second, there's the claim that $EXPL$ eliminates some of the doubts in question, and as a result increases the probability of H .

What doubts make $\Pr(H \mid O)$ less than maximal? Lange initially mentions these:

- (2) The strand of hair found in the safe wasn't left by the robber.
- (3) The strand of hair found in the safe was left by the robber, but the robber wasn't Jones and instead was one of the other serious suspects with blonde hair.

He later points to this:

- (4) Jones is the robber, but the strand of hair found in the safe doesn't belong to him because it wasn't left by the robber.

Lange claims that $EXPL$ eliminates (4), and because of this, $EXPL$ increases the probability of H . This is strange, since (4) is a possibility in which H is true. How is it that by eliminating a possibility in which H is true, $EXPL$ increases H 's probability? Instead, why not think that by eliminating a possibility in which H is true, $EXPL$ decreases H 's probability?

What Lange is describing can happen, but we doubt that it is true in his robbery example. To illustrate how an observation can raise the probability of a hypothesis and also eliminate a possibility in which the hypothesis is true, consider this example. A card is randomly drawn from a standard well-shuffled deck of cards. Let A , D , and S be understood as follows:

- A : The card drawn is an Ace.
- D : The card drawn is a Diamond.
- S : The card drawn is a Spade.

Given that $\sim(A \& D) \& \sim S$ entails $\sim(A \& D)$, it follows that $\sim(A \& D) \& \sim S$ eliminates $A \& D$ and thus eliminates a possibility in which D is true. Yet $\sim(A \& D) \& \sim S$ nonetheless increases D 's probability from $1/4$ to $12/38$.

The thing to notice here is this: although $\sim(A\&D)\&\sim S$ eliminates a possibility in which D is true, it also eliminates several possibilities in which D is *false* (for example, $A\&S$).

Is something similar true in Lange's robbery case? That is, is it the case that, though *EXPL* eliminates (4), it also eliminates various possibilities in which H is *false*? There's no doubt that *EXPL* eliminates *some* possibilities in which H is false. It eliminates, for example, $\sim H\&\sim EXPL$. But this, by itself, is insignificant, since, at the same time, and for the same reason, it also eliminates $H\&\sim EXPL$.

Consider, instead, these possibilities:

- (5) Smith, not Jones, is the robber. The strand of hair found in the safe doesn't belong to Smith, and was instead planted by him as a distraction to the police.
- (6) Smith, not Jones, is the robber. The strand of hair found in the safe was left by the owner of the jewel a few days before the robbery while she was taking something other than the jewel out of the safe.

Clearly, assuming, with Lange, that the background information on hand is realistic, *EXPL* doesn't eliminate (5) or (6).

It might be that Lange can flesh out this example of his so that it is clear that $\Pr(H \mid O\&EXPL) > \Pr(H \mid O)$. We return to this possibility in Section II.3.

II.2. Lange's Physics Example

Lange's second attempt to provide a realistic example in which $\Pr(H \mid O\&EXPL) > \Pr(H \mid O)$ is taken from physics, where H is the hypothesis that the light quantum hypothesis is empirically adequate, and O is an equation specifying the black-body spectrum.⁴ First, he argues that if H is true but the light quantum hypothesis is false, then it's a mere coincidence that various phenomena behave as if the light quantum hypothesis is true:

If H holds but light is not quantized, then it is just a coincidence that various phenomena behave as if light is quantized: the fundamental natural laws of light (which do not include that light is quantized) entail the particular derivative laws (such as O) that govern various phenomena, and the light-quantum hypothesis also entails those laws, but there is no common reason why these two facts hold: this combination is 'nothing more than a

curious property of light, without any physical significance' [Lange (2017), p. 309].

Second, he considers the possibility that H is true because the light quantum hypothesis is true too:

On the other hand, if H holds because light is indeed quantized, then it is no coincidence that various phenomena behave as if light is quantized. Rather, there is a common reason why the fundamental laws of light and the light-quantum hypothesis are alike in sharing the property of entailing the derivative laws. The common reason is that the light-quantum hypothesis is one of those fundamental laws. If H holds, then if there are light-quanta, H can explain O ; the explanation is roughly that since there are light-quanta, everything behaves as if there are (i.e., H holds), including the black-body spectrum, and O is what the black-body spectrum would be if H . However, if H holds, then H cannot explain O if there are no light-quanta, since then H is just a fluke [Lange (2017), p. 6].

Third, he argues that $\Pr(H \mid O \& EXPL) > \Pr(H \mid O \& \sim EXPL)$:

... compare $\Pr(H \mid O \& EXPL)$ to $\Pr(H \mid O \& \sim EXPL)$. Both $O \& EXPL$ and $O \& \sim EXPL$ confirm H to some degree. But $O \& \sim EXPL$ confirms H only by removing one way in which H could have gone wrong—that is, only by confirming one part of H (that the light-quantum hypothesis entails the correct equation for the black-body spectrum) and having no bearing on the rest of H (that the light-quantum hypothesis entails the correct equations for the photoelectric effect, for the Volta effect, ...). In contrast, $O \& EXPL$ confirms H not just by confirming that the light-quantum hypothesis gets the black-body spectrum right, but also by confirming that the light-quantum hypothesis gets various other phenomena right. Therefore, $\Pr(H \mid O \& EXPL) > \Pr(H \mid O \& \sim EXPL)$ [Lange (2017), pp. 309-310].

