A Second Look at the Logic of Explanatory Power (with Two Novel Representation Theorems)

Vincenzo Crupi and Katya Tentori*†

We discuss the probabilistic analysis of explanatory power and prove a representation theorem for posterior ratio measures recently advocated by Schupbach and Sprenger. We then prove a representation theorem for an alternative class of measures that rely on the notion of relative probability distance. We end up endorsing the latter, as relative distance measures share the properties of posterior ratio measures that are genuinely appealing, while overcoming a feature that we consider undesirable. They also yield a telling result concerning formal accounts of explanatory power versus inductive confirmation, thereby bridging our discussion to a so-called no-miracle argument.

1. Introduction. *Explanation* is a central notion in human reasoning and a traditional topic in the philosophy of science. In a recent article, Schupbach and Sprenger (2011) have put forward a probabilistic analysis of how successful a hypothesis or theory h is in explaining an event or state of affairs of interest e. In our view, the general philosophical assumptions of Schupbach and Sprenger's approach are sensible, and the theoretical endeavor that they pursue is highly valuable. Thus, in what follows, we will rely on how they define their framework of inquiry. However, we will subject their specific proposal, results, and interpretations to critical discussion.

Received June 2011; revised October 2011.

*To contact the authors, please write to: Vincenzo Crupi, University of Turin, via S. Ottavio 20, 10124 Torino, Italy; e-mail: vincenzo.crupi@unito.it. Katya Tentori, University of Trento, corso Bettini 31, 38068 Rovereto, Italy; e-mail: katya.tentori@unitn.it.

†We thank Jonah Schupbach, Jan Sprenger, Valeriano Iranzo, and an anonymous reviewer for useful comments and exchanges concerning this work. Our research has been supported by the Alexander von Humboldt Foundation, the Italian Ministry of Scientific Research (PRIN grant 20083NAH2L_001), and the Spanish Department of Science and Innovation (grant FFI2008-01169/FISO).

Philosophy of Science, 79 (July 2012) pp. 365–385. 0031-8248/2012/7903-0003$10.00

We will use E to label a probabilistic measure of explanatory power and write $E_P(e, h)$, with e and h denoting contingent statements for the explanandum and candidate explanans, respectively, and P a relevant probability function, which is assumed to be regular (i.e., such that, for any contingent statement α, $0 < P(\alpha) < 1$).[1] Notably, $E_P(e, h)$ is meant as a measure of how successful h is in explaining e assuming that h qualifies as potentially relevant to e in explanatory terms. Following Schupbach and Sprenger (2011, 106–8), we will see the latter caveat as implied in the use of $E_P(e, h)$ but not encoded in $E_P(e, h)$ itself. Accordingly, $E_P(e, h)$ will reflect the strength of an explanans-explanandum relation between h and e, if any is present. A systematic characterization of when such a relation per se obtains, although of course relevant and much deserving, is seen here as a distinct topic of investigation.

2. Reviewing Adequacy Conditions. Schupbach and Sprenger's favorite probabilistic measure of explanatory power is as follows:

$$\varepsilon(e, h) = \frac{P(h|e) - P(h|\neg e)}{P(h|e) + P(h|\neg e)}.$$

In their main formal result, they prove that $E_P(e, h) = f[\varepsilon(e, h)]$ must hold (where f is a strictly increasing function), provided that four conditions CA1–CA4 are satisfied (see their theorem 1; 2011, 111). This pinpoints a class of ordinally equivalent measures that are represented by some increasing function of the posterior ratio $P(h|e)/P(h|\neg e)$. (To appreciate this, consider the following rendition of $\varepsilon(e, h) : \varepsilon(e, h) = \tanh(1/2 \{\ln[P(h|e)/P(h|\neg e)]\})$.) We will now briefly comment on each of Schupbach and Sprenger's adequacy conditions in turn.

> **CA1**. There exists an analytic function g such that, for any contingent e, h and any regular P, $E_P(e, h) = g[P(h|e), P(h|\neg e), P(e)]$. Values of $E_P(e, h)$ range in $[-1, +1]$.

Schupbach and Sprenger label this condition *formal structure* and describe it as "rather uncontentious" (2011, 109). Some cautionary notes are in order, though. As they show, all measures satisfying their CA1–CA4 are ordinally equivalent posterior ratio measures, as defined above. However, it is not the case that all posterior ratio measures satisfy CA1–CA4, and this is especially due to CA1, which goes beyond constraining the ordinal structure. For instance, there is no doubt that it is mathematically elegant to have $E_P(e, h)$

1. As usual, a further term B could be included to represent relevant background knowledge and assumptions, thus having $E_P(e, h|B)$. Such a term will be omitted from our notation for simple reasons of convenience, as it is inconsequential for our discussion.

normalized to range within $[-1, +1]$. If that specific range is imposed, however, we have that posterior ratio measures such as $E_P(e,h) = n\varepsilon(e,h) + n$ (with $n > 0$) are ruled out just because the range becomes $[0, 2n]$. Similar consequences arise from the analyticity requirement in CA1, for there exist posterior ratio measures that are not analytic, thus violating CA1.[2] By this restrictive character, CA1 prevents Schupbach and Sprenger's main result from being a proper *representation theorem* for posterior ratio measures. We will come back to this point soon.

> **CA2**. Ceteris paribus, the greater the degree of statistical relevance between e and h, the greater the value of $E_P(e,h)$.

Schupbach and Sprenger's label for this condition is *positive relevance*. Our sole remark here is that CA2 is left rather unspecified, as statistical relevance can be measured in various ways. This is innocuous for their proof. Yet, to allow for a discussion of the content and plausibility of this assumption, a sharper rendition would be helpful.

> **CA3**. If h_2 is probabilistically independent from e, h_1, and their conjunction (i.e., $P(e \wedge h_2) = P(e)P(h_2)$, $P(h_1 \wedge h_2) = P(h_1)P(h_2)$, and $P(e \wedge h_1 \wedge h_2) = P(e \wedge h_1)P(h_2)$), then $E_P(e,h_1) = E_P(e, h_1 \wedge h_2)$.

Schupbach and Sprenger's line of argument for this condition, which we do not dispute, goes as follows. Statement h_2 is assumed to be irrelevant, so we have $P(e|h_1) = P(e|h_1 \wedge h_2)$. But then, as adding h_2 leaves the degree to which e is expected unchanged, it does not alter the degree to which e is explained either. Their label for CA3 is *irrelevant conjunctions*. We suggest *conjunction of irrelevant explanantes*, though, for we will have to deal with the conjunction of irrelevant explananda as a separate issue later on.

> **CA4**. If $\neg h$ entails e, then the values of $E_P(e,h)$ do not depend on the values of $P(h)$. Formally, there exists a function f so that, if $\neg h \vDash e$, then either $E_P(e,h) = f[P(h|e)]$ or $E_P(e,h) = f[P(e)]$.

We struggle to see a clear connection between the informal and the formal clause in CA4 and, thus, to interpret this condition as conveying *irrelevance of priors*, as Schupbach and Sprenger label it.[3] In fact, we decidedly concur

2. Here is an example: $E_P(e,h) = 1 - [P(h|\neg e)/P(h|e)]$ if $P(e|h) \geq P(e)$; $E_P(e,h) = [P(h|e)/P(h|\neg e)] - 1$ if $P(e|h) < P(e)$.

3. Let us articulate our worry more precisely. CA4 is meant to say that $E_P(e,h)$ is a function of either just $P(h|e)$ or just $P(e)$ in the target class of cases. Yet none of these quantities can be said to make the prior $P(h)$ "irrelevant," lacking additional details. In

with them that intuition is hardly “strong enough to make a conclusive case for CA4” (2011, 111).

3. A Representation Theorem for Posterior Ratio Measures. The above remarks were meant to motivate a partly different approach to the derivation of posterior ratio measures of explanatory power to be pursued in the present section. To this purpose, let us introduce a propositional language L and the set L_c of the contingent formulas in L. Further, let $\boldsymbol{P}$ be the set of all regular probability functions that can be defined over L (so that for any $P \in \boldsymbol{P}$ and any $\alpha \in L_c$, $0 < P(\alpha) < 1$) and posit $E_P : \{L_c \times L_c \times \boldsymbol{P}\} \rightarrow \Re$. Each $P \in \boldsymbol{P}$ can be seen as representing a possible state of belief concerning a domain described in L. As a first move, we propose the following explication of CA2 above:

> **E1**. *Positive relevance*. For any $e, h_1, h_2 \in L_c$ and any $P \in \boldsymbol{P}$, $E_P(e, h_1) \gtreqless E_P(e, h_2)$ if and only if (iff) $P(e|h_1) \gtreqless P(e|h_2)$.

In essence, E1 states that $E_P(e, h)$ is a strictly increasing function of $P(e|h)$ when $P(e)$ is kept fixed. Indeed, this assumption closely fits Schupbach and Sprenger’s remark that (positive) explanatory power reflects “the strength” of the inequality $P(e|h) > P(e)$ (see 2011, 110). An additional feature of E1 is that we immediately get an analogue of CA3 above, that is, an appropriate principle concerning the conjunction of irrelevant explanantes. Indeed, E1 implies that, for any $e, h_1, h_2 \in L_c$ and any $P \in \boldsymbol{P}$, if h_2 is probabilistically independent from e, h_1, and their conjunction, then $E_P(e, h_1) = E_P(e, h_1 \wedge h_2)$. As it is redundant on our E1, this condition does not need to be stated on its own.