Fourth, and finally, he concludes that $\Pr(H \mid O \& EXPL) > \Pr(H \mid O)$.⁵ Lange (2017), p. 311, says that inequalities of the following form are central to his argument:

(7) $\Pr(\text{the ... phenomenon behaves as if there were light-quanta} \mid O \& EXPL)$
 $>> \Pr(\text{the ... phenomenon behaves as if there were light-quanta} \mid O \& \sim EXPL)$.

Consider, for example, the following, where O^* is the proposition that the photoelectric effect behaves as if there were light-quanta:

$$(8) \Pr(O^* \mid O \& EXPL) \gg \Pr(O^* \mid O \& \sim EXPL)$$

It's not immediately obvious, however, how inequalities like (8) figure in his overall argument. The problem is that (8) makes no mention of H , whereas the claim that $\Pr(H \mid O \& EXPL) > \Pr(H \mid O)$ does.

It might be that Lange is tacitly assuming Hempel's (1965) Converse Consequence Condition (here understood in terms of confirmation in the sense of increase in probability):

CCC: For any propositions X , Y , Z , and Z^* , if (i) $\Pr(Z \mid Y \& X) > \Pr(Z \mid Y)$ and (ii) Z^* entails Z , then $\Pr(Z^* \mid Y \& X) > \Pr(Z^* \mid Y)$.

It follows from (8) that

$$(9) \Pr(O^* \mid O \& EXPL) > \Pr(O^* \mid O \& \sim EXPL),$$

and this entails that:

$$(10) \Pr(O^* \mid O \& EXPL) > \Pr(O^* \mid O).$$

Given (10), and given, suppose, that H entails O^* , it follows by CCC that $\Pr(H \mid O \& EXPL) > \Pr(H \mid O)$. However, if this is how the argument is supposed to work, then the argument fails. As is well known, CCC has counterexamples.⁶

We now set this problem aside and suppose, for the sake of argument, that there's a legitimate way to get from inequalities such as (8) to the claim that $\Pr(H \mid O \& EXPL) > \Pr(H \mid O)$.⁷ Why should inequalities like (8) be accepted?

We don't understand Lange's answer here, but we have a conjecture based in part on what he says about a slight variant of an example described by a former time slice of one of the authors of the paper you're now reading [Sober (2015)]. Suppose that two of the students in a seminar you are teaching turn in word-for-word identical essays. You consider two hypotheses. CC (short for "Common Cause") says that they searched the Internet together, found a paper suited to the assignment, and agreed to plagiarize it. SC (short for "Separate Causes") says that they worked separately and independently. Sober (2015), pp. 103-104, formulates the following thesis, and says that it is a "first pass" in need of refinement:

$$(11) \Pr(\text{the papers match} \mid CC) \gg \Pr(\text{the papers match} \mid SC)$$

Lange modifies the example slightly by letting “w” be a certain long sequence of words, and replacing (11) with the following:

- (12) $\Pr(\text{Smith's paper contains } w \ \& \ \text{Jones's paper contains } w \mid CC)$
 $\gg \Pr(\text{Smith's paper contains } w \ \& \ \text{Jones's paper contains } w \mid SC)$

Lange then claims that Sober would agree to (12) because he would also agree that:

- (13) $\Pr(\text{Smith's paper contains } w \mid \text{Jones's paper contains } w \ \& \ CC)$
 $\gg \Pr(\text{Smith's paper contains } w \mid \text{Jones's paper contains } w \ \& \ SC)$

Here is Lange’s explanation of why Sober would endorse (12):

For Smith’s paper to contain w, given that Jones’s paper contains w but *SC*, would be extremely unlikely. ‘According to *SC*, the matching is a coincidence; according to *CC*, it is anything but’ (Sober 2015: 103) [Lange (2017) p. 310].

The idea here (and elsewhere in Lange’s discussion) seems to be that Sober would accept (12) because he would accept:

- (14) Given *SC*, it would be a coincidence if both Smith’s paper and Jones’s paper were to contain w, and so given *SC* and that Jones’s paper contains w, it is highly unlikely that Smith’s paper also contains w, whereas given *CC*, it would not be a coincidence if both Smith’s paper and Jones’s paper were to contain w, and so, given *CC* and that Jones’s paper contains w, it is not highly unlikely that Smith’s paper also contains w.

Note the transitions from “would be a coincidence” to “is highly unlikely”, and from “would not be a coincidence” to “is not highly unlikely”.⁸

Lange seems to agree with Sober (as he reads him) on all of this, and further seems to think that analogous points hold in his physics case. This suggests that Lange accepts (8), for example, because he accepts:

- (15) Given $\sim EXPL$, it would be a coincidence if both *O* and *O** were true, and so given $\sim EXPL$ and *O*, it is highly unlikely that *O** is true, whereas given *EXPL*, it would not be a coincidence if both *O* and *O** were true, and so, given *EXPL* and *O*, it is not highly unlikely that *O** is true.

This, like (14), has transitions from “would be a coincidence” to “is highly unlikely”, and from “would not be a coincidence” to “is not highly unlikely”.

However, Sober (2015) *denies* that a common cause explanation of a “matching” between two events always has a higher likelihood than a separate cause explanation of that matching. Sometimes the matching of the two events favors a common cause explanation that says that the matching is not a coincidence, but sometimes it does not. Everything depends on the background assumptions that pertain. In the example of the student essays, it’s easy to see how (12) could be false. Suppose that if the students work together and plagiarize, they will be loathed to include word sequence w , but if they work separately and independently, the chances of them including that sentence in their essays is much greater. Notice that our point here does not depend on this supposition’s being realistic.⁹

If, as it seems, Lange’s view is that (8) should be accepted because (15) should be accepted, then his argument is in trouble. For, (15), like (14), should be rejected.

We noted at the end of Section II.1 that it might be that Lange can flesh out his robbery example so that it’s clear that $\Pr(H \mid O \& EXPL) > \Pr(H \mid O)$. The same is true with respect to his physics example.

II.3. SOT’s Scope

It might seem that SOT *universally* quantifies over all logically possible cases:

TOO STRONG: For any logically possible case in which $\Pr(H \mid O \& EXPL)$ and $\Pr(H \mid O)$ are well-defined, O screens-off $EXPL$ from H in that $\Pr(H \mid O \& EXPL) = \Pr(H \mid O)$.

Alternatively, it might seem that SOT is much more modest, in that it *existentially* quantifies over all logically possible cases:

TOO WEAK: There are logically possible cases in which O screens-off $EXPL$ from H in that $\Pr(H \mid O \& EXPL) = \Pr(H \mid O)$.

In fact, neither reading is right. TOO STRONG is obviously false. Suppose, for instance, that $\Pr(H \mid O)$ is less than unity, and that the background information codified in $\Pr(-)$ includes the assumption that:

$$(16) (EXPL \& H) \vee (\sim EXPL \& \sim H)$$

It follows that $\Pr(H \mid O \& EXPL) = 1 > \Pr(H \mid O)$. TOO WEAK, in turn, is obviously true but utterly boring. If, for example, $\Pr(H \mid O)$ equals unity, and $\Pr(H \mid O \& EXPL)$ is well-defined, then, trivially, $\Pr(H \mid O \& EXPL) = 1 = \Pr(H \mid O)$.

How, then, should SOT be understood? Inspired by Goldilocks, we understand it to be saying this:

JUST RIGHT: There are many realistic cases in which the background information codified in $\Pr(-)$ includes frequency data such that O screens-off $EXPL$ from H in that $\Pr(H \mid O \& EXPL) = \Pr(H \mid O)$.

Our smoking example is a case of this sort, but there are many – very many – other examples of the same kind.

We acknowledge, though, that there are some potentially misleading passages in Roche and Sober (2013). Here is one:

Our screening-off thesis is related to Van Fraassen's (1989) thesis that inference to the best explanation (IBE) is probabilistically incoherent, and therefore subject to a Dutch book. Van Fraassen thinks that IBE proposes a two-step rule for updating: if the evidence O increases H 's probability, then H receives a further boost in probability if H would provide a good explanation of O . Our argument aims to show that the explanatoriness of H *cannot* provide this additional boost; in addition, it side-steps the question of how the apparently prudential considerations introduced by Dutch book arguments are relevant to a non-prudential notion of rational degree of belief [Roche and Sober (2013), p. 665, emphasis added]

Our use of the word “cannot” may suggest that we meant TOO STRONG as opposed to JUST RIGHT, but that wasn't our intent. We meant *cannot* (*in cases like our smoking case where there's abundant frequency data on hand*). We regret not making this completely clear.

Given that SOT should be understood as JUST RIGHT, it follows that even if Lange fleshed out his robbery example or his physics example so that it's clear that $\Pr(H \mid O \& EXPL) > \Pr(H \mid O)$, SOT would remain unscathed. SOT allows for realistic cases in which $\Pr(H \mid O \& EXPL) > \Pr(H \mid O)$. It even allows for realistic cases in which the background information codified in $\Pr(-)$ includes frequency data such that $\Pr(H \mid O \& EXPL) > \Pr(H \mid O)$.¹⁰

Is SOT, understood as JUST RIGHT, trivial in the way that TOO WEAK is trivial? We think not, and now will explain why.

III. SOT AND IBE*

IBE can be formulated in different ways. Here is one:

IBE*: If (i) O , (ii) H is a potential explanation of O , and (iii) H is better overall in terms of explanatory virtues v_1, v_2, \dots , and v_n than each of the available rival potential explanations of O , then it's rational to believe H and disbelieve each of the rivals.

This is a relatively standard formulation, but there are others. We discuss other formulations in Section V.