Second, it will prove convenient to retain a different statement of CA4 that Schupbach and Sprenger had employed in earlier versions of their work.[4] The informal wording simply read “values of $E_P(e, h)$ do not depend on the values of $P(h)$” (i.e., without the restrictive caveat about $\neg h$ implying e that is involved in CA4 as spelled out above). Upon scrutiny, however, the proper formal rendition of the assumption that was intended by this statement

fact, it all depends on what is or is not being concurrently kept fixed. For instance, if we happen to keep $P(e|h)$ and $P(e)$ fixed, then of course $P(h|e)$ and $P(h)$ vary together. If $P(h|e)$ and $P(h|\neg e)$ are kept fixed instead, then it is $P(e)$ and $P(h)$ that vary together. Surely, it could also be the case that, say, $P(e)$ varies “independently” from $P(h)$ (i.e., with the latter possibly remaining fixed), provided that room for variation is allowed for somewhere else. But condition CA4 is silent on all this. For this reason, we think that the connection with the “irrelevance of priors” is not well specified. (We thank Jan Sprenger for prompting this clarification.)

4. See http://philsci-archive.pitt.edu/5521/1/ExplanatoryPower.pdf, 7.

turns out to yield a substantial constraint on the functional form of $E_P(e,h)$.[5] For lack of anything better, we will label this condition *posteriors* (its implications and plausibility will be discussed later on).

> **E2**. *Posteriors*. There exists a function g such that, for any $e, h \in L_c$ and any $P \in \boldsymbol{P}$, $E_P(e,h) = g[P(h|e), P(h|\neg e)]$.

It is then possible to prove the following theorem (see app. A):

> **Theorem 1**. E1 and E2 hold iff there exists a strictly increasing function f such that, for any $e, h \in L_c$ and any $P \in \boldsymbol{P}$, $E_P(e,h) = f[\varepsilon(e,h)]$.

As a bidirectional implication, theorem 1 shows that E1 and E2 are both sufficient and necessary to single out the set of posterior ratio measures of explanatory power, thus providing a representation theorem, that is, an axiomatic foundation carving out distinctive traits of these ordinally equivalent measures taken together. Notably, this has not been achieved by invoking a more demanding set of conditions than CA1–CA4, for the range of $E_P(e,h)$ is left entirely unconstrained, and no analyticity, differentiability, or continuity requirement is involved.[6] We now feel in a better position to precisely discuss how appealing posterior ratio measures of explanatory power are in philosophical terms.

4. Explanatory Justice. A clear and general axiomatization of posterior ratio measures fosters insight into their distinctive properties and thus a focused discussion of their implications. As a starting point for such a discussion, we will consider a quote from Schupbach and Sprenger (2011, 115) in which they criticize an important competing measure of explanatory power—that is, $I(e,h) = \ln[P(e|h)/P(e)]$—originally adopted by Good (1960) and more recently by McGrew (2003):

> Consider the following example: let e be a general description of the Brownian motion observed in some particles suspended in a particular liquid, and let h be Einstein's atomic explanation of this motion. Of course, h constitutes a lovely explanation of e. . . . However, take any irrelevant new statement e^* and conjoin it to e; for example, let e^* be the proposition that the mating season for an American green tree frog takes place from mid-April to mid-August. In this case, measure $I(e,h)$ judges

5. We are grateful to Jonah Schupbach and Jan Sprenger for an e-mail exchange that helped us get this point clear.

6. Analyticity implies differentiability, which in turn implies continuity. Notice that the requirement of continuity has been criticized as unduly demanding in similar results concerning the derivation of probabilistic measures of confirmation (see Fitelson 2006, 506 n. 12).

> that Einstein hypothesis explains Brownian motion to the same extent that it explains Brownian motion and this fact about tree frogs. Needless to say, this result is deeply unsettling.

On the basis of this remark, Schupbach and Sprenger put forward the requirement that $E_P(e \wedge e^*, h) < E_P(e, h)$ in such circumstances (i.e., when e is an explanatory success of h and e^* is probabilistically independent from e, h, and their conjunction), which is in fact implied by any posterior ratio measure. Presumably, the intuitive motivation here reflects a sense of "explanatory justice," to be informally understood as follows: a certain amount of explanatory success of h concerning e, no matter how large, cannot be extended "for free." That is, it would not make sense to claim the same amount of explanatory success concerning any logically stronger statement $e \wedge e^*$—unless of course some positive explanatory import of h on e^* independently exists.[7] Along this line of argument, we urge, careful and separate consideration should be given to the case of explanatory failure.

Let us illustrate with a mundane example. Assume there is a fair coin that can only be of one of two kinds: either it is a normal coin or it has tails on both sides. Label h the former hypothesis (normal coin) and $\neg h$ the latter (no heads, two tails). Let e be a streak of tosses of, say, 10 tails. Clearly, $\neg h$ explains e perfectly well, while h fares poorly on this account. But now let e^* be involved once again (i.e., the American green tree frogs' mating season). Should we say that the explanatory failure of h is at all mitigated if e^* is conjoined to e? We think not. Again it seems a matter of "explanatory justice." For, should it be the case that $E_P(e \wedge e^*, h) > E_P(e, h)$ in these circumstances, then one would be allowed to indefinitely relieve a lack of explanatory power, no matter how large, by adding more and more irrelevant explananda, simply at will. A strongly unpalatable outcome—yet this is precisely what any posterior ratio measure of explanatory power implies, including of course $\varepsilon(e, h)$. Schupbach and Sprenger are aware of this property (see 2011, 115), but they do not offer a discussion dispelling what we see as its highly counterintuitive character. On the basis of the foregoing, we put forward the following principle:

7. An anonymous reviewer suggested that Schupbach and Sprenger's quote is unconvincing to begin with, for in their view a measure of explanatory power only applies when a genuine explanatory connection is assumed to be present, which might fail to be the case when irrelevant statement e^* is conjoined to explanandum e, so that the behavior of $E_P(e, h)$ in these cases would become quite inconsequential. We find this worry legitimate, although by no means conclusive. Space limitations force us to postpone more thorough discussion to another occasion, but a relevant remark will arise in n. 10 below.

E2*. *Explanatory justice* (*upward and downward*).

i) For any e, e^*, $h \in L_c$ and any $P \in \boldsymbol{P}$, if $P(e|h) > P(e)$ and e^* is probabilistically independent from e, h, and their conjunction, then $E_P(e \wedge e^*, h) < E_P(e, h)$.
ii) For any e, e^*, $h \in L_c$ and any $P \in \boldsymbol{P}$, if $P(e|h) \leq P(e)$ and e^* is probabilistically independent from e, h, and their conjunction, then $E_P(e \wedge e^*, h) = E_P(e, h)$.

Recall that Schupbach and Sprenger criticized Good's measure of explanatory power $I(e, h)$ for violating E2*i, which $\varepsilon(e, h)$ satisfies (along with all posterior ratio measures). For the sake of brevity and vividness, we will say that posterior ratio measures are *just upward*, while Good's measure is not. Notably, the pattern reverses with E2*ii, which is implied by Good's measure $I(e, h)$ but violated by $\varepsilon(e, h)$ (and all posterior ratio measures). Good's measure is thus *just downward*, while posterior ratio measures are not. The question naturally arises, then, whether E2*i and E2*ii can ever be satisfied at once, that is, whether there exist measures displaying overall explanatory justice.[8] This is the topic of the next section.

5. A Further Representation Theorem. Let us go back to the properties of posterior ratio measures and pinpoint the culprit for their lack of downward explanatory justice. Based on our theorem 1, the problem is easily located in E2. To show this, we will now drop that assumption and replace it in such a way as to obtain a new and different representation theorem by which both clauses in E2* are fulfilled.

First, to dispense with E2, we will rely on another technical assumption about $E_P(e, h)$, that is,

E0. *Formality*. There exists a function g such that, for any $e, h \in L_c$ and any $P \in \boldsymbol{P}$, $E_P(e, h) = g[P(e \wedge h), P(e), P(h)]$.

8. One could wonder how the strict equality in E2*ii can be motivated, instead of "equal to or lower than." Here is one possible argument. Consider the limiting case of deductive explanation, i.e., such that $h \vDash e$, and so that h and $\neg e$ are inconsistent. If and only if this holds, for each of Schupbach and Sprenger's posterior ratio measures and each of our relative distance measures alike (see below), $E_P(e, h)$ gets a fixed maximum value and $E_P(\neg e, h)$ a fixed minimum value, which seems rather appealing (also see Schupbach and Sprenger 2011, 111, corollary 1). Accordingly, if e^* is irrelevant, one also has $E_P(e \wedge e^*, h) < E_P(e, h)$ (for the conjunctive explanandum $e \wedge e^*$, unlike e, is not implied by h) along with $E_P(\neg e \wedge e^*, h) = E_P(\neg e, h)$ (for the conjunctive explanandum $\neg e \wedge e^*$ is still inconsistent with h, just as $\neg e$ was). While the strict inequality in E2*i neatly generalizes the former fact, the strict equality in E2*ii does so with the latter. (We thank an anonymous reviewer for raising this issue.)

For a purely probabilistic theory of explanatory power, assumption E0 is almost as undemanding as it can be. It simply states that $E_P(h, e)$ depends on the probability distribution over the algebra generated by h and e, which is entirely determined by $P(e \wedge h)$, $P(e)$, and $P(h)$.[9]

Now we only need to include what follows:

> **E3**. *Symmetry*. For any $e_1, e_2, h \in L_c$ and any $P \in \boldsymbol{P}$, $E_P(e_1, h) \gtreqless E_P(e_2, h)$ iff $E_P(\neg e_1, h) \lesseqgtr E_P(\neg e_2, h)$.