Where does probability come into play in IBE*? Let *BEST* be the proposition that H is the best overall available rival explanation of O in terms of virtues v_1, v_2, \dots , and v_n . We assume that IBE* entails that there are no cases in its scope where O , *EXPL*, and *BEST* are true but:

$$(17) \Pr(H \mid O \& EXPL \& BEST) \leq 0.5$$

For, presumably, if $\Pr(H \mid O \& EXPL \& BEST) \leq 0.5$, then it isn't rational to believe H and disbelieve each of the rivals.¹¹

We further assume that IBE* entails that at least some cases in its scope where O , *EXPL*, and *BEST* are true are such that:

$$(18) \Pr(H \mid O \& EXPL \& BEST) > t > \Pr(H \mid O)$$

Here " t " is the threshold for high probability, and should be understood so that it is less than 1 and greater than or equal to 0.5.¹² If *EXPL* & *BEST* never increases H 's probability (given O) from low (not high) to high, then why build a theory of inference around *EXPL* & *BEST*?¹³

Now consider:

$$(19) \Pr(H \mid O \& EXPL) > \Pr(H \mid O)$$

Should IBE* be understood so that every case in its scope where O and *EXPL* are true is a case where (19) holds? SOT is relevant here. The claim that H is a potential explanation of O is in effect *EXPL*. This, at any rate, is how IBE-ists typically construe the notion of a potential explanation. Consider, for instance, the following:

‘Explains’ in the second premise [i.e., the premise, in our notation, that H explains O] cannot, without begging the question, mean ‘actually explains’; rather, it is used in the sense of ‘would explain if true’ [Lycan (2002), p. 413].

A *potential* explanation of the evidence is anything that *would* explain the evidence *if it were true* [Williamson (2016), p. 266, emphasis original].

Lycan doesn’t use the expression “potentially explains”, but that’s the intended contrast with “actually explains”.¹⁴ If IBE* should be understood so that every case in its scope where O and $EXPL$ are true is a case where (19) holds, and if IBE*’s scope includes all realistic cases, then SOT entails that IBE* is false.¹⁵

We aren’t insisting, though, that IBE* should be understood in that manner. We are simply addressing one potential way of understanding it.

IV. SOT* AND IBE*

The present task is to consider the possibility that IBE* does *not* require that every case in its scope where O and $EXPL$ are true is a case where (19) holds. The first point to note is that SOT has a cousin:

SOT*: There are many realistic cases in which the background information codified in $\text{Pr}(-)$ includes frequency data such that O screens-off $BEST$ from H in that $\text{Pr}(H \mid O \& BEST) = \text{Pr}(H \mid O)$.

There’s no explicit mention here of $EXPL$. But it’s there implicitly, since $BEST$ is logically equivalent to $EXPL \& BEST$, which means that $\text{Pr}(H \mid O \& BEST) = \text{Pr}(H \mid O \& EXPL \& BEST)$. Is SOT* true, and if so, does it undermine IBE*?

It might seem that SOT* follows from SOT. For, $BEST$ is logically stronger than $EXPL$, and it might seem that for any propositions X , X^* , Y , and Z , if (i) $\text{Pr}(Z \mid Y \& X) = \text{Pr}(Z \mid Y)$ and (ii) X^* is logically stronger than X , then $\text{Pr}(Z \mid Y \& X^*) = \text{Pr}(Z \mid Y)$. But, as with CCC, there are exceptions, for example, where $1 > \text{Pr}(Z \mid Y \& X) = \text{Pr}(Z \mid Y)$, and X^* is the conjunction of X and Z .

There are many different explanatory virtues noted in the extant literature on IBE.¹⁶ Some commonly cited examples are:

- (a) empirical adequacy
- (b) explanatory power
- (c) fit with background data

- (d) fertility
- (e) internal consistency
- (f) internal coherence
- (g) mechanism
- (h) parsimony
- (i) precision
- (j) scope
- (k) unification

Different sets of explanatory virtues lead to different versions of IBE*.¹⁷ We want to focus on versions of IBE* on which unification is included in v_1, v_2, \dots , and v_n . We assume for definiteness that a common cause explanation is always superior in unification to a separate cause explanation. This way of understanding unification has some *prima facie* appeal, and has been explicitly endorsed in the extant literature on unification.¹⁸

Suppose, adapting a case introduced in Lange (2004) and later modified in Blanchard (2018), that your background information includes the following frequency data:

$$(20) \text{Freq}(\text{pleuritis \& malar rash} \mid \text{lupus}) = 0.891 > 0.0495 = \text{Freq}(\text{pleuritis \& malar rash})$$

$$(21) \text{Freq}(\text{lupus}) = 0.005$$

$$(22) \text{Freq}(\text{lupus} \mid \text{pleuritis \& malar rash}) = 0.09$$

Let L , M , and P be understood as follows:

L : Jones has lupus.

M : Jones has a malar rash.

P : Jones has pleuritis.

Given (20), (21), and (22), and given that, suppose, Jones is a random member of the population from your perspective, you should have the following probabilities:

$$(23) \text{Pr}(P\&M \mid L) = 0.891 > 0.0495 = \text{Pr}(P\&M)$$

$$(24) \text{Pr}(L) = 0.005$$

$$(25) \Pr(L \mid P\&M) = 0.09$$

You then learn that $P\&M$, and as a result rightly increase your credence in L from 0.005 to 0.09.

Now let B and F be understood as follows:

B : Jones has Bloom's disease.

F : Jones has the flu.

Suppose that you subsequently come to learn that pleuritis is caused both by lupus and by the flu, and that malar rashes are caused both by lupus and by Bloom's disease. You then have two potential explanations of $P\&M$. One is L , which is a common cause explanation. The other is $F\&B$, which is a separate cause explanation. Which is better in terms of v_1, v_2, \dots , and v_n ?

Given the assumption that a common cause explanation is always more unifying than a separate cause explanation, it follows that L is superior in unification to $F\&B$. Suppose that L is equal or superior to $F\&B$ in terms of each of the remaining explanatory virtues included in v_1, v_2, \dots , and v_n , and that you learn this and thus further learn that L is better overall in terms of v_1, v_2, \dots , and v_n than $F\&B$.¹⁹ Should you increase your credence in L from 0.09 to some higher value? It seems clear that the answer is negative; your credence in L should remain at 0.09.

What if you learned not just that L surpasses $F\&B$ in terms of v_1, v_2, \dots , and v_n , but also that since there are no additional available potential explanations of $P\&M$, L surpasses each of its rival available potential explanations in terms of v_1, v_2, \dots , and v_n ? What if, in other words, you learned *BEST*? The answer, it seems, is the same: *BEST* is screened-off in that your credence in L should remain at 0.09.

This verdict can be bolstered by adding some further details to the example. Suppose that your frequency data goes beyond (20), (21), and (22). Suppose in particular that it includes:

$$(26) \text{Freq}(\text{pleuritis \& malar rash} \mid \text{flu \& Bloom's disease}) = 0.891$$

$$(27) \text{Freq}(\text{flu \& Bloom's disease}) = 0.005$$

$$(28) \text{Freq}(\text{flu \& Bloom's disease} \mid \text{pleuritis \& malar rash}) = 0.09$$

It follows that your frequency data is neutral between L and $F\&B$. For, you know that Jones has pleuritis and a malar rash (and know nothing else relevant about his symptoms), and you know that the frequency of

lupus among people who have pleuritis and a malar rash is equal to the frequency of the flu and Bloom's disease among such people. But then you, like your frequency data, should be neutral between L and $F\&B$.

Things could have turned out differently. Your background information could have included things in addition to the frequency data given in (20), (21), (22), (26), (27), and (28), and this extra information could have had the result that your credence in L should be greater than your credence in $F\&B$. You could have known, for instance, that Jones's partner has lupus, and that lupus is easily passed from person to person. We were supposing, though, that initially Jones was a random member of the population from your perspective, so you didn't have any such extra information.

There is nothing special about our lupus example. It's typical of many realistic cases in which the background information codified in $\text{Pr}(-)$ includes frequency data such that, although $BEST$ is true (because H is superior in unification to each of the available rival potential explanations of O), $BEST$ is screened-off from H by O in that $\text{Pr}(H \mid O\&BEST) = \text{Pr}(H \mid O)$. Hence SOT^* is true.

How does all this bear on IBE^* (or, more specifically, on the versions of IBE^* under consideration)? If we are right about our lupus case, and if IBE^* 's scope includes all realistic cases, then at least some of the cases in virtue of which SOT^* is true are cases where (17) is true and (18) is false because:

$$(29) t > \text{Pr}(H \mid O\&EXPL\&BEST) = 0.09 = \text{Pr}(H \mid O)$$

But then IBE^* is false.

Denying that $BEST$ is screened-off in our lupus case doesn't save IBE^* . Even if $\text{Pr}(H \mid O\&EXPL\&BEST)$ were greater than 0.09, surely it wouldn't be greater than 0.5, and thus surely it wouldn't be greater than t . Hence (17) would still be true.

We noted above that L is superior in unification to $F\&B$, and then simply supposed, without argument, that L is equal or superior to $F\&B$ in terms of each of the remaining explanatory virtues included in v_1 , v_2 , ..., and v_n . This seemed legitimate then, and still seems legitimate now. Nothing in the case as specified to this point requires that L be *inferior* to $F\&B$ in terms of any of empirical adequacy, explanatory power, fit with background data, or any of the other explanatory virtues noted above (or any additional ones for that matter).

V. SOT** and IBE**

IBE* is a relatively standard formulation of IBE, but there are alternatives. Here is one:

IBE**: If (i) O , (ii) H is a potential explanation of O , (iii) H is better overall (in terms of explanatory virtues v_1, v_2, \dots , and v_n) than each of the available rival potential explanations of O , and (iv) H 's overall score in terms of v_1, v_2, \dots , and v_n is high, then it's rational to believe H and disbelieve each of the rivals.²⁰

Let *HIGH* be the proposition that H 's overall score in terms of v_1, v_2, \dots , and v_n is high. If *HIGH* holds in cases like our lupus case, then:

SOT**: There are many realistic cases in which the background information codified in $\text{Pr}(-)$ includes frequency data such that O screens-off *BEST&HIGH* from H in that $\text{Pr}(H \mid O \& \text{BEST} \& \text{HIGH}) = \text{Pr}(H \mid O)$.

It might be argued, however, that *HIGH* is false in cases like our lupus case. What then?

There would still be problems. The idea behind IBE** is that though *O&EXPL&BEST* (IBE*'s antecedent) leads to the *fully comparative* claim that H 's probability is *greater than* the probabilities of the other potential explanations in question, it doesn't lead to the *partially non-comparative* claim that H 's probability is *high*. *HIGH* is supposed to connect the two.²¹ Our lupus example, though, shows that sometimes *O&EXPL&BEST* does *not* lead to the fully comparative claim that H 's probability is greater than the probabilities of the other potential explanations in question. This undermines the idea behind IBE**.