E3 postulates an inverse (ordinal) correlation between the explanatory success and explanatory failure of a hypothesis with regards to complementary events. Notably, it is purely ordinal in nature and, thus, a weaker version of a homonymous condition that Schupbach and Sprenger themselves advocate and motivate with the remark that "the less surprising (more expected) the truth of e is in light of a hypothesis, the more surprising (less expected) is e's falsity" (2011, 113).

The announced result is thus as follows (see app. B for a proof):

> **Theorem 2**. E0, E1, E2*, and E3 hold iff there exists a strictly increasing function f such that, for any $e, h \in L_c$ and any $P \in \boldsymbol{P}$, $E_P(e, h) = f[\varepsilon^*(e, h)]$, where

$$\varepsilon^*(e,h) = \begin{cases} \dfrac{P(e|h) - P(e)}{1 - P(e)} & \text{if } P(e|h) \geq P(e) \\ \dfrac{P(e|h) - P(e)}{P(e)} & \text{if } P(e|h) < P(e) \end{cases}.$$

The answer to our question above concerning overall explanatory justice is then in the positive. There exists a class of ordinally equivalent measures of explanatory power that—unlike either Schupbach and Sprenger's $\varepsilon(e, h)$ or Good's $I(e, h)$—displays justice both upward and downward, that is, for both explanatory success and failure. The crucial step is to assume E2* straight away as an axiom and to get rid of E2.[10]

9. In essence, E0 is nothing else than Schupbach and Sprenger's CA1 itself without the restrictions concerning analyticity and range. The label *formality* is adopted after Tentori, Crupi, and Osherson (2007, 2010).

10. In personal communication, Schupbach and Sprenger have disputed E2* by suggesting that conjoining irrelevant explananda should water down degrees of explanatory success and failure alike (also see Schupbach and Sprenger 2011, 115), as follows: (W) For any $e, e^*, h \in L_c$ and any $P \in \boldsymbol{P}$, if e^* is probabilistically independent from e, h, and their conjunction, then $|E_P(e \wedge e^*, h)| < |E_P(e, h)|$. Note, however, that $\varepsilon(e, h)$ does not imply W, as it gives $\varepsilon(e \wedge e^*, h) = \varepsilon(e, h) = -1$ whenever $h \vDash \neg e$ (also see

Even beyond E0, E1, and E3, measures $\varepsilon(e,h)$ and $\varepsilon^*(e,h)$ share further desirable properties spelled out by Schupbach and Sprenger. For instance, both of them take maximum (minimum) value iff h implies e ($\neg e$). Yet given the conflicting consequences of E2 and E2*, our theorems 1 and 2 imply that $\varepsilon(e,h)$ and $\varepsilon^*(e,h)$ are nonequivalent, not even in purely ordinal terms. For instance, ε may rank e_1 rather than e_2 as a stronger explanatory success of a given h, while ε^* implies the opposite; that is, there exist cases in which $\varepsilon(e_1,h) > \varepsilon(e_2,h)$, while $\varepsilon^*(e_1,h) < \varepsilon^*(e_2,h)$. So $\varepsilon(e,h)$ and $\varepsilon^*(e,h)$ represent two genuinely alternative ways to explicate the notion of explanatory power. As a historical note, it should be pointed out that $\varepsilon^*(e,h)$ is not an entirely new idea, as its positive branch already appears in Niiniluoto and Tuomela (1973, 66, 89; also see Pietarinen 1970). The next section will discuss some further properties of this measure.

6. Explanatory Power, Confirmation, and Miracles. There exists a rather effective rule of thumb to generate a decent candidate measure of explanatory power; that is, take a plausible probabilistic measure of (incremental) confirmation and invert the positions of e and h. Indeed, Schupbach and Sprenger readily notice that their measure $\varepsilon(e,h)$ entertains this relationship with a very interesting measure of confirmation or inductive support originally defined by Kemeny and Oppenheim (1952) and more recently revived by Fitelson (2005); that is,

$$l(h,e) = \frac{P(e|h) - P(e|\neg h)}{P(e|h) + P(e|\neg h)}.$$

(The label l is employed here because this quantity is a strictly increasing function of the likelihood ratio $P(e|h)/P(e|\neg h)$.)

As shown below, the same kind of structural analogy holds between $\varepsilon^*(e,h)$ and yet another measure of confirmation, $z(h,e)$, which is based on the notion of the "relative distance" between the initial and final probability of a hypothesis (see Crupi, Tentori, and Gonzalez 2007; Crupi, Festa, and Buttasi 2010; Crupi and Tentori 2010, forthcoming):

n. 8). Thus, posterior ratio measures do not convey the idea of watering down in a fully coherent fashion. Second, by allowing for watering down in (nonextreme) cases of explanatory failure, $\varepsilon(e,h)$ also allows for e being a worse explanatory failure than $e \wedge e^*$, despite e^* being itself a (mild) explanatory failure for a given h, and even if h screens off e^* from e (proof omitted). So, e.g., let h be the hypothesis that Lee Harvey Oswald acted alone in the assassination of JFK. There exist sensible subjective probability estimates by which, according to a posterior ratio measure, h would fail to explain the sequence of shots in Dallas (e) but fail to a lesser extent with the conjunctive explanandum consisting in both that sequence of shots and the subsequent killing of Oswald ($e \wedge e^*$). We find it hard to see how rankings of this kind could be rationalized and highlight as a virtue of explanatory justice (E2*) that it prevents them from arising.

$$z(h,e) = \begin{cases} \dfrac{P(h|e) - P(h)}{1 - P(h)} & \text{if } P(h|e) \geq P(h) \\ \dfrac{P(h|e) - P(h)}{P(h)} & \text{if } P(h|e) < P(h) \end{cases}.$$

$$\varepsilon^*(e,h) = \begin{cases} \dfrac{P(e|h) - P(e)}{1 - P(e)} & \text{if } P(e|h) \geq P(e) \\ \dfrac{P(e|h) - P(e)}{P(e)} & \text{if } P(e|h) < P(e) \end{cases}.$$

In considering pairs of analogous expressions for explanatory power and inductive confirmation, such as those above, one should always keep in mind a crucial theoretical difference. In the measurement of explanatory power, the hypothesis *h* at issue is meant to be in some explanatory relation (to be separately defined) with evidence *e*—a caveat with no apparent counterpart in probabilistic confirmation theory. That said, $\varepsilon^*(e,h)$ is itself a measure of relative distance, the latter notion involving the background probability of the explanandum, $P(e)$, and the associated likelihood of the candidate explanans, $P(e|h)$. In particular, when *h* enjoys some amount of explanatory success relative to *e* (so that $P(e|h) > P(e)$), $\varepsilon^*(e,h)$ assesses that explanatory success by the proportion of the background surprisingness of *e* (i.e., $1 - P(e)$) that is removed by assuming *h* (i.e., by covering the positive difference between $P(e|h)$ and $P(e)$). However, in case *e* is at odds with *h* (so that $P(e|h) < P(e)$), $\varepsilon^*(e,h)$ rates explanatory failure by a negative value that is lower and lower, the higher the proportion of the background expectedness of *e* (i.e., $P(e)$) that is removed by assuming *h* (i.e., by covering the negative difference between $P(e|h)$ and $P(e)$).[11]

In view of these structural analogies, investigating the connections between probabilistic explanatory power $E_P(e,h)$ and confirmation $C_P(h,e)$ appears appropriate (see Crupi 2012). Interestingly, if relative distance measures of confirmation and explanatory power are concurrently adopted—that is, if one posits both $C_P(h,e) = f[z(h,e)]$ and $E_P(e,h) = g[\varepsilon^*(e,h)]$, where *f* and *g* are strictly increasing functions—then one can easily derive what follows (proof omitted):

> **M**. *No miracle (retail version)*. For any $e_1, e_2, h_1, h_2 \in L_c$ and any $P \in \boldsymbol{P}$, if $P(h_1|e_1) > P(h_1)$ and $P(h_2|e_2) > P(h_2)$, then $C_P(h_1,e_1) \gtreqless C_P(h_2,e_2)$ iff $E_P(e_1,\neg h_1) \lesseqgtr E_P(e_2,\neg h_2)$.

11. As suggested by an anonymous reviewer, $\varepsilon^*(e,h)$ can be written in yet another way that is possibly instructive, as follows: $\varepsilon^*(e,h) = [P(e|h) - P(e)]/[1 - P(e)]$ if $P(e|h) \geq P(e)$; $\varepsilon^*(e,h) = -[P(\neg e|h) - P(\neg e)]/[1 - P(\neg e)]$ if $P(\neg e|h) > P(\neg e)$.

M postulates an inverse (ordinal) correlation between the degree of positive confirmation that a successful explanatory hypothesis h receives from the occurrence of explanandum e and the degree to which e fails to be explained by $\neg h$. So explanatory hypothesis h is confirmed by e to the extent that e appears inexplicable (a miracle, as it were), assuming the falsity of h. Notably, if $z(h, e)$ and $\varepsilon^*(e, h)$ are chosen among the corresponding classes of ordinal equivalence, a straightforward quantitative counterpart of M obtains; that is $z(h, e) = -\varepsilon^*(e, \neg h)$ just in case $P(e|h) > P(e)$.