Let's set aside this worry and turn to the question of whether there's a legitimate way to motivate the claim that *HIGH* is false in our lupus example. Recall that given the frequency data on hand, $\text{Pr}(P \& M \mid L) = 0.891$, and $\text{Pr}(L) = 0.005$. It might be argued that L 's explanatory power with respect to $P \& M$ is given by $\text{Pr}(P \& M \mid L)$, that L 's fit with background data is given by $\text{Pr}(L)$, and that neither of these is high enough for *HIGH* to be true. What now?

It turns out that there are variants of our example in which each of the following holds:

$$(30) \text{Freq}(\text{pleuritis \& malar rash} \mid \text{lupus}) = 1 = \text{Freq}(\text{pleuritis \& malar rash} \mid \text{flu \& Bloom's disease})$$

$$(31) \text{Freq}(\text{lupus}) = 0.4 = \text{Freq}(\text{flu \& Bloom's disease})$$

$$(32) \text{Freq}(\text{lupus} \mid \text{pleuritis \& malar rash}) = \text{Freq}(\text{flu \& Bloom's disease} \mid \text{pleuritis \& malar rash}) \approx 0.493$$

Now $\text{Pr}(P\&M \mid L) = 1$ (up from 0.891) and $\text{Pr}(L) = 0.4$ (up from 0.005), so that L 's ability to predict $P\&M$ (the observation to be explained), as given by $\text{Pr}(P\&M \mid L)$, is maximal, and though L 's fit with the background information on hand, as given by $\text{Pr}(L)$, isn't maximal, it is nonetheless relatively high and much greater than 0.005. If this means that *HIGH* is true, then since your frequency data is still neutral between L and $F\&B$, and since there are many additional examples like this variant of our lupus example, it follows that IBE^{**} is false and SOT^{**} is true.

Friends of IBE^{**} could restrict IBE^{**} to realistic cases where the background information on hand includes no relevant frequency data. Then our lupus examples would pose no threat, but friends of IBE^{**} would be obliged to furnish an independent rationale for this restriction. If the sole reason for this restriction is that otherwise the theory would be open to counterexample, the restriction would be *ad hoc*. Furthermore, friends of IBE^{**} should be worried about more than just frequency-data cases. For, arguably, things other than frequency data, for example, background theories, can enable O to screen-off *BEST\&HIGH* from H .²²

What if friends of IBE^{**} simply excised unification from the set of virtues (or else simply denied the assumption that a common cause explanation is always superior in unification to a separate cause explanation)? Would they then be in the clear? Not necessarily. Our argument carries over to any version of IBE^{**} on which the set of virtues includes a virtue ν such that there can be realistic cases where H is superior in ν to each of its rival available potential explanations, and yet substantial frequency data at hand is neutral between H and at least one of the rivals.²³

VI. CONCLUSION

We have defended three screening-off theses, which we'll now repeat for convenience:

- SOT: There are many realistic cases in which the background information codified in $\text{Pr}(-)$ includes frequency data such that O screens-off $EXPL$ from H in that $\text{Pr}(H \mid O \& EXPL) = \text{Pr}(H \mid O)$.
- SOT*: There are many realistic cases in which the background information codified in $\text{Pr}(-)$ includes frequency data such that O screens-off $BEST$ from H in that $\text{Pr}(H \mid O \& BEST) = \text{Pr}(H \mid O)$.
- SOT**: There are many realistic cases in which the background information codified in $\text{Pr}(-)$ includes frequency data such that O screens-off $BEST \& HIGH$ from H in that $\text{Pr}(H \mid O \& BEST \& HIGH) = \text{Pr}(H \mid O)$.

These theses are all existential, but unlike TOO WEAK, they are far from trivial. First, if IBE^* is understood so that every case in its scope where O and $EXPL$ are true is a case where $\text{Pr}(H \mid O \& EXPL) > \text{Pr}(H \mid O)$, and if IBE^* 's scope includes all realistic cases, then SOT refutes IBE^* . Second, if IBE^* and IBE^{**} are understood so that unification is an explanatory virtue such that a common cause explanation is always more unifying than a separate cause explanation, and if IBE^* 's and IBE^{**} 's scopes include all realistic cases, then SOT* refutes IBE^* , and SOT** refutes IBE^{**} .

*Department of Philosophy
Texas Christian University
Fort Worth, TX 76129, USA
E-mail: w.roche@tcu.edu*

*Department of Philosophy
University of Wisconsin, Madison
Madison, WI, 53706 USA
E-mail: ersober@wisc.edu*

NOTES

¹ McCain and Poston (2017) have responded in kind to our response to their earlier objections. We don't have the space here to respond back.

² All references in this section to Lange are to Lange (2017).

³ Strictly speaking, Lange uses " E " as opposed to " $EXPL$ ". We have changed his notation in the quoted passages below, so that it conforms to our notation.