While admittedly only a first step in a formal analysis of the subtle relations between confirmation and explanatory power, this result strikes us nonetheless as a telling implication of relative distance measures. In fact, the label for M is freely adapted after Worrall (2006), who contrasts "retail" instances of a so-called no-miracle argument with its "wholesale" variant. In Worrall's terms, retail instances of the argument involve specific (scientific) hypotheses, while the wholesale argument (which he criticizes) would concern the global philosophical stance of a realist view of "all" science. Despite the differences between Worrall's and our theoretical framework and objectives, we believe that the meaning of M vindicates our terminological suggestion.[12]

7. Conclusion. In our discussion, we have retained much of Schupbach and Sprenger's original approach in the inquiry on the notion of explanatory power. We have proved a representation theorem for posterior ratio measures, which Schupbach and Sprenger favor, along with a different representation theorem for an alternative set of measures relying on the notion of relative distance in probability. We end up advocating the latter and $\varepsilon^*(e, h)$ as a suitable exemplar of its class of ordinal equivalence. We have motivated this preference through several remarks. First, $\varepsilon^*(e, h)$ shares the properties set out by Schupbach and Sprenger that are genuinely appealing, while overcoming an undesirable trait of posterior ratio measures, that is, departure from downward justice upon the conjunction of irrelevant explananda. Second, $\varepsilon^*(e, h)$ conveys in probabilistic terms a view of explanatory power that is conceptually sound, as it transparently involves how the background surprisingness/expectedness of explanandum e is reduced by assuming candidate explanans h. Finally, relative distance measures yield a telling

12. Our result M is clearly not meant to bear direct consequences on the philosophical debate on scientific realism but only to pinpoint a sharp probabilistic rendition of (retail) no-miracle arguments. Many realists would add that, for h and $\neg h$ to have a genuinely explanatory connection with e, they could not be instrumentalistically construed to begin with. In virtue of its formal nature, however, M does not fix a specific account of the relationship between explanatory relevance and the status of scientific hypotheses. (We thank Jan Sprenger for prompting this remark.)

preliminary result concerning the relation between formal representations of explanatory power and inductive confirmation, thus bridging discussions in this area with a so-called no-miracle argument.

Appendix A

Theorem 1. E1 and E2 hold iff there exists a strictly increasing function f such that, for any $e, h \in L_c$ and any $P \in \boldsymbol{P}$, $E_P(e,h) = f[\varepsilon(e,h)]$.

Proof.

Right-to-Left Implication.

E1. Let O denote *odds*; that is, $O(\alpha) = P(\alpha)/P(\neg\alpha)$. By the probability calculus (Bayes's theorem), $P(h|e)/P(h|\neg e) = O(e|h)/O(e)$. Keeping in mind that f is an increasing function and $\varepsilon(e,h)$ is an increasing function of the posterior ratio $P(h|e)/P(h|\neg e)$, we have $E_P(e,h_1) = f[\varepsilon(e,h_1)] \gtreqless f[\varepsilon(e,h_2)] = E_P(e,h_2)$ iff $O(e|h_1)/O(e) \gtreqless O(e|h_2)/O(e)$ iff $O(e|h_1) \gtreqless O(e|h_2)$ iff $P(e|h_1) \gtreqless P(e|h_2)$.

E2. Fulfillment of E2 immediately follows from $\varepsilon(e,h) = \tanh(1/2\ \{\ln[P(h|e)/P(h|\neg e)]\})$.

Left-to-Right Implication.

As a trivial consequence of E2, there exists a function j such that, for any $h, e \in L_c$ and any $P \in \boldsymbol{P}$, $E_P(e,h) = j[P(h|e)/P(h|\neg e), P(h|\neg e)]$. With no loss of generality, we will convey probabilistic coherence and regularity by constraining the domain of j to include pairs of values (x,y) such that (i) $0 < y < 1$, and (ii) $0 \leq x \leq 1/y$. We thus posit $j : \{(x,y) \in \{\Re^+ \cup \{0\}\} \times (0,1) | x \leq 1/y\} \rightarrow \Re$ and denote the domain of j as D_j.

Lemma 1.1. For any x, y_1, y_2 such that $x \in \Re^+ \cup \{0\}$, y_1, $y_2 \in (0,1)$, and $x \leq 1/y_1, 1/y_2$, there exist $e, h_1, h_2 \in L_c$ and $P' \in \boldsymbol{P}$ such that $P'(h_1|e)/P'(h_1|\neg e) = P'(h_2|e)/P'(h_2|\neg e) = x, P'(h_1|\neg e) = y_1$, and $P'(h_2|\neg e) = y_2$.

Proof. Let $w \in (0,1)$ be given. The equalities in lemma 1.1 arise from the following scheme of probability assignments:

$P'(h_1 \wedge h_2 \wedge e) = x^2 y_1 y_2 w$; $P'(\neg h_1 \wedge h_2 \wedge e) = (1 - xy_1)xy_2 w$;
$P'(h_1 \wedge h_2 \wedge \neg e) = y_1 y_2 (1-w)$; $P'(\neg h_1 \wedge h_2 \wedge \neg e) = (1-y_1)y_2(1-w)$;
$P'(h_1 \wedge \neg h_2 \wedge e) = xy_1(1 - xy_2)w$; $P'(\neg h_1 \wedge \neg h_2 \wedge \neg e) = (1 - xy_1)(1 - xy_2)w$;
$P'(h_1 \wedge \neg h_2 \wedge \neg e) = y_1(1-y_2)(1-w)$; $P'(\neg h_1 \wedge \neg h_2 \wedge \neg e) = (1-y_1)(1-y_2)(1-w)$.

Suppose there exist $(x, y_1), (x, y_2) \in D_j$ such that $j(x, y_1) \neq j(x, y_2)$. Then, by lemma 1.1 and the definition of D_j, there exist $e, h_1, h_2 \in L_c$ and $P' \in \boldsymbol{P}$ such that $P'(h_1|e)/P'(h_1|\neg e) = P'(h_2|e)/P'(h_2|\neg e) = x$, $P'(h_1|\neg e) = y_1$, and $P'(h_2|\neg e) = y_2$. By the probability calculus, if the latter equalities hold, then $P'(e|h_1) = P'(e|h_2)$. Thus, there exist $e,\ h_1,\ h_2 \in L_c$ and $P' \in \boldsymbol{P}$ such that $E_{P'}(e, h_1) = j(x, y_1) \neq j(x, y_2) = E_{P'}(e, h_2)$, even if $P'(e|h_1) = P'(e|h_2)$, contradicting E1. Conversely, E1 implies that, for any (x, y_1), $(x, y_2) \in D_j$, $j(x, y_1) = j(x, y_2)$. So, for E1 to hold, there must exist f such that, for any $e,\ h \in L_c$ and any $P \in \boldsymbol{P}$, $E_P(e, h) = f[P(h|e)/P(h|\neg e)]$, and $f(x) = j(x, y)$. We thus posit $f : \{\Re^+ \cup \{0\}\} \to \Re$ and denote the domain of f as D_f.

Lemma 1.2. For any $x_1, x_2 \in \Re^+ \cup \{0\}$, there exist $e, h_1, h_2 \in L_c$ and $P'' \in \boldsymbol{P}$ such that $P''(h_1|e)/P''(h_1|\neg e) = x_1$ and $P''(h_2|e)/P''(h_2|\neg e) = x_2$.

Proof. Let $y_1, y_2 \in (0, 1)$ and $w \in (0, 1)$ be given so that $y_1 \leq 1/x_1$ (as the latter quantity must be positive, y_1 exists) and $y_2 \leq 1/x_2$ (as the latter quantity must be positive, y_2 exists). The equalities in lemma 1.2 arise from the following scheme of probability assignments:

$P''(h_1 \wedge h_2 \wedge e) = x_1 x_2 y_1 y_2 w$; $P''(\neg h_1 \wedge h_2 \wedge e) = (1 - x_1 y_1) x_2 y_2 w$;
$P''(h_1 \wedge h_2 \wedge \neg e) = y_1 y_2 (1 - w)$; $P''(\neg h_1 \wedge h_2 \wedge \neg e) = (1 - y_1) y_2 (1 - w)$;
$P''(h_1 \wedge \neg h_2 \wedge e) = x y_1 (1 - x_2 y_2) w$; $P''(\neg h_1 \wedge \neg h_2 \wedge e) = (1 - x y_1)(1 - x_2 y_2) w$;
$P''(h_1 \wedge \neg h_2 \wedge \neg e) = y_1 (1 - y_2)(1 - w)$; $P''(\neg h_1 \wedge \neg h_2 \wedge \neg e) = (1 - y_1)(1 - y_2)(1 - w)$.

Suppose there exist $x_1, x_2 \in D_f$ such that $x_1 > x_2$ and $f(x_1) \leq f(x_2)$. Then, by lemma 1.2 and the definition of D_f, there exist $e, h_1, h_2 \in L_c$ and $P'' \in \boldsymbol{P}$ such that $P''(h_1|e)/P''(h_1|\neg e) = x_1$ and $P''(h_2|e)/P''(h_2|\neg e) = x_2$. By the probability calculus, if the latter equalities hold, then $P''(e|h_1) > P''(e|h_2)$. Thus, there exist $e, h_1, h_2 \in L_c$ and $P'' \in \boldsymbol{P}$ such that $E_{P''}(e, h_1) = f(x_1) \leq f(x_2) = E_{P''}(e, h_2)$, even if $P''(e|h_1) > P''(e|h_2)$, contradicting E1. Conversely, E1 implies that, for any $x_1, x_2 \in D_f$, if $x_1 > x_2$ then $f(x_1) > f(x_2)$. By a similar argument, E1 also implies that, for any $x_1, x_2 \in D_f$, if $x_1 = x_2$ then $f(x_1) = f(x_2)$. So, for E1 to hold, it must be that, for any $e, h \in L_c$ and any $P \in \boldsymbol{P}$, $E_P(e, h) = f[P(h|e)/P(h|\neg e)]$, and f is a strictly increasing function. This completes the proof of theorem 1.

Appendix B

Theorem 2. E0, E1, E2*, and E3 hold iff there exists a strictly increasing function f such that, for any $e, h \in L_c$ and any $P \in \boldsymbol{P}$, $E_P(e, h) = f[\varepsilon^*(e, h)]$, where

$$\varepsilon^*(e,h) = \begin{cases} \dfrac{P(e|h)-P(e)}{1-P(e)} & \text{if } P(e|h) \geq P(e) \\ \dfrac{P(e|h)-P(e)}{P(e)} & \text{if } P(e|h) < P(e) \end{cases}.$$

Proof.

Right-to-Left Implication.

E0. If there exists a strictly increasing function f such that $E_P(e,h) = f[\varepsilon^*(e,h)]$, then E0 is trivially satisfied.

E1. Let $e,\ h_1,\ h_2 \in L_c$ be given. Three classes of cases can obtain. (i) Let $P \in \boldsymbol{P}$ be such that $P(e|h_1) \gtreqless P(e)$ and $P(e|h_2) \lesseqgtr P(e)$. It is easy to verify that, for any $e, h \in L_c$, $P(e|h) \gtreqless P(e)$ iff $\varepsilon^*(e,h) \gtreqless 0$. So we have that, for any $e, h_1, h_2 \in L_c$, $P(e|h_1) \gtreqless P(e)$ iff $\varepsilon^*(e,h_1) \gtreqless 0$ and $P(e|h_2) \lesseqgtr P(e)$ iff $\varepsilon^*(e,h_2) \lesseqgtr 0$. It follows that, for any $e,\ h_1, h_2 \in L_c$, $P(e|h_1) \gtreqless P(e|h_2)$ iff $\varepsilon^*(e,h_1) \gtreqless \varepsilon^*(e,h_2)$. (ii) Let $P \in \boldsymbol{P}$ be such that $P(e|h_1) \geq P(e)$ and $P(e|h_2) \geq P(e)$. Then we have that, for any $e,\ h_1,\ h_2 \in L_c$, $P(e|h_1) \gtreqless P(e|h_2)$ iff $P(\neg e|h_1) \lesseqgtr P(\neg e|h_2)$ iff $P(\neg e|h_1)/P(\neg e) \lesseqgtr P(\neg e|h_2)/P(\neg e)$ iff $1 - P(\neg e|h_1)/P(\neg e) \gtreqless 1 - P(\neg e|h_2)/P(\neg e)$ iff $\varepsilon^*(e,h_1) \gtreqless \varepsilon^*(e,h_2)$. (iii) Finally, let $P \in \boldsymbol{P}$ be such that $P(e|h_1) \leq P(e)$ and $P(e|h_2) \leq P(e)$. Then we have that, for any $e, h_1, h_2 \in L_c$, $P(e|h_1) \gtreqless P(e|h_2)$ iff $P(e|h_1)/P(e) \gtreqless P(e|h_2)/P(e)$ iff $P(e|h_1)/P(e) - 1 \gtreqless P(e|h_2)/P(e) - 1$ iff $\varepsilon^*(e,h_1) \gtreqless \varepsilon^*(e,h_2)$. As i–iii are exhaustive, for any $e, h_1, h_2 \in L_c$ and any $P \in \boldsymbol{P}$, $P(e|h_1) \gtreqless P(e|h_2)$ iff $\varepsilon^*(e,h_1) \gtreqless \varepsilon^*(e,h_2)$. By ordinal equivalence, if there exists a strictly increasing function f such that $E_P(e,h) = f[\varepsilon^*(e,h)]$, then E1 follows.

E2*. Let $e,\ e^*,\ h \in L_c$ be given. (i) Let $P \in \boldsymbol{P}$ be given so that $P(e|h) > P(e)$. Then, for any $e,\ e^*h \in L_c$, if e^* is probabilistically independent from e, h, and their conjunction, $P(h|e) = P(h|e \wedge e^*)$ iff $P(h|e) - P(h) = P(h|e \wedge e^*) - P(h)$, by which $[P(h|e) - P(h)]P(e) > [P(h|e \wedge e^*) - P(h)]P(e \wedge e^*)$, which obtains iff $P(h \wedge e) - P(h)P(e) > P(h \wedge e \wedge e^*) - P(h)P(e \wedge e^*)$ iff $P(e|h) - P(e) > P(e \wedge e^*|h) - P(e \wedge e^*)$, by which $[P(e|h) - P(e)]/[1 - P(e)] > [P(e \wedge e^*|h) - P(e \wedge e^*)]/[1 - P(e \wedge e^*)]$, which obtains iff $\varepsilon^*(e,h) > \varepsilon^*(e \wedge e^*, h)$. By ordinal equivalence, if there exists a strictly increasing function f such that $E_P(e,h) = f[\varepsilon^*(e,h)]$, then E2*i follows. (ii) Let $P \in \boldsymbol{P}$ be given so that $P(e|h) \leq P(e)$. Then, for any $e,\ e^*h \in L_c$, if e^* is probabilistically independent from e, h, and their conjunction, $P(e|h)/P(e) = P(e \wedge e^*|h)/P(e \wedge e^*)$ iff $P(e|h)/P(e) - 1 = P(e \wedge e^*|h)/P(e \wedge e^*) - 1$ iff $\varepsilon^*(e,h) = \varepsilon^*(e \wedge e^*, h)$.

By ordinal equivalence, if there exists a strictly increasing function f such that $E_P(e,h) = f[\varepsilon^*(e,h)]$, then E2*ii follows.

E3. Let $e_1, e_2, h \in L_c$ and $P \in \boldsymbol{P}$ be given. Three classes of cases can obtain. (i) Let $P \in \boldsymbol{P}$ be such that $P(e_1|h) \gtreqless P(e_1)$ and $P(e_2|h) \lesseqgtr P(e_2)$. It is easy to verify that, for any $e, h \in L_c$, $P(e|h) \gtreqless P(e)$ iff $\varepsilon^*(e,h) \gtreqless 0$ iff $P(\neg e|h) \lesseqgtr P(\neg e)$ iff $\varepsilon^*(\neg e, h) \lesseqgtr 0$. So we have that, for any e_1, $e_2, h \in L_c$, $P(e_1|h) \gtreqless P(e_1)$ iff $\varepsilon^*(e_1,h) \gtreqless 0$ iff $P(\neg e_1|h) \lesseqgtr P(\neg e_1)$ iff $\varepsilon^*(\neg e_1, h) \lesseqgtr 0$ and $P(e_2 \mid h) \gtreqless P(e_2)$ iff $\varepsilon^*(e_2, h) \gtreqless 0$ $\neg P(\neg e_2 \mid h) \lesseqgtr P(\neg e_2)$ iff $\varepsilon^*(\neg e_2, h) \lesseqgtr 0$. It follows that, for any $e_1, e_2, h \in L_c$, $\varepsilon^*(e_1,h) \gtreqless \varepsilon^*(e_2,h)$ iff $\varepsilon^*(\neg e_1, h) \lesseqgtr \varepsilon^*(\neg e_2, h)$. (ii) Let $P \in \boldsymbol{P}$ be such that $P(e_1|h) \geq P(e_1)$ and $P(e_2|h) \geq P(e_2)$. Then we have that, for any $e_1, e_2, h \in L_c$, $\varepsilon^*(e_1,h) \gtreqless \varepsilon^*(e_2,h)$ iff $1 - P(\neg e_1|h)/P(\neg e_1) \gtreqless 1 - P(\neg e_2|h)/P(\neg e_2)$ iff $P(\neg e_1|h)/P(\neg e_1) \lesseqgtr P(\neg e_2|h)/P(\neg e_2)$ iff $P(\neg e_1|h)/P(\neg e_1) - 1 \lesseqgtr P(\neg e_2|h)/P(\neg e_2) - 1$ iff $\varepsilon^*(\neg e_1, h) \lesseqgtr \varepsilon^*(\neg e_2, h)$. (iii) Finally, let $P \in \boldsymbol{P}$ be such that $P(e_1|h) \leq P(e_1)$ and $P(e_2|h) \leq P(e_2)$. Then we have that, for any $e_1, e_2, h \in L_c$, $\varepsilon^*(e_1,h) \gtreqless \varepsilon^*(e_2,h)$ iff $P(e_1|h)/P(e_1) - 1 \gtreqless P(e_2|h)/P(e_2) - 1$ iff $P(e_1|h)/P(e_1) \gtreqless P(e_2 \mid h)/P(e_2)$ iff $1 - P(e_1 \mid h)/P(e_1) \lesseqgtr 1 - P(e_2 \mid h)/P(e_2)$ iff $\varepsilon^*(\neg e_1, h) \lesseqgtr \varepsilon^*(\neg e_2, h)$. As i–iii are exhaustive, for any e_1, e_2, $h \in L_c$ and any $P \in \boldsymbol{P}$, $\varepsilon^*(e_1,h) \gtreqless \varepsilon^*(e_2,h)$ iff $\varepsilon^*(\neg e_1, h) \lesseqgtr \varepsilon^*(\neg e_2, h)$. By ordinal equivalence, if there exists a strictly increasing function f such that $E_P(e,h) = f[\varepsilon^*(e,h)]$, then E3 follows.