⁴ Lange construes the light quantum hypothesis as the hypothesis "that light comes in discrete quantities rather than continuous waves" [Lange (2017), p. 308]. But since he claims that the light quantum hypothesis entails O , which is an equation, we take it that he means for the light quantum hypothesis to be something more than just the hypothesis that light comes in discrete quantities ra-

ther than continuous waves. We shall assume for the sake of argument that the light quantum hypothesis when *fully specified* has the entailments claimed by Lange.

⁵ It's a theorem of the probability calculus that for any propositions X , Y , and Z , $\Pr(Z \mid Y \& X) > \Pr(Z \mid Y)$ precisely when $\Pr(Z \mid Y \& X) > \Pr(Z \mid Y \& \sim X)$.

⁶ See Roche (2017) for discussion of whether CCC and related theses can be repaired by modifying them in terms of explanation.

⁷ We also want to set aside a further problem. In the second displayed passage in this subsection, Lange seems to assume that flukes are explanatorily impotent. However, that assumption is dubious. If Smith and Jones run into each other on State Street at noon on Tuesday by coincidence, then their running into each other is a fluke. But their running into each other can nonetheless explain why they each are smiling then.

⁸ It is natural to think that two events that happen at the same time comprise a coincidence precisely when they are causally/explanatorily unconnected (neither causes/explains the other and there is no common cause/explanation). This does not mean that coincidences are inexplicable; that's what separate cause explanations provide [see Sober (2012), p. 362 for discussion].

⁹ In addition, Sober (2015) does not commit to the thesis that if the common cause explanation has the higher likelihood, then its value is much bigger than 0.5 whereas the likelihood of the separate cause explanation is much smaller than 0.5.

¹⁰ Lange argues not just that there are realistic case where $\Pr(H \mid O \& EXPL) > \Pr(H \mid O)$, but also that there are cases where $\Pr(H \mid O \& EXPL) = \Pr(H \mid O)$ because $EXPL$ is a necessary truth. He writes:

The problem with cases where $EXPL$ is a logical necessity is that in such cases, although Roche and Sober are correct that $\Pr(H \mid O \& EXPL) = \Pr(H \mid O)$, this equality is trivial since $\Pr(EXPL) = 1$. The equality then fails to show that H 's explanatoriness counts for nothing in its confirmation [Lange (2017), p. 308].

We have three comments. First, even if Lange is right that there are cases where $\Pr(H \mid O \& EXPL) = \Pr(H \mid O)$ because $EXPL$ is a necessary truth, this leaves it open, as per SOT, that there are also many realistic cases in which the background information codified in $\Pr(-)$ includes frequency data such that O screens-off $EXPL$ from H in that $\Pr(H \mid O \& EXPL) = \Pr(H \mid O)$. Second, $EXPL$ isn't a necessary truth in cases like our smoking case. Third, as we explain in Sections III IV, and V, SOT and its cousins SOT* and SOT** are far from trivial, given their implications with respect to various versions of IBE.

¹¹ Some theorists argue that belief and acceptance are distinct in that a subject can accept a given hypothesis without believing it. See, e.g., Elliott and Willmes (2013). We leave it open whether they are right; maybe there are cases in which $\Pr(H \mid O \& EXPL \& BEST) \leq 0.5$ where it's rational to accept H (though not to believe it).

¹² It might be that t varies from context to context. We take no stand on this.

¹³ Cabrera (2017) has pointed out that some of the explanatory virtues are antithetical to high probability. For example, if theory T_1 entails theory T_2 , then T_1 can't have a higher posterior probability than T_2 , no matter what the evidence is. That said, T_1 may have wider scope than T_2 .

¹⁴ It might be that, strictly speaking, there can be cases where H is a potential explanation of O , and yet it's false that if H and O were true, then H would explain O , because something else would explain it [see Lipton (2004), Ch. 4 for discussion]. We're ignoring this possibility here (as is standard).

¹⁵ This objection is distinct from van Fraassen's (1989) "best of a bad lot" objection. See Okasha (2000) for discussion of the latter.

¹⁶ For a recent taxonomy of explanatory virtues, and for references, see Keas (2018). See also Beebe (2009); Douven (2017), sec. 2; Harman (1965); Lipton (2004), Chs. 7 and 8; Lycan (2002), sec. 3; McMullin (2008); and Psillos (2002).

¹⁷ There are 2047 sets of one or more of (a)-(k). ¹⁸ See, for example, Blanchard (2018), Lange (2004), and Patrick (2018). There are ways of understanding unification on which our assumption is false. This is true, for example, of Friedman's (1974) account of unification, since it's restricted to propositions about laws of nature. For discussion of Friedman's ideas about unification, and, more generally, his unificationist theory of explanation, see Roche and Sober (2017a).

¹⁹ We are assuming, as is natural, that for any rival available potential explanations H and H^* (of O), if (i) H is superior to H^* in at least one of v_1, v_2, \dots , and v_n and (ii) H isn't inferior to H^* in any of v_1, v_2, \dots , and v_n , then H is better overall in terms of v_1, v_2, \dots , and v_n than H^* .

²⁰ There's a variant of IBE** where the fourth condition in the antecedent is the condition that H 's overall score in terms of v_1, v_2, \dots , and v_n is *significantly greater than* the overall scores of the other potential explanations in question [see Lycan (2002), p. 414 for discussion]. If this condition can be met even though *HIGH* is false, then there can be cases where IBE** and this variant of it come apart. Even so, what we say below about the former carries over to the latter.

²¹ See Douven (2017), sec. 2, for further discussion and references.

²² We suspect that the genetics example in Roche and Sober (2017b) can be adapted to show this.

²³ See Roche (2018) for related discussion on parsimony and background information.

REFERENCES

- BEEBE, J. (2009), "The Abductivist Reply to Skepticism"; *Philosophy and Phenomenological Research*, 79, pp. 605-636.
- BLANCHARD, T. (2018), "Bayesianism and Explanatory Unification: A Compatibilist Account"; *Philosophy of Science*, 85, 682-703.
- CABRERA, F. (2017), "Can there be a Bayesian Explanationism? On the Prospects of a Productive Partnership"; *Synthese*, 194, pp. 1245-1272.

- CLIMENHAGA, N. (2017), "How Explanation Guides Confirmation"; *Philosophy of Science*, 84, pp. 359-368.
- DOUVEN, I. (2017), "Abduction"; in E. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (summer ed.). URL = <<http://plato.stanford.edu/archives/sum2017/entries/abduction/>>.
- ELLIOTT, K., and D. WILLMES (2013), "Cognitive Attitudes and Values in Science"; *Philosophy of Science*, 80, pp. 807-817.
- FRIEDMAN, M. (1974), "Explanation and Scientific Understanding"; *Journal of Philosophy*, 71, pp. 5-19.
- HARMAN, G. (1965), "The Inference to the Best Explanation"; *Philosophical Review*, 74, pp. 88-95.
- HEMPEL, C. (1965), "Studies in the Logic of Confirmation"; in C. Hempel, *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*, New York: Free Press, pp. 3-46.
- KEAS, M. (2018), "Systematizing the Theoretical Virtues"; *Synthese*, 195, pp. 2761-2793.
- LANGE, M. (2004), "Bayesianism and Unification: A reply to Wayne Myrvold"; *Philosophy of Science*, 71, pp. 205-215.
- (2017), "The Evidential Relevance of Explanatoriness: A Reply to Roche and Sober"; *Analysis*, 77, pp. 303-312.
- LIPTON, P. (2004), *Inference to the Best Explanation* (2nd ed.); London: Routledge.
- LYCAN, W. (2002), "Explanation and Epistemology"; in P. Moser (Ed.), *The Oxford Handbook of Epistemology*; Oxford: Oxford University Press, pp. 408-433.
- MCCAIN, K., and T. POSTON (2014), "Why Explanatoriness is Evidentially Relevant"; *Thought*, 3, pp. 145-153.
- (2017), "The Evidential Impact of Explanatory Considerations"; in K. McCain and T. Poston (Eds.), *Best Explanations: New Essays on Inference to the Best Explanation*; Oxford: Oxford University Press, pp. 121-129.
- MCMULLIN, E. (2008), "The Virtues of Good Theories"; in S. Psillos and M. Curd (Eds.), *The Routledge Companion to Philosophy of Science*, London: Routledge, pp. 498-508.
- OKASHA, S. (2000), "Van Fraassen's Critique of Inference to the Best Explanation"; *Studies in the History and Philosophy of Science*, 31, pp. 691-710.
- PATRICK, K. (2018), "Unity as an Epistemic Virtue"; *Erkenntnis*, 83, pp. 983-1002.
- PSILLOS, S. (2002), "Simply the Best: A Case for Abduction"; in A. Kakas and F. Sadri (Eds.), *Computational Logic: Logic Programming and Beyond*; Berlin: Springer-Verlag, pp. 605-625.
- ROCHE, W. (2017), "Explanation, Confirmation, and Hempel's Paradox"; in K. McCain and T. Poston (Eds.), *Best Explanations: New Essays on Inference to the Best Explanation*; Oxford: Oxford University Press, pp. 219-241.
- (2018), "The Perils of Parsimony"; *Journal of Philosophy*, 115, pp. 485-505.

- ROCHE, W., and E. SOBER (2013), “Explanatoriness is Evidentially Irrelevant, or Inference to the Best Explanation Meets Bayesian Confirmation Theory”; *Analysis*, 73, pp. 659-668.
- (2014), “Explanatoriness and Evidence: A Reply to McCain and Poston”; *Thought*, 3, pp. 193-199.
- (2017a), “Explanation = Unification? A New Criticism of Friedman’s Theory and a Reply to an Old One”; *Philosophy of Science*, 84, pp. 391-413.
- (2017b), “Is Explanatoriness a Guide to Confirmation? A Reply to Climenhaga”; *Journal for General Philosophy of Science*, 48, pp. 581-590.
- SOBER, E. (2012), “Coincidences and How to Reason About Them”; *European Philosophy of Science Association Proceedings*, 1, pp. 355-374.
- (2015), *Ockham’s Razors – A User’s Manual*; Cambridge: Cambridge University Press.
- VAN FRAASSEN, B. (1989), *Laws and Symmetry*; Oxford: Oxford University Press.
- WILLIAMSON, T. (2016), “Abductive Philosophy”; *Philosophical Forum*, 47, pp. 263-280.