Left-to-Right Implication.

The Case of Explanatory Failure $(P(e|h)) \leq P(e))$. Note that $P(e \wedge h) = [P(e|h)/P(e)]P(e)P(h)$. As a consequence, by E0, there exists a function j such that, for any $e, h \in L_c$ and any $P \in \boldsymbol{P}$, $E_P(e,h) = j[P(e|h)/P(e), P(e), P(h)]$. With no loss of generality, we will convey probabilistic coherence, regularity, and explanatory failure by constraining the domain of j to include triplets of values (x, y, w) such that the following conditions are jointly satisfied:

- $0 < y, w < 1$;
- $x \geq 0$, by which $x = P(e|h)/P(e) \geq 0$, so that $P(e|h) \geq 0$, and thus $P(e \wedge h) \geq 0$;
- $x \leq 1$ (conveying explanatory failure; that is, $P(e|h) \leq P(e)$), by which $xy = P(e|h) < 1$, so that $P(e \wedge h) < P(h)$, and thus $P(\neg e \wedge h) > 0$, and $xw = P(h|e) < 1$, so that $P(e \wedge h) < P(e)$, and thus $P(\neg e \wedge h) > 0$;
- $x \geq (y + w - 1)/yw$, by which $xyw = P(e \wedge h) \geq P(e) + P(h) - 1 = y + w - 1$, and thus $P(e \wedge h) + P(\neg e \wedge h) + P(e \wedge \neg h) \leq 1$.

We thus posit $j : \{(x, y, w) \in [0, 1] \times (0, 1)^2 | x \geq (y + w - 1)/yw\} \rightarrow \Re$ and denote the domain of j as D_j.

Lemma 2.1. For any x, y, w_1, w_2 such that $x \in [0,1]$, $y, w_1, w_2 \in (0, 1)$, and $x \geq (y + w_1 - 1)/yw_1, (y + w_2 - 1)/yw_2$, there exist $e, h_1, h_2 \in L_c$ and $P' \in \boldsymbol{P}$ such that $P'(e|h_1)/P'(e) = P'(e|h_2)/P'(e) = x$, $P'(e) = y$, $P'(h_1) = w_1$, and $P'(h_2) = w_2$.

Proof. The equalities in lemma 2.1 arise from the following scheme of probability assignments:

$$P'(h_1 \wedge h_2 \wedge e) = (xw_1)(xw_2)y; \qquad P'(\neg h_1 \wedge h_2 \wedge e) = (1 - xw_1)(xw_2)y;$$
$$P'(h_1 \wedge h_2 \neg e) = \frac{(1 - xy)^2 w_1 w_2}{(1 - y)}; \qquad P'(\neg h_1 \wedge h_2 \wedge \neg e) = \left[1 - \frac{(1 - xy)w_1}{(1 - y)}\right](1 - xy)w_2;$$
$$P'(h_1 \wedge \neg h_2 \wedge e) = (xw_1)(1 - xw_2)y; \qquad P'(\neg h_1 \wedge \neg h_2 \wedge e) = (1 - xw_1)(1 - xw_2)y;$$
$$P'(h_1 \wedge \neg h_2 \wedge \neg e) = (1 - xy)w_1\left[1 - \frac{(1 - xy)w_2}{(1 - y)}\right]; \qquad P'(\neg h_1 \wedge \neg h_2 \wedge \neg e) = \left[1 - \frac{(1 - xy)w_1}{(1 - y)}\right]\left[1 - \frac{(1 - xy)w_2}{(1 - y)}\right](1 - y).$$

Suppose there exist (x, y, w_1), $(x, y, w_2) \in D_j$ such that $j(x, y, w_1) \neq j(x, y, w_2)$. Then, by lemma 2.1 and the definition of D_j, there exist $e, h_1, h_2 \in L_c$ and $P' \in \boldsymbol{P}$ such that $P'(e|h_1)/P'(e) = P'(e|h_2)/P'(e) = x$, $P'(e) = y$, $P'(h_1) = w_1$, and $P'(h_2) = w_2$. Clearly, if the latter equalities hold, then $P'(e|h_1) = P'(e|h_2)$. Thus, there exist e, h_1, $h_2 \in L_c$ and $P' \in \boldsymbol{P}$ such that $E_{P'}(e, h_1) = j(x, y, w_1) \neq j(x, y, w_2) = E_{P'}(e, h_2)$, even if $P'(e|h_1) = P'(e|h_2)$, contradicting E1. Conversely, E1 implies that, for any (x, y, w_1), $(x, y, w_2) \in D_j$, $j(x, y, w_1) = j(x, y, w_2)$. So, for E1 to hold, there must exist k such that, for any $e, h \in L_c$ and any $P \in \boldsymbol{P}$, if $P(e|h) \leq P(e)$, then $E_p(e, h) = k[P(e|h)/P(e), P(e)]$, and $k(x, y) = j(x, y, w)$. We thus posit $k : \{(x, y) \in [0, 1] \times (0, 1)\} \rightarrow \Re$ and denote the domain of k as D_k.

Lemma 2.2. For any x, y_1, y_2 such that $x \in [0, 1]$, y_1, $y_2 \in (0, 1)$, and $y_1 > y_2$, there exist $e, e^*, h \in L_c$ and $P'' \in \boldsymbol{P}$ such that $P''(e|h)/P''(e) = x$, $P''(e) = y_1$, $P''(e \wedge e^*) = y_2$, $P''(e \wedge e^*) = P''(e)P''(e^*)$, $P''(h \wedge e^*) = P''(h)P''(e^*)$, and $P''(h \wedge e \wedge e^*) = P''(h \wedge e)P''(e^*)$.

Proof. Let $w \in (0, 1)$ be given so that $w \leq (1 - y_1)/(1 - xy_1)$ (as the latter quantity must be positive, w exists). The equalities in lemma 2.2 arise from the following scheme of probability assignments:

$$P''(h \wedge e \wedge e^*) = xwy_2; \qquad P''(\neg h \wedge e \wedge e^*) = (1 - xw)y_2;$$
$$P''(h \wedge e \wedge \neg e^*) = xw(y_1 - y_2); \qquad P''(\neg h \wedge e \wedge \neg e^*) = (1 - xw)(y_1 - y_2);$$
$$P''(h \wedge \neg e \wedge e^*) = \left[\frac{(1 - xy_1)w}{(1 - y_1)}\right]\left[\frac{(1 - y_1)y_2}{y_1}\right]; \qquad P''(\neg h \wedge \neg e \wedge \neg e^*) = \left[1 - \frac{(1 - xy_1)w}{(1 - y_1)}\right]\left[\frac{(1 - y_1)y_2}{y_1}\right];$$
$$P''(h_1 \wedge \neg e \wedge \neg e^*) = \left[\frac{(1 - xy_1)w}{(1 - y_1)}\right]\left[\frac{(1 - y_1)(y_1 - y_2)}{y_1}\right]; \qquad P''(\neg h \wedge \neg e \wedge \neg e^*) = \left[1 - \frac{(1 - xy_1)w}{(1 - y_1)}\right]\left[\frac{(1 - y_1)(y_1 - y_2)}{y_1}\right].$$

Suppose there exist (x, y_1), $(x, y_2) \in D_k$ such that $k(x, y_1) \neq k(x, y_2)$. Assume $y_1 > y_2$ with no loss of generality. Then, by lemma 2.2 and the definition of D_k, there exist e, e^*, $h \in L_c$ and $P'' \in \boldsymbol{P}$ such that $P''(e|h)/P''(e) = x$, $P''(e) = y_1$, $P''(e \wedge e^*) = y_2$, $P''(e \wedge e^*) = P''(e)P''(e^*)$, $P''(h \wedge e^*) = P''(h)P(e)$, and $P''(h \wedge e \wedge e^*) = P''(h \wedge e)P''(e^*)$. If the latter equalities hold, then $P''(e|h) \leq P''(e)$, and e* is probabilistically independent from e, h, and their conjunction. Thus, there exist e, e^*, $h \in L_c$ and $P'' \in \boldsymbol{P}$ such that $E_{p''}(e, h) = k(x, y_1) \neq k(x, y_2) = E_{p''}(e \wedge e^*, h)$, even if $P''(e|h) \leq P''(e)$ and e* is probabilistically independent from e, h, and their conjunction, contradicting E2*ii. Conversely, E2* implies that, for any $(x, y_1), (x, y_2) \in D_k$, $k(x, y_1) = k(x, y_2)$. So, for E2* to hold, there must exist m such that, for any $e, h, \in L_c$ and any $P \in \boldsymbol{P}$, if $P(e|h) \leq P(e)$, then $E_p(e, h) = m[P(e\,|\,h)\,/\,P(e)]$, and $m(x) = k(x, y)$. We thus posit $m : [0, 1] \rightarrow \Re$ and denote the domain of m as D_m.

Lemma 2.3. For any $x_1,\ x_2 \in [0, 1]$, there exist $e,\ h_1,\ h_2 \in L_c$ and $P''' \in \boldsymbol{P}$ such that $P'''(e|h_1)/P'''(e) = x_1$ and $P'''(e|h_2)/P'''(e) = x_2$.

Proof. Let $y, w_1, w_2 \in (0, 1)$ be given so that $w_1 \leq (1-y)/(1-x_1y)$ (as the latter quantity must be positive, w_1 exists) and $w_2 \leq (1-y)/(1-x_2y)$ (as the latter quantity must be positive, w_2 exists). The equalities in lemma 2.3 arise from the following scheme of probability assignments:

$$P'''(h_1 \wedge h_2 \wedge e) = (x_1w_1)(x_2w_2)y; \qquad P'''(\neg h_1 \wedge h_2 \wedge e) = (1-x_1w_1)(x_2w_2)y;$$
$$P'''(h_1 \wedge h_2 \wedge \neg e) = \frac{(1-x_1y)(1-{}_2y)w_1w_2}{(1-y)}; \qquad P'''(\neg h_1 \wedge h_2 \wedge \neg e) = \left[1-\frac{(1-x_1y)w_1}{(1-y)}\right](1-x_2y)w_2;$$
$$P'''(h_1 \wedge \neg h_2 \wedge e) = (x_1w_1)(x_2w_2)y; \qquad P'''(\neg h_1 \wedge \neg h_2 \wedge e) = (1-x_1w_1)(x_2w_2)y;$$
$$P'''(h_1 \wedge \neg h_2 \wedge \neg e) = (1-x_1y)w_1\left[1-\frac{(1-x_2y)w_2}{(1-y)}\right]; \qquad P'''(\neg h_1 \wedge \neg h_2 \wedge \neg e) = \left[1-\frac{(1-x_1y)w_1}{(1-y)}\right]\left[1-\frac{(1-x_2y)w_2}{(1-y)}\right](1-y).$$

Suppose there exist $x_1, x_2 \in D_m$ such that $x_1 > x_2$ and $m(x_1) \leq m(x_2)$. Then, by lemma 2.3 and the definition of D_m, there exist $e, h_1, h_2 \in L_c$ and $P''' \in \boldsymbol{P}$ such that $P'''(e|h_1)/P'''(e) = x_1$ and $P'''(e|h_2)/P'''(e) = x_2$. Clearly, if the latter equalities hold, then $P'''(e|h_1) > P'''(e|h_2)$. Thus, there exist $e,\ h_1,\ h_2 \in L_c$ and $P''' \in \boldsymbol{P}$ such that $E_{p'''}(e, h_1) = m(x_1) \leq m(x_2) = E_{p'''}(e, h_2)$, even if $P'''(e|h_1) > P'''(e|h_2)$, contradicting E1. Conversely, E1 implies that, for any $x_1,\ x_2 \in D_m$, if $x_1 > x_2$, then $m(x_1) > m(x_2)$. By a similar argument, E1 also implies that, for any $x_1,\ x_2 \in D_m$, if $x_1 = x_2$, then $m(x_1) = m(x_2)$. So, for E1 to hold, it must be that, for any $e, h, \in L_c$ and any $P \in \boldsymbol{P}$, if $P(e|h) \leq P(e)$, then $E_p(e, h) = m[P(e|h)/\,P(e)]$, and m is a strictly increasing function.

The Case of Explanatory Success $(P(e|h) > P(e))$. Notice that $P(e \wedge h) = \{1 - [P(\neg e|h)/P(\neg e)]P(\neg e)\}P(h)$ and $P(e) = 1 - P(\neg e)$. As a consequence, by E0, there exists a function r such that, for any e,

$h, \in L_c$ and any $P \in \boldsymbol{P}$, $E_p(e,h) = r[P(\neg e|h)/P(\neg e), P(\neg e), P(h)]$. With no loss of generality, we will convey probabilistic coherence, regularity, and explanatory success by constraining the domain of r to include triplets of values (x, y, w) such that the following conditions are jointly satisfied:

- $0 < y, w < 1$;
- $x \geq 0$, by which $x = P(\neg e|h)/P(\neg e) \geq 0$, so that $P(\neg e|h) \geq 0$, and thus $P(\neg e \wedge h) \geq 0$;
- $x < 1$ (conveying explanatory success; that is, $P(e|h) > P(e)$), by which $xy = P(\neg e|h) < 1$, so that $P(\neg e \wedge h) < P(h)$, and thus $P(e \wedge h) > 0$, and $xw = P(h|\neg e) < 1$, so that $P(\neg e \wedge h) < P(\neg e)$, and thus $P(\neg e \wedge \neg h) > 0$;
- $x \geq (y + w - 1)/yw$, by which $xyw = P(\neg e \wedge h) \geq P(\neg e) + P(h) - 1 = y + w - 1$, and thus $P(\neg e \wedge h) + P(e \wedge h) + P(\neg e \wedge \neg h) \leq 1$.

We thus posit $r : \{(x, y, w) \in [0, 1) \times (0, 1)^2 | x \geq (y + w - 1\}/yw\} \rightarrow \Re$ and denote the domain of r as D_r.

Lemma 2.4. For any x, y, w_1, w_2 such that $x \in [0, 1)$, $y, w_1, w_2 \in (0, 1)$, and $x \geq (y + w_1 - 1)/yw_1$, $(y + w_2 - 1)/yw_2$, there exist e, h_1, $h_2 \in L_c$ and $P' \in \boldsymbol{P}$ such that $P'(\neg e|h_1)/P'(\neg e) = P'(\neg e|h_2)/P'(\neg e) = x$, $P'(\neg e) = y$, $P'(h_1) = w_1$, and $P'(h_2) = w_2$.

Proof. The equalities in lemma 2.4 arise from the following scheme of probability assignments:

$$P'(h_1 \wedge h_2 \wedge e) = \frac{(1-xy)^2 w_1 w_2}{(1-y)}; \qquad P'(\neg h_1 \wedge h_2 \wedge e) = \left[1 - \frac{(1-xy)w_1}{(1-y)}\right](1-xy)w_2;$$
$$P'(h_1 \wedge h_2 \wedge \neg e) = (xw_1)(xw_2)y; \qquad P'(\neg h_1 \wedge h_2 \wedge \neg e) = (1-xw_1)(xw_2)y;$$
$$P'(h_1 \wedge \neg h_2 \wedge e) = (1-xy)w_1\left[1 - \frac{(1-xy)w_2}{(1-y)}\right]; \qquad P'(\neg h_1 \wedge \neg h_2 \wedge e) = \left[1 - \frac{(1-xy)w_1}{(1-y)}\right]\left[1 - \frac{(1-xy)w_2}{(1-y)}\right](1-y);$$
$$P'(h_1 \wedge \neg h_2 \wedge \neg e) = (xw_1)(1-xw_2)y; \qquad P'(\neg h_1 \wedge \neg h_2 \wedge \neg e) = (1-xw_1)(1-xw_2)y.$$

Suppose there exist $(x, y, w_1), (x, y, w_2) \in D_r$ such that $r(x, y, w_1) \neq r(x, y, w_2)$. Then, by lemma 2.4 and the definition of D_r, there exist e, h_1, $h_2 \in L_c$ and $P' \in \boldsymbol{P}$ such that $P'(\neg e|h_1)/P'(\neg e) = P'(\neg e|h_2)/P'(\neg e) = x$, $P'(\neg e) = y$, $P'(h_1) = w_1$, and $P'(h_2) = w_2$. By the probability calculus, if the latter equalities hold, then $P'(e|h_1) = P'(e|h_2)$. Thus, there exist e, h_1, $h_2 \in L_c$ and $P' \in \boldsymbol{P}$ such that $E_{p'}(e, h_1) = r(x, y, w_1) \neq r(x, y, w_2) = E_{p'}(e, h_2)$, even if $P'(e|h_1) = P'(e|h_2)$, contradicting E1. Conversely, E1 implies that, for any $(x, y, w_1), (w, y, w_2) \in D_r$, $r(x, y, w_1) = r(x, y, w_2)$. So, for E1 to hold, there must exist s such that, for any e, h, $\in L_c$ and any $P \in \boldsymbol{P}$, if $P(e|h) > P(e)$, then $E_p(e, h) = s[P(\neg e|h)/$

$P(\neg e), P(\neg e)]$, and $s(x,y) = r(x,y,w)$. We thus posit $s : \{(x,y) \in [0,1) \times (0,1)\} \rightarrow \Re$ and denote the domain of s as D_s.

Lemma 2.5. For any x, y_1, y_2 such that $x \in [0,1)$ and $y_1, y_2 \in (0,1)$, there exist e_1, e_2, $h \in L_c$ and $P'' \in \boldsymbol{P}$ such that $P''(\neg e_1|h)/P''(\neg e_1) = P''(\neg e_2|h)/P''(\neg e_2) = x$, $P''(\neg e_1) = y_1$, and $P''(\neg e_2) = y_2$.

Proof. Let $w \in (0,1)$ be given so that $w \leq (1-y_1)/(1-xy_1)$ (as the latter quantity must be positive, w exists) and $w \leq (1-y_2)/(1-xy_2)$ (as the latter quantity must be positive, w exists). The equalities in lemma 2.5 arise from the following scheme of probability assignments:

$$P''(h \wedge e_1 \wedge e_2) = (1-xy_1)(1-xy_2)w; \quad P''(\neg h \wedge e_1 \wedge e_2) = \left[1 - \frac{(1-xw)y_1}{(1-w)}\right]\left[1 - \frac{(1-xw)y_2}{(1-w)}\right](1-w);$$

$$P''(h \wedge e_1 \wedge \neg e_2) = (1-xy_1)(xy_2)w; \quad P''(\neg h \wedge e_1 \wedge \neg e_2) = \left[1 - \frac{(1-xw)y_1}{(1-w)}\right](1-xw)y_2;$$

$$P''(h \wedge \neg e_1 \wedge e_2) = (xy_1)(1-xy_2)w; \quad P''(\neg h \wedge e_1 \wedge e_2) = (1-xw)y_1\left[1 - \frac{(1-xw)y_2}{(1-w)}\right];$$

$$P''(h \wedge \neg e_1 \wedge \neg e_2) = (xy_1)(xy_2)w; \quad P''(\neg h \wedge \neg e_1 \wedge \neg e_2) = \frac{(1-xw)^2 y_1 y_2}{(1-w)}.$$

Suppose there exist $(x, y_1)(x, y_2) \in D_s$ such that $s(x,y_1) \neq s(x,y_2)$. Then, by lemma 2.5 and the definition of D_s, there exist e_1, e_2, $h \in L_c$ and $P'' \in \boldsymbol{P}$ such that $P''(\neg e_1|h)/P''(\neg e_1) = P''(\neg e_2|h)/P''(\neg e_2) = x$, $P''(\neg e_1) = y_1$, and $P''(\neg e_2) = y_2$. If the latter equalities hold, then $E_{P''}(\neg e_1, h) = m[P''(\neg e_1|h)/P''(\neg e_1)] = m[P''(\neg e_2|h)/P''(\neg e_2)] = E_{P''}(\neg e_2, h)$. Thus, there exist e_1, e_2, $h \in L_c$ and $P'' \in \boldsymbol{P}$ such that $E_{P''}(e_1, h) = s(x,y_1) \neq s(x,y_2) = E_{P''}(e_2, h)$, even if $E_{P''}(\neg e_1, h) = E_{P''}(\neg e_2, h)$, contradicting E3. So, for E3 to hold, there must exist t such that, for any e, $h \in L_c$ and any $P \in \boldsymbol{P}$, if $P(e|h) > P(e)$, then $E_P(e,h) = t[P(\neg e|h)/P(\neg e)]$, and $t(x) = s(x,y)$. We thus posit $t : [0,1) \rightarrow \Re$ and denote the domain of t as D_t.

Lemma 2.6. For any $x_1, x_2 \in [0,1)$, there exist $e, h_1, h_2 \in L_c$ and $P''' \in \boldsymbol{P}$ such that $P'''(\neg e|h_1)/P'''(\neg e) = x_1$, and $P'''(\neg e|h_2)/P'''(\neg e) = x_2$.

Proof. Let $y, w_1, w_2 \in (0,1)$ be given so that $w_1 \leq (1-y)/(1-x,y)$ (as the latter quantity must be positive, w_1 exists) and $w_2 \leq (1-y)/(1-x_2 y)$ (as the latter quantity must be positive, w_2 exists). The equalities in lemma 2.6 arise from the following scheme of probability assignments:

$$P'''(h_1 \wedge {}_2 \wedge e) = \frac{(1-x_1 y)(1-x_2 y)w_1 w_2}{(1-y)} \quad P'''(h_1 \wedge h_2 \wedge e = \left[1 - \frac{(1-x_1 y)w_1}{(1-y)}\right](1-x_2 y)w_2;$$

$$P'''(h_1 \wedge h_2 \wedge \neg e) = (x_1 w_1)(x_2 w_2)y; \quad P'''(\neg h_1 \wedge h_2 \wedge \neg e) = (1-x_1 w_1)(x_2 w_2)y;$$

$$P'''(h_1 \wedge \neg h_2 \wedge e) = (1-x_1 y)w_1\left[1 - \frac{(1-x_2 y)w_2}{(1-y)}\right]; \quad P'''(\neg h_1 \wedge \neg h_2 \wedge e) = \left[1 - \frac{(1-x_1 y)w_1}{(1-y)}\right]\left[1 - \frac{(1-x_2 y)w_2}{(1-y)}\right](1-y);$$

$$P'''(h_1 \wedge \neg h_2 \wedge \neg e) = (x_1 w_1)(x_2 w_2)y; \quad P'''(\neg h_1 \wedge \neg h_2 \wedge \neg e) = (1-x_1 w_1)(x_2 w_2)y.$$

Suppose there exist $x_1, x_2 \in D_t$ such that $x_1 > x_2$ and $t(x_1) \geq t(x_2)$. Then, by lemma 2.6 and the definition of D_t, there exist e, h_1, $h_2 \in L_c$ and $P''' \in \boldsymbol{P}$ such that $P'''(\neg e|h_1)/P'''(\neg e) = x_1$ and $P'''(\neg e|h_2)/P'''(\neg e) = x_2$. By the probability calculus, if the latter equalities hold, then $P'''(e|h_1) < P'''(e|h_2)$. Thus, there exist e, h_1, $h_2 \in L_c$ and $P''' \in \boldsymbol{P}$ such that $E_{P'''}(e, h_1) = t(x_1) \geq t(x_2) = E_{P'''}(e, h_2)$, even if $P'''(e|h_1) < P'''(e|h_2)$, contradicting E1. Conversely, E1 implies that, for any x_1, $x_2 \in D_t$, if $x_1 > x_2$, then $t(x_1) < t(x_2)$. By a similar argument, E1 also implies that, for any x_1, $x_2 \in D_t$, if $x_1 = x_2$, then $t(x_1) = t(x_2)$. So, for E1 to hold, it must be that, for any $e, h \in L_c$ and any $P \in \boldsymbol{P}$, if $P(e|h) > P(e)$, then $E_P(e, h) = t[P(\neg e|h)/P(\neg e)]$, and t is a strictly decreasing function.

Summing up, if E0, E1, E2*, and E3 hold, then for any e, $h \in L_c$ and any $P \in \boldsymbol{P}$, (i) if $P(e|h) \leq P(e)$, then $E_P(e, h) = m[P(e|h)/P(e)]$ and m is a strictly increasing function, thus $E_P(e, h)$ is a strictly increasing function of $\varepsilon^*(e, h)$, and (ii) if $P(e|h) > P(e)$, then $E_P(e, h) = t[P(\neg e|h)/P(\neg e)]$ and t is a strictly decreasing function, thus $E_P(e, h)$ is a strictly increasing function of $\varepsilon^*(e, h)$. As i and ii are exhaustive, for E0, E1, E2*, and E3 to hold, it must be that, for any $e, h \in L_c$ and any $P \in \boldsymbol{P}$, $E_P(e, h) = f[\varepsilon^*(e, h)]$, and f is a strictly increasing function. This completes the proof of theorem 2.

REFERENCES

Crupi, Vincenzo. 2012. “An Argument for Not Equating Confirmation and Explanatory Power.” *Reasoner* 6 (3): 39–40. Erratum, *Reasoner* 6 (4): 68

Crupi, Vincenzo, Roberto Festa, and Carlo Buttasi. 2010. “Towards a Grammar of Bayesian Confirmation.” In *Epistemology and Methodology of Science*, ed. Mauricio Suárez, Mauro Dorato, and Miklós Rédei, 73–93. Dordrecht: Springer.

Crupi, Vincenzo, and Katya Tentori. 2010. “Irrelevant Conjunction: Statement and Solution of a New Paradox.” *Philosophy of Science* 77:1–13.

———. Forthcoming. “Confirmation as Partial Entailment: A Representation Theorem in Inductive Logic.” *Journal of Applied Logic*.

Crupi, Vincenzo, Katya Tentori, and Michel Gonzalez. 2007. “On Bayesian Measures of Evidential Support: Theoretical and Empirical Issues.” *Philosophy of Science* 74:229–52.

Fitelson, Branden. 2005. “Inductive Logic.” In *Philosophy of Science: An Encyclopedia*, ed. Sahotra Sarkar and Jessica Pfeifer, 384–93. New York: Routledge.

———. 2006. “Logical Foundations of Evidential Support.” *Philosophy of Science* 73:500–512.

Good, Irving John. 1960. “Weight of Evidence, Corroboration, Explanatory Power, Information and the Utility of Experiments.” *Journal of the Royal Statistical Society* B 22:319–31.

Kemeny, John, and Paul Oppenheim. 1952. “Degrees of Factual Support.” *Philosophy of Science* 19:307–24.

McGrew, Timothy. 2003. “Confirmation, Heuristics, and Explanatory Reasoning.” *British Journal for the Philosophy of Science* 54:553–67.

Niiniluoto, Ilkka, and Raimo Tuomela. 1973. *Theoretical Concepts and Hypothetico-Inductive Inference*. Dordrecht: Reidel.

Pietarinen, Juhani. 1970. “Quantitative Tools for Evaluating Scientific Systematizations.” In *Information and Inference*, ed. Jaakko Hintikka and Patrick Suppes, 123–47. Dordrecht: Reidel.

Schupbach, Jonah, and Jan Sprenger. 2011. "The Logic of Explanatory Power." *Philosophy of Science* 78:105–27.

Tentori, Katya, Vincenzo Crupi, and Daniel Osherson. 2007. "Determinants of Confirmation." *Psychonomic Bulletin and Review* 14:877–83.

———. 2010. "Second-Order Probability Affects Hypothesis Confirmation." *Psychonomic Bulletin and Review* 17:129–34.

Worrall, John. 2006. "Miracles, Pessimism, and Scientific Realism." Unpublished manuscript, Department of Philosophy, Logic, and Scientific Method, London School of Economics. http://www2.lse.ac.uk/philosophy/WhosWho/staffhomepages/Publications/NMAandPIfinal.pdf